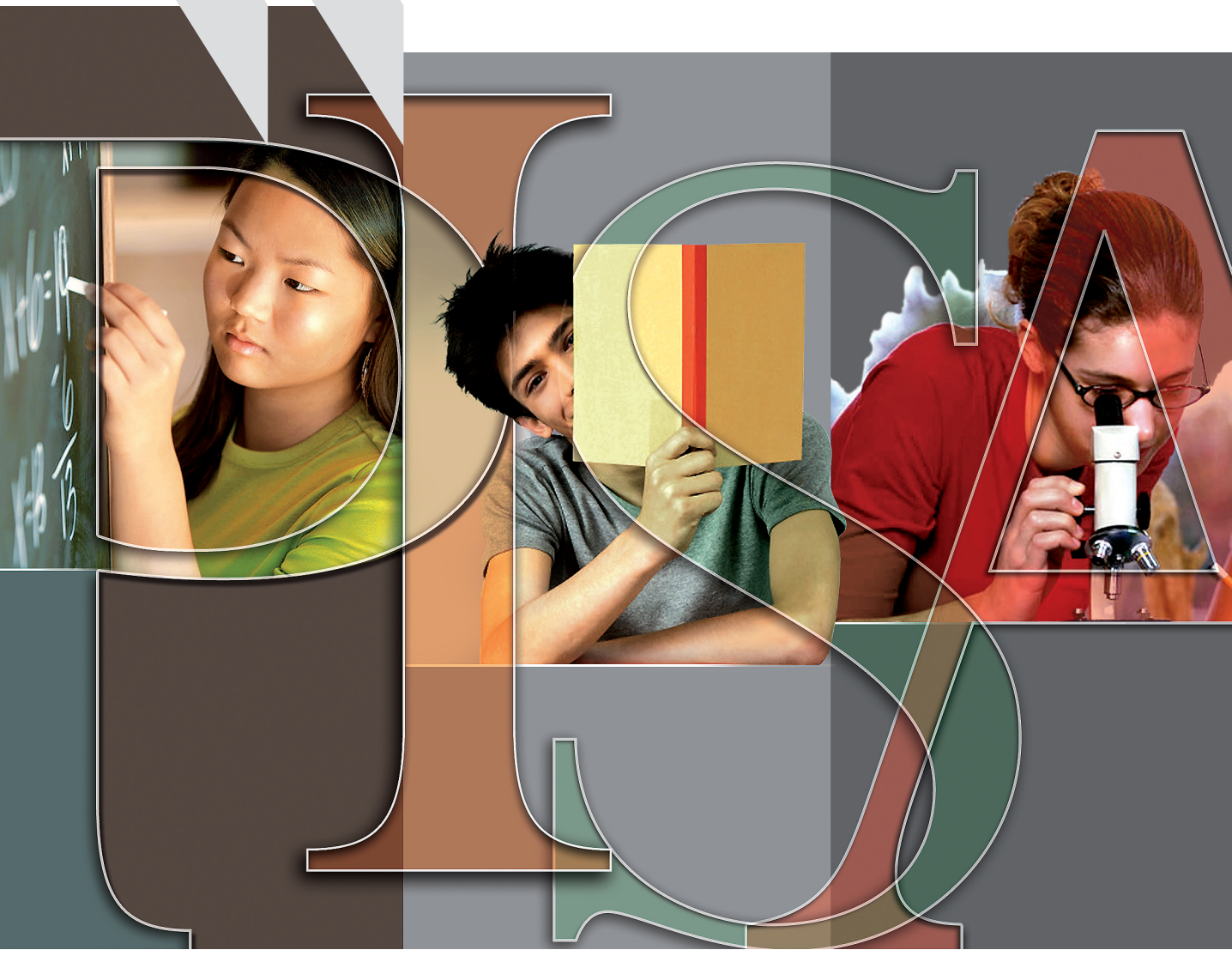
PISA 2006
Technical Report



Programme for International Student Assessment

OECD

PISA 2006
Technical Report

Programme for **I**nternational **S**tudent **A**ssessment



ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies 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, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities 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.

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

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

PISA™, OECD/PISA™ and the PISA logo are trademarks of the Organisation for Economic Co-operation and Development (OECD). All use of OECD trademarks is prohibited without written permission from the OECD.

© OECD 2009

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



Foreword

The OECD's 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 2006 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 2006 database, which includes information on nearly 400,000 students from 57 countries. The database includes not only information on student performance in the three main areas of assessment – reading, mathematics and science – but also their responses to the student questionnaire that they completed as part of the assessment. Data from the school principals of participating schools are also included. The PISA 2006 database was used to generate information and to act as a base for analysis for the production of the PISA 2006 initial report, *PISA 2006: Science Competencies for Tomorrow's World* (OECD, 2007).

The information in this report complements the *PISA 2006 Data Analysis Manuals* (OECD, 2009) which give detailed accounts of how to carry out the analyses of the information in the database.

PISA is a collaborative effort by the participating countries, and guided by their governments on the basis of shared policy-driven interests. Representatives of each country form the PISA Governing Board which decides on the assessment and reporting of results in PISA.

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 who acted as editor for this report, and his team, Steve Dept, Andrea Ferrari, Susan Fuss, Eveline Gebhardt, Beatrice Halleux, Sheila Krawchuk, Ron Martin, Martin Murphy, Alla Routitsky, Keith Rust, Wolfram Schulz, Ross Turner, and Erin Wilson. A full list of the contributors to the PISA project is included in Appendix 8 of this report. The editorial work at the OECD Secretariat was carried out by John Cresswell, Sophie Vayssettes and Elisabeth Villoutreix.



Ryo Watanabe
Chair of the PISA Governing Board



Barbara Ischinger
Director for Education, OECD

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



Table of contents

FOREWORD **3**

CHAPTER 1 **PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT: AN OVERVIEW** **19**

Participation 21

Features of PISA 22

Managing and implementing PISA 23

Organisation of this report 23

READER'S GUIDE **25**

CHAPTER 2 **TEST DESIGN AND TEST DEVELOPMENT** **27**

Test scope and format 28

Test design 28

Test development centres 29

Development timeline 30

The PISA 2006 scientific literacy framework 30

Test development – cognitive items 31

- Item development process 31
- National item submissions 33
- National review of items 34
- International item review 35
- Preparation of dual (English and French) source versions 35

Test development – attitudinal items 35

Field trial 38

- Field trial selection 38
- Field trial design 39
- Despatch of field trial instruments 40
- Field trial coder training 40
- Field trial coder queries 40
- Field trial outcomes 41
- National review of field trial items 42

Main study 42

- Main study science items 42
- Main study reading items 44
- Main study mathematics items 45
- Despatch of main study instruments 46
- Main study coder training 46
- Main study coder query service 46
- Review of main study item analyses 47

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



CHAPTER 3 **THE DEVELOPMENT OF THE PISA CONTEXT QUESTIONNAIRES** **49**

Overview 50

The conceptual structure 51
- A conceptual framework for PISA 2006 51

Research areas in PISA 2006 55

The development of the context questionnaires 57

The coverage of the questionnaire material 58
- Student questionnaire 58
- School questionnaire 59
- International options 59
- National questionnaire material 60

The implementation of the context questionnaires 60

CHAPTER 4 **SAMPLE DESIGN** **63**

Target population and overview of the sampling design 64

Population coverage, and school and student participation rate standards 65
- Coverage of the PISA international target population 65
- Accuracy and precision 66
- School response rates 66
- Student response rates 68

Main study school sample 68
- Definition of the national target population 68
- The sampling frame 69
- Stratification 70
- Assigning a measure of size to each school 74
- School sample selection 74
- PISA and TIMSS or PIRLS overlap control 76
- Student samples 82

CHAPTER 5 **TRANSLATION AND CULTURAL APPROPRIATENESS OF THE TEST AND SURVEY MATERIAL** **85**

Introduction 86

Development of source versions 86

Double translation from two source languages 87

PISA translation and adaptation guidelines 88

Translation training session 89

Testing languages and translation/adaptation procedures 89

International verification of the national versions 91
- VegaSuite 93
- Documentation 93
- Verification of test units 93
- Verification of the booklet shell 94
- Final optical check 94
- Verification of questionnaires and manuals 94
- Final check of coding guides 95
- Verification outcomes 95

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



Translation and verification outcomes – national version quality 96
- Analyses at the country level 96
- Analyses at the item level 103
- Summary of items lost at the national level, due to translation, printing or layout errors 104

CHAPTER 6 **FIELD OPERATIONS** **105**

Overview of roles and responsibilities 106
- National project managers 106
- School coordinators 107
- Test administrators 107
- School associates 108

The selection of the school sample 108

Preparation of test booklets, questionnaires and manuals 108

The selection of the student sample 109

Packaging and shipping materials 110

Receipt of materials at the national centre after testing 110

Coding of the tests and questionnaires 111
- Preparing for coding 111
- Logistics prior to coding 113
- Single coding design 115
- Multiple coding 117
- Managing the process coding 118
- Cross-national coding 120
- Questionnaire coding 120

Data entry, data checking and file submission 120
- Data entry 120
- Data checking 120
- Data submission 121
- After data were submitted 121

The main study review 121

CHAPTER 7 **QUALITY ASSURANCE** **123**

PISA quality control 124
- Comprehensive operational manuals 124
- National level implementation planning document 124

PISA quality monitoring 124
- Field trial and main study review 124
- Final optical check 126
- National centre quality monitor (NCQM) visits 126
- PISA quality monitor (PQM) visits 126
- Test administration 127
- Delivery 128

CHAPTER 8 **SURVEY WEIGHTING AND THE CALCULATION OF SAMPLING VARIANCE** **129**

Survey weighting 130

The school base weight 131
- The school weight trimming factor 132

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



- The student base weight 132
- School non-response adjustment 132
- Grade non-response adjustment 134
- Student non-response adjustment 135
- Trimming student weights 136
- Comparing the PISA 2006 student non-response adjustment strategy with the strategy used for PISA 2003 136
- The comparison 138

Calculating sampling variance 139
- The balanced repeated replication variance estimator 139
- Reflecting weighting adjustments 141
- Formation of variance strata 141
- Countries where all students were selected for PISA 141

CHAPTER 9 **SCALING PISA COGNITIVE DATA** **143**

The mixed coefficients multinomial logit model 144
- The population model 145
- Combined model 146

Application to PISA 146
- National calibrations 146
- National reports 147
- International calibration 153
- Student score generation 153

Booklet effects 155

Analysis of data with plausible values 156

Developing common scales for the purposes of trends 157
- Linking PISA 2003 and PISA 2006 for reading and mathematics 158
- Uncertainty in the link 158

CHAPTER 10 **DATA MANAGEMENT PROCEDURES** **163**

Introduction 164

KeyQuest 167

Data management at the national centre 167
- National modifications to the database 167
- Student sampling with *KeyQuest* 167
- Data entry quality control 167

Data cleaning at ACER 171
- Recoding of national adaptations 171
- Data cleaning organisation 171
- Cleaning reports 171
- General recodings 171

Final review of the data 172
- Review of the test and questionnaire data 172
- Review of the sampling data 172

Next steps in preparing the international database 172

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



CHAPTER 11 **SAMPLING OUTCOMES** **175**

Design effects and effective sample sizes 187
- Variability of the design effect 191
- Design effects in PISA for performance variables 191

Summary analyses of the design effect 203
- Countries with outlying standard errors 205

CHAPTER 12 **SCALING OUTCOMES** **207**

International characteristics of the item pool 208
- Test targeting 208
- Test reliability 208
- Domain inter-correlations 208
- Science scales 215

Scaling outcomes 216
- National item deletions 216
- International scaling 219
- Generating student scale scores 219

Test length analysis 219

Booklet effects 221
- Overview of the PISA cognitive reporting scales 232
- PISA overall literacy scales 234
- PISA literacy scales 234
- Special purpose scales 234

Observations concerning the construction of the PISA overall literacy scales 235
- Framework development 235
- Testing time and item characteristics 236
- Characteristics of each of the links 237

Transforming the plausible values to PISA scales 246
- Reading 246
- Mathematics 246
- Science 246
- Attitudinal scales 247

Link error 247

CHAPTER 13 **CODING AND MARKER RELIABILITY STUDIES** **249**

Homogeneity analyses 251

Multiple marking study outcomes (variance components) 254
- Generalisability coefficients 254

International coding review 261
- Background to changed procedures for PISA 2006 261
- ICR procedures 261
- Outcomes 264
- Cautions 270

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



CHAPTER 14 **DATA ADJUDICATION** **271**

Introduction 272
- Implementing the standards – quality assurance 272
- Information available for adjudication 273
- Data adjudication process 273

General outcomes 274
- Overview of response rate issues 274
- Detailed country comments 275

CHAPTER 15 **PROFICIENCY SCALE CONSTRUCTION** **283**

Introduction 284

Development of the described scales 285
- Stage 1: Identifying possible scales 285
- Stage 2: Assigning items to scales 286
- Stage 3: Skills audit 286
- Stage 4: Analysing field trial data 286
- Stage 5: Defining the dimensions 287
- Stage 6: Revising and refining with main study data 287
- Stage 7: Validating 287

Defining proficiency levels 287

Reporting the results for PISA science 290
- Building an item map 290
- Levels of scientific literacy 292
- Interpreting the scientific literacy levels 299

CHAPTER 16 **SCALING PROCEDURES AND CONSTRUCT VALIDATION OF CONTEXT QUESTIONNAIRE DATA** **303**

Overview 304

Simple questionnaire indices 304
- Student questionnaire indices 304
- School questionnaire indices 307
- Parent questionnaire indices 309

Scaling methodology and construct validation 310
- Scaling procedures 310
- Construct validation 312
- Describing questionnaire scale indices 314

Questionnaire scale indices 315
- Student scale indices 315
- School questionnaire scale indices 340
- Parent questionnaire scale indices 342
- The PISA index of economic, social and cultural status (ESCS) 346

CHAPTER 17 **VALIDATION OF THE EMBEDDED ATTITUDINAL SCALES** **351**

Introduction 352

International scalability 353
- Analysis of item dimensionality with exploratory and confirmatory factor analysis 353
- Fit to item response model 353

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



- Reliability 355
- Differential item functioning 355
- Summary of scalability 357

Relationship and comparisons with other variables 357
- Within-country student level correlations with achievement and selected background variables 358
- Relationships between embedded scales and questionnaire 360
- Country level correlations with achievement and selected background variables 361
- Variance decomposition 363
- Observations from other cross-national data collections 363
- Summary of relations with other variables 364

Conclusion 364

CHAPTER 18 **INTERNATIONAL DATABASE** **367**

Files in the database 368
- Student files 368
- School file 370
- Parent file 370

Records in the database 371
- Records included in the database 371
- Records excluded from the database 371

Representing missing data 371

How are students and schools identified? 372

Further information 373

REFERENCES **375**

APPENDICES **379**

Appendix 1 PISA 2006 main study item pool characteristics 380

Appendix 2 Contrast coding used in conditioning 389

Appendix 3 Design effect tables 399

Appendix 4 Changes to core questionnaire items from 2003 to 2006 405

Appendix 5 Mapping of ISCED to years 411

Appendix 6 National household possession items 412

Appendix 7 Exploratory and confirmatory factor analyses for the embedded items 414

Appendix 8 PISA consortium, staff and consultants 416

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



LIST OF BOXES

Box 1.1 Core features of PISA 2006 22

LIST OF FIGURES

Figure 2.1 Main study Interest in Science item 36
Figure 2.2 Main study Support for Scientific Enquiry item 36
Figure 2.3 Field trial Match-the-opinion Responsibility item 37

Figure 3.1 Conceptual grid of variable types 52
Figure 3.2 The two-dimensional conceptual matrix with examples of variables collected or available from other sources 54

Figure 4.1 School response rate standard 67

Figure 6.1 Design for the single coding of science and mathematics 115
Figure 6.2 Design for the single coding of reading 116

Figure 9.1 Example of item statistics in Report 1 148
Figure 9.2 Example of item statistics in Report 2 149
Figure 9.3 Example of item statistics shown in Graph B 150
Figure 9.4 Example of item statistics shown in Graph C 151
Figure 9.5 Example of item statistics shown in Table D 151
Figure 9.6 Example of summary of dodgy items for a country in Report 3a 152
Figure 9.7 Example of summary of dodgy items in Report 3b 152

Figure 10.1 Data management in relation to other parts of PISA 164
Figure 10.2 Major data management stages in PISA 166
Figure 10.3 Validity reports - general hierarchy 170

Figure 11.1 Standard error on a mean estimate depending on the intraclass correlation 188
Figure 11.2 Relationship between the standard error for the science performance mean and the intraclass correlation within explicit strata (PISA 2006) 205

Figure 12.1 Item plot for mathematics items 210
Figure 12.2 Item plot for reading items 211
Figure 12.3 Item plot for science items 212
Figure 12.4 Item plot for interest items 213
Figure 12.5 Item plot for support items 214
Figure 12.6 Scatter plot of per cent correct for reading link items in PISA 2000 and PISA 2003 238
Figure 12.7 Scatter plot of per cent correct for reading link items in PISA 2003 and PISA 2006 240
Figure 12.8 Scatter plot of per cent correct for mathematics link items in PISA 2003 and PISA 2006 242
Figure 12.9 Scatter plot of per cent correct for science link items in PISA 2000 and PISA 2003 244
Figure 12.10 Scatter plot of per cent correct for science link items in PISA 2003 and PISA 2006 245

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



Figure 13.1 Variability of the homogeneity indices for science items in field trial 250
Figure 13.2 Average of the homogeneity indices for science items in field trial and main study 251
Figure 13.3 Variability of the homogeneity indices for each science item in the main study 252
Figure 13.4 Variability of the homogeneity indices for each reading item in the main study 252
Figure 13.5 Variability of the homogeneity indices for each mathematics item 252
Figure 13.6 Variability of the homogeneity indices for the participating countries in the main study 253
Figure 13.7 Example of ICR report (reading) 269

Figure 14.1 Attained school response rates 274

Figure 15.1 The relationship between items and students on a proficiency scale 285
Figure 15.2 What it means to be at a level 289
Figure 15.3 A map for selected science items 291
Figure 15.4 Summary descriptions of the six proficiency levels on the science scale 294
Figure 15.5 Summary descriptions of six proficiency levels in *identifying scientific issues* 295
Figure 15.6 Summary descriptions of six proficiency levels in *explaining phenomena scientifically* 297
Figure 15.7 Summary descriptions of six proficiency levels in *using scientific evidence* 300

Figure 16.1 Summed category probabilities for fictitious item 314
Figure 16.2 Fictitious example of an item map 315
Figure 16.3 Scatterplot of country means for ESCS 2003 and ESCS 2006 347

Figure 17.1 Distribution of item fit mean square statistics for embedded attitude items 354
Figure 17.2 An example of the ESC plot for item S408RNA 356
Figure 17.3 Scatterplot of mean mathematics interest against mean mathematics for PISA 2003 363

LIST OF TABLES

Table 1.1 PISA 2006 participants 21

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

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



Table 3.1 Themes and constructs/variables in PISA 2006 56

Table 4.1 Stratification variables 71
Table 4.2 Schedule of school sampling activities 78

Table 5.1 Countries sharing a common version with national adaptations 90
Table 5.2 PISA 2006 translation/adaptation procedures 91
Table 5.3 Mean deviation and root mean squared error of the item by country interactions for each version 97
Table 5.4 Correlation between national item parameter estimates for Arabic versions 99
Table 5.5 Correlation between national item parameter estimates for Chinese versions 99
Table 5.6 Correlation between national item parameter estimates for Dutch versions 99
Table 5.7 Correlation between national item parameter estimates for English versions 99
Table 5.8 Correlation between national item parameter estimates for French versions 99
Table 5.9 Correlation between national item parameter estimates for German versions 100
Table 5.10 Correlation between national item parameter estimates for Hungarian versions 100
Table 5.11 Correlation between national item parameter estimates for Italian versions 100
Table 5.12 Correlation between national item parameter estimates for Portuguese versions 100
Table 5.13 Correlation between national item parameter estimates for Russian versions 100
Table 5.14 Correlation between national item parameter estimates for Spanish versions 100
Table 5.15 Correlation between national item parameter estimates for Swedish versions 100
Table 5.16 Correlation between national item parameter estimates within countries 101
Table 5.17 Variance estimate 102
Table 5.18 Variance estimates 103

Table 6.1 Design for the multiple coding of science and mathematics 118
Table 6.2 Design for the multiple coding of reading 118

Table 8.1 Non-response classes 133

Table 9.1 Deviation contrast coding scheme 154

Table 10.1 Double entry discrepancies per country: field trial data 169

Table 11.1 Sampling and coverage rates 178
Table 11.2 School response rates before replacement 182
Table 11.3 School response rates after replacement 184
Table 11.4 Student response rates after replacement 185
Table 11.5 Standard errors for the PISA 2006 combined science scale 189
Table 11.6 Design effect 1 by country, by domain and cycle 193
Table 11.7 Effective sample size 1 by country, by domain and cycle 194
Table 11.8 Design effect 2 by country, by domain and cycle 195
Table 11.9 Effective sample size 2 by country, by domain and cycle 196
Table 11.10 Design effect 3 by country, by domain and by cycle 197

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



Table 11.11 Effective sample size 3 by country, by domain and cycle 198
Table 11.12 Design effect 4 by country, by domain and cycle 199
Table 11.13 Effective sample size 4 by country, by domain and cycle 200
Table 11.14 Design effect 5 by country, by domain and cycle 201
Table 11.15 Effective sample size 5 by country, by domain and cycle 202
Table 11.16 Median of the design effect 3 per cycle and per domain across the 35 countries that participated in every cycle 203
Table 11.17 Median of the standard errors of the student performance mean estimate for each domain and PISA cycle for the 35 countries that participated in every cycle 203
Table 11.18 Median of the number of participating schools for each domain and PISA cycle for the 35 countries that participated in every cycle 204
Table 11.19 Median of the school variance estimate for each domain and PISA cycle for the 35 countries that participated in every cycle 204
Table 11.20 Median of the intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle 204
Table 11.21 Median of the within explicit strata intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle 205
Table 11.22 Median of the percentages of school variances explained by explicit stratification variables, for each domain and PISA cycle for the 35 countries that participated in every cycle 205

Table 12.1 Number of sampled student by country and booklet 209
Table 12.2 Reliabilities of each of the four overall scales when scaled separately 215
Table 12.3 Latent correlation between the five domains 215
Table 12.4 Latent correlation between science scales 215
Table 12.5 Items deleted at the national level 216
Table 12.6 Final reliability of the PISA scales 216
Table 12.7 National reliabilities for the main domains 217
Table 12.8 National reliabilities for the science subscales 218
Table 12.9 Average number of not-reached items and missing items by booklet 219
Table 12.10 Average number of not-reached items and missing items by country 220
Table 12.11 Distribution of not-reached items by booklet 221
Table 12.12 Estimated booklet effects on the PISA scale 221
Table 12.13 Estimated booklet effects in logits 221
Table 12.14 Variance in mathematics booklet means 222
Table 12.15 Variance in reading booklet means 224
Table 12.16 Variance in science booklet means 226
Table 12.17 Variance in interest booklet means 228
Table 12.18 Variance in support booklet means 230
Table 12.19 Summary of PISA cognitive reporting scales 233
Table 12.20 Linkage types among PISA domains 2000-2006 235
Table 12.21 Number of unique item minutes for each domain for each PISA assessments 237
Table 12.22 Numbers of link items between successive PISA assessments 237
Table 12.23 Per cent correct for reading link items in PISA 2000 and PISA 2003 238
Table 12.24 Per cent correct for reading link items in PISA 2003 and PISA 2006 239
Table 12.25 Per cent correct for mathematics link items in PISA 2003 and PISA 2006 241

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



Table 12.26 Per cent correct for science link items in PISA 2000 and PISA 2003 243
Table 12.27 Per cent correct for science link items in PISA 2003 and PISA 2006 245
Table 12.28 Link error estimates 247

Table 13.1 Variance components for mathematics 255
Table 13.2 Variance components for science 256
Table 13.3 Variance components for reading 257
Table 13.4 Generalisability estimates for mathematics 258
Table 13.5 Generalisability estimates for science 259
Table 13.6 Generalisability estimates for reading 260
Table 13.7 Examples of flagged cases 263
Table 13.8 Count of analysis groups showing potential bias, by domain 264
Table 13.9 Comparison of codes assigned by verifier and adjudicator 265
Table 13.10 Outcomes of ICR analysis part 1 265
Table 13.11 ICR outcomes by country and domain 266

Table 15.1 Scientific literacy performance band definitions on the PISA scale 293

Table 16.1 ISCO major group white-collar/blue-collar classification 306
Table 16.2 ISCO occupation categories classified as science-related occupations 307
Table 16.3 OECD means and standard deviations of WL estimates 311
Table 16.4 Median, minimum and maximum percentages of between-school variance for student-level indices across countries 313
Table 16.5 Household possessions and home background indices 316
Table 16.6 Scale reliabilities for home possession indices in OECD countries 317
Table 16.7 Scale reliabilities for home possession indices in partner countries/economies 318
Table 16.8 Item parameters for interest in science learning (INTSCIE) 318
Table 16.9 Item parameters for enjoyment of science (JOYSCIE) 319
Table 16.10 Model fit and estimated latent correlations for interest in and enjoyment of science learning 319
Table 16.11 Scale reliabilities for interest in and enjoyment of science learning 320
Table 16.12 Item parameters for instrumental motivation to learn science (INSTSCIE) 320
Table 16.13 Item parameters for future-oriented science motivation (SCIEFUT) 321
Table 16.14 Model fit and estimated latent correlations for motivation to learn science 321
Table 16.15 Scale reliabilities for instrumental and future-oriented science motivation 322
Table 16.16 Item parameters for science self-efficacy (SCIEEFF) 322
Table 16.17 Item parameters for science self-concept (SCSCIE) 323
Table 16.18 Model fit and estimated latent correlations for science self-efficacy and science self-concept 323
Table 16.19 Scale reliabilities for science self-efficacy and science self-concept 324
Table 16.20 Item parameters for general value of science (GENSCIE) 324
Table 16.21 Item parameters for personal value of science (PERSCIE) 325
Table 16.22 Model fit and estimated latent correlations for general and personal value of science 325
Table 16.23 Scale reliabilities for general and personal value of science 326
Table 16.24 Item parameters for science activities (SCIEACT) 326

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



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

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



Table 16.67 Scale reliabilities for parent questionnaire scales 345
Table 16.68 Factor loadings and internal consistency of ESCS 2006 in OECD countries 347
Table 16.69 Factor loadings and internal consistency of ESCS 2006 in partner countries/economies 348

Table 17.1 Student-level latent correlations between mathematics, reading, science, embedded interest and embedded support 354
Table 17.2 Summary of the IRT scaling results across countries 355
Table 17.3 Gender DIF table for embedded attitude items 357
Table 17.4 Correlation amongst attitudinal scales, performance scales and HISEI 358
Table 17.5 Correlations for science scale 359
Table 17.6 Loadings of the achievement, interest and support variables on three varimax rotated components 360
Table 17.7 Correlation between embedded attitude scales and questionnaire attitude scales 361
Table 17.8 Rank order correlation five test domains, questionnaire attitude scales and HISEI 362
Table 17.9 Intra-class correlation (rho) 362

Table A1.1 2006 Main study reading item classification 380
Table A1.2 2006 Main study mathematics item classification 381
Table A1.3 2006 Main study science item classification (cognitive) 383
Table A1.4 2006 Main study science embedded item classification (interest in learning science topics) 387
Table A1.5 2006 Main study science embedded item classification (support for scientific enquiry) 388

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables 389
Table A2.2 2006 Main study contrast coding used in conditioning for the ICT questionnaire variables 396
Table A2.3 2006 Main study contrast coding used in conditioning for the parent questionnaire variables and other variables 397

Table A3.1 Standard errors of the student performance mean estimate by country, by domain and cycle 399
Table A3.2 Sample sizes by country and cycle 400
Table A3.3 School variance estimate by country, by domain and cycle 401
Table A3.4 Intraclass correlation by country, by domain and cycle 402
Table A3.5 Within explicit strata intraclass correlation by country, by domain and cycle 403
Table A3.6 Percentages of school variance explained by explicit stratification variables, by domain and cycle 404

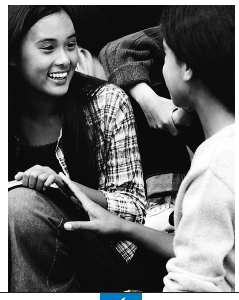
Table A4.1 Student questionnaire 405
Table A4.2 ICT familiarity questionnaire 407
Table A4.3 School questionnaire 408

Table A5.1 Mapping of ISCED to accumulated years of education 411

Table A6.1 National household possession items 412

Table A7.1 Exploratory and confirmatory factor analyses (EFA and CFA) for the embedded items 414

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



1

Programme for International Student Assessment: An Overview

Participation 21
Features of PISA 22
Managing and implementing PISA 23
Organisation of this report 23

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



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 2001), the second in 2003 and the third in 2006; the results of these surveys have been published in a series of reports (OECD, 2001, 2003, 2004, 2007) and a wide range of thematic and technical reports. The next survey will occur in 2009. For each assessment, one of the three areas (science, reading and mathematics) is chosen as the major domain and given greater emphasis. The remaining two areas, the minor domains, are assessed less thoroughly. In 2000 the major domain was reading; in 2003 it was mathematics and in 2006 it was science.

PISA is an age-based survey, assessing 15-year-old students in school in grade seven 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 2006 domain definitions are fully articulated in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD, 2006).

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. Further, PISA 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 2006 a number of countries[1] also administered a parent questionnaire to the parents of the students participating in PISA.

Using the data from these questionnaires, analyses linking contextual information with student achievement could address:

- Differences between countries in the relationships between student-level factors (such as gender and social 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;

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



- Differences in education systems and national context that are related to differences in student achievement across countries;
- Through links to PISA 2000 and PISA 2003, 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.

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 2001. PISA 2000 surveyed reading, mathematical and scientific literacy, with a primary focus on reading. The second PISA survey, conducted in 2003 in 41 countries, assessed reading, mathematical and scientific literacy, and problem solving with a primary focus on mathematical literacy. The third survey covered reading, mathematical and scientific literacy, with a primary focus on scientific literacy, and was conducted in 2006 in 57 countries. In some countries it was decided to carry out detailed analysis of some regions. In these 24 sub-national regions sufficient data were collected and quality control mechanisms implemented at a level that would permit OECD endorsement of their results. The participants in PISA 2006 are listed in Table 1.1. This report is concerned with the technical aspects of PISA 2006.

Table 1.1
PISA 2006 participants

OECD countries	Partner countries/economies
Australia	Argentina
Austria	Azerbaijan
Belgium	Brazil
Canada	Bulgaria
Czech Republic	Chile
Denmark	Colombia
Finland	Croatia
France	Estonia
Germany	Hong Kong-China
Greece	Indonesia
Hungary	Israel
Iceland	Jordan
Ireland	Kyrgyzstan
Italy	Latvia
Japan	Liechtenstein
Korea	Lithuania
Luxembourg	Macao-China
Mexico	Montenegro
Netherlands	Qatar
New Zealand	Romania
Norway	Russian Federation
Poland	Serbia
Portugal	Slovenia
Slovak Republic	Chinese Taipei
Spain	Thailand
Sweden	Tunisia
Switzerland	Uruguay
Turkey	
United Kingdom	
United States	

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



FEATURES OF PISA

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

- 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;
- The multilingual nature of the test, which involves rules and procedures designed to guarantee the equivalence of the different language versions used within and between participating countries, and taking 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;
- 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.

Box 1.1 Core features of PISA 2006

Sample size

- Nearly 400,000 students, representing almost 20 million 15-year-olds enrolled in the schools of the 57 participating countries and economies, were assessed in 2006.

Content

- PISA 2006 covered three domains: reading, mathematics and science.
- PISA 2006 looked at young people's ability to use their knowledge and skills in order to meet real-life challenges rather than how well they had mastered a specific school curriculum. The emphasis was placed on the mastery of processes, the understanding of concepts, and the ability to function in various situations within each domain.

Methods

- PISA 2006 used pencil-and-paper assessments, lasting two hours for each student.
- PISA 2006 used both multiple-choice items and questions requiring students to construct their own answers. Items were typically organised in units based on a passage describing a real-life situation.
- A total of six and a half hours of assessment items was included, with different students taking different combinations of the assessment items.
- Students answered a background questionnaire that took about 30 minutes to complete and, as part of an international option, completed questionnaires on their educational careers as well as familiarity with computers.
- School principals completed questionnaires about their schools.

Outcomes

- A profile of knowledge and skills among 15-year-olds.
- Contextual indicators relating results to student and school characteristics.
- A knowledge base for policy analysis and research.
- Trend indicators showing how results change over time.

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



This report describes the above-mentioned methodologies as they have been implemented in PISA 2006. Further, it 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 2006.

The ambitious goals of PISA come at a cost: PISA is both resource 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.

MANAGING AND IMPLEMENTING PISA

The design and implementation of PISA for the 2000, 2003 and 2006 data collections has been the responsibility of an international consortium led by the Australian Council for Educational Research (ACER) with Ray Adams as international project director. The other partners in this consortium have been the National Institute for Educational Measurement (Cito Group) in the Netherlands, Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe) at Université de Liège in Belgium, Westat and the Educational Testing Service (ETS) in the United States and the National Institute for Educational Research (NIER) in Japan. Appendix 8 lists the consortium staff and consultants who have made significant contributions to the development and implementation of the project.

The consortium implements PISA 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. These expert groups were referred to as Subject Matter Expert Groups (SMEGs) (*see* Appendix 8 for members). By participating in these expert groups and regularly reviewing outcomes of the groups' meetings, countries ensured that the instruments were internationally valid and that they took into account the cultural and educational contexts of the different OECD member countries, 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 NPM 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 had overall responsibility for managing 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 consortium.

ORGANISATION 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 2006 which have been published in *PISA 2006: Science Competencies for Tomorrow's World* (OECD, 2007). A bibliography of other PISA related reports is included in Appendix 9.

There are five sections in this report:

- *Section One – Instrument design* (Chapters 1-4): Describes the design and development of both the questionnaires and achievement tests;
- *Section Two – Operations* (Chapters 5-7): 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;
- *Section Three – Data processing* (Chapters 8-10): 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;
- *Section Four – Quality indicators and outcomes* (Chapters 11-14): Covers the results of the scaling and weighting, reports 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 2006 data adjudication; that is, the overall analysis of data quality for each country;
- *Section Five – Scale construction and data products* (Chapters 15-18): Describes the construction of the PISA 2006 levels of proficiency and the construction and validation of questionnaire-related indices. The final chapter briefly describes the contents of the PISA 2006 database;
- *Appendices*: Detailed appendices of results pertaining to the chapters of the report are provided.

Notes

1. The PISA 2006 parent questionnaire was administered in Denmark, Germany, Iceland, Italy, Luxembourg, New Zealand, Poland, Portugal, Korea and Turkey, as well as in the partner countries/economies Bulgaria, Colombia, Croatia, Hong Kong-China, Macao-China and Qatar.



Reader's Guide

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

OECD countries

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

Partner countries and economies

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

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



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

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

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



2

Test design and test development

Test scope and format 28
Test design 28
Test development centres 29
Development timeline 30
The PISA 2006 scientific literacy framework 30
Test development – cognitive items 31
- Item development process 31
- National item submissions 33
- National review of items 34
- International item review 35
- Preparation of dual (English and French) source versions 35

Test development – attitudinal items 35
Field trial 38
- Field trial selection 38
- Field trial design 39
- Despatch of field trial instruments 40
- Field trial coder training 40
- Field trial coder queries 40
- Field trial outcomes 41
- National review of field trial items 42

Main study 42
- Main study science items 42
- Main study reading items 44
- Main study mathematics items 45
- Despatch of main study instruments 46
- Main study coder training 46
- Main study coder query service 46
- Review of main study item analyses 47

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



This chapter describes the test design for PISA 2006 and the processes by which the PISA consortium, led by ACER, developed the PISA 2006 paper-and-pencil test.

TEST SCOPE AND FORMAT

In PISA 2006 three subject domains were tested, with science as the major domain for the first time in a PISA administration and reading and mathematics as minor domains.

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 up to four items assessing students' scientific competencies and knowledge. In addition, for PISA 2006 about 60% of the science units contained one or two items designed to assess aspects of students' attitudes towards science. Throughout this chapter the terms "cognitive items" and "attitudinal items" will be used to distinguish these two separate types of items.

There were 37 science units, comprising a total of 108 cognitive items and 31 embedded attitudinal items, representing approximately 210 minutes of testing time for science in PISA 2006. The same amount of time was allocated to the major domain for 2003 (mathematics), but there were no attitudinal items in the 2003 test. The reading assessment consisted of the same 31 items (8 units) as in 2003, representing approximately 60 minutes of testing time, and the mathematics assessment consisted of 48 items (31 units), representing approximately 120 minutes of testing time. The mathematics items were selected from the 167 items used in 2003.

The 108 science cognitive items used in the main study included 22 items from the 2003 test. The remaining 86 items were selected from a pool of 222 newly-developed items that had been tested in a field trial conducted in all countries in 2005, one year prior to the main study. There was no new item development for reading and mathematics.

Item formats employed with science cognitive items were multiple-choice, short closed-constructed response, and open- (extended) constructed response. Multiple-choice items were either standard multiple-choice with four 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 several possible responses (yes/no, true/false, correct/incorrect, etc.). Closed-constructed response items required students to construct a numeric response within very limited constraints, or only required a word or short phrase as the answer. Open-constructed response items required more extensive writing and frequently required some explanation or justification.

Each attitudinal item required students to express their level of agreement on a four-point scale with two or three statements expressing either interest in science or support for science. Each attitudinal item was formatted distinctively and appeared in a shaded box – see Figure 2.1 and Figure 2.2.

Pencils, erasers, rulers, and in some cases calculators, were provided. 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 some students.

TEST DESIGN

The main study items were allocated to thirteen item clusters (seven science clusters, two reading clusters and four mathematics clusters) with each cluster representing 30 minutes of test time. The items were presented to students in thirteen test booklets, with each booklet being composed of four clusters according



to the rotation design shown in Table 2.1. S1 to S7 denote the science clusters, R1 and R2 denote the reading clusters, and M1 to M4 denote the mathematics clusters. R1 and R2 were the same two reading clusters as in 2003, but the mathematics clusters were not intact clusters from 2003. The eight science link units from 2003 were distributed across the seven science clusters, in first or second position.

The fully-linked design is a balanced incomplete block design. Each cluster appears in each of the four possible positions within a booklet once and so each test item appeared in four of the test booklets. Another feature of the design is that each pair of clusters appears in one (and only one) booklet.

Each sampled student was randomly assigned one of the thirteen booklets, which meant each student undertook two hours of testing. Students were allowed a short break after one hour. The directions to students emphasised that there were no correct answers to the attitudinal questions, and that they would not count in their test scores, but that it was important to answer them truthfully.

Table 2.1
Cluster rotation design used to form test booklets for PISA 2006

Booklet	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	S1	S2	S4	S7
2	S2	S3	M3	R1
3	S3	S4	M4	M1
4	S4	M3	S5	M2
5	S5	S6	S7	S3
6	S6	R2	R1	S4
7	S7	R1	M2	M4
8	M1	M2	S2	S6
9	M2	S1	S3	R2
10	M3	M4	S6	S1
11	M4	S5	R2	S2
12	R1	M1	S1	S5
13	R2	S7	M1	M3

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 exclusively to students with special needs. The UH booklet contained about half as many items as the other booklets, with about 50% of the items being science items, 25% reading and 25% mathematics. The items were selected from the main study items taking into account their suitability for students with special educational needs.

TEST DEVELOPMENT CENTRES

Experience gained in the two previous PISA assessments showed the importance of using diverse centres of test development expertise to help achieve conceptually rigorous material that has the highest possible levels of cross-cultural and cross-national diversity. Accordingly, to prepare new science items for PISA 2006 the consortium expanded its number of test development centres over the number used for PISA 2003. Test development teams were established in five culturally-diverse and well-known institutions, namely ACER (Australia), CITO (the Netherlands), ILS (University of Oslo, Norway), IPN (University of Kiel, Germany) and NIER (Japan) (see Appendix 9).

In addition, for PISA 2006 the test development teams were encouraged to do initial development of items, including cognitive laboratory activities, in their local language. Translation to the OECD source languages (English and French) took place only after items had reached a well-formed state. The work of the test development teams was coordinated and monitored overall at ACER by the consortium's manager of test and framework development for science.

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



DEVELOPMENT TIMELINE

The PISA 2006 project started formally in September 2003, and concluded in December 2007. Planning for item development began in July 2003, with preparation of material for a three-day meeting of test developers from each team, which was held in Oslo in September, 2003. The meeting had the following purposes:

- To become familiar with the draft PISA 2006 scientific literacy framework, especially its implications for test development;
- To discuss the requirements for item development, including item presentation and formats, use of templates and styles, cognitive laboratory procedures and timelines;
- To be briefed on detailed guidelines, based on experience from the first two PISA administrations, for avoiding potential translation and cultural problems when developing items;
- To review sample items prepared for the meeting by each of the test development teams;
- To prepare advice to the PISA 2006 Science Expert Group (SEG) on the adequacy of the draft science framework as a basis for item development.

Test development began in earnest after the first PISA 2006 SEG meeting which was held in Las Vegas in October 2003. The main phase of test development finished when the items were distributed for the field trial in December 2004. 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 2005, after which items were selected for the main study and distributed to countries in December 2005. Table 2.2 shows the major milestones and activities of the PISA 2006 test development timeline.

Table 2.2
Test development timeline for PISA 2006

Activity	Period
Initial framework development by OECD	December 2002 – June 2003
Framework development by ACER consortium	October 2003 – August 2004
Item development	July 2003 – October 2004
Item submission from countries	February – June 2004
Distribution of field trial material	November – December 2004
Translation into national languages	December 2004 – April 2005
Field trial coder training	February 2005
Field trial in participating countries	March – August 2005
Selection of items for main study	August – October 2005
Preparation of final source versions of all main study materials, in English and French	October – December 2005
Distribution of main study material	December 2005
Main study coder training	February 2006
Main study in participating countries	From March 2006

THE PISA 2006 SCIENTIFIC LITERACY FRAMEWORK

For each PISA subject domain, an assessment framework is produced to guide the PISA assessments in accordance with the policy requirements of the OECD's PISA Governing Board (PGB). 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.

In December 2002 the OECD invited national experts to a science forum as the first step in the preparation of a revised and expanded science framework for PISA 2006. The forum established a working group which met in January 2003 and prepared a draft framework for review at a second science forum held in February.



A further draft was then produced and considered by the PGB at its meeting in Mexico City in March. After the PGB meeting a Science Framework Expansion Committee was established to continue development of the framework until the PISA 2006 contract was let. This committee, like the forums and working group, was chaired by Rodger Bybee who would subsequently be appointed chair of the PISA 2006 Science Expert Group.

Many sections of the science framework presented to the consortium in October 2003 were well developed – especially those concerning the definition of the domain and its organisational aspects (in particular, the discussions of contexts, competencies and knowledge). Other sections, however, were not so well developed. Over the next 10 months, through its Science Expert Group and test developers, and in consultation with national centres and the science forum, the consortium further developed the framework and a final draft was submitted to the OECD in August 2004. In the latter part of 2005, following the field trial, some revisions were made to the framework and in early 2006 it was prepared for publication along with an extensive set of example items. All three PISA 2006 frameworks were published in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD, 2006). The reading and mathematics frameworks were unchanged from 2003.

TEST DEVELOPMENT – COGNITIVE ITEMS

The test development process commenced with preparations for the meeting of test developers held in Oslo in September 2003. This included the preparation of documentation to guide all parts of the process for the development of cognitive items. The process continued with the calling of submissions from participating countries, writing and reviewing items, carrying out 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 study.

Item development process

Cognitive item development was guided by a comprehensive set of guidelines prepared at the start of the project and approved by the first meeting of the PISA 2006 Science Expert Group. The guidelines included an overview of the development process and timelines, a specification of item requirements, including the importance of framework fit, and a discussion of issues affecting item difficulty. A number of sample items were also provided. These guidelines were expected to be followed by item developers at each of the five test development centres.

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. As in PISA 2000 and 2003, double-digit coding was used in some items to distinguish between cognitively distinct ways of achieving the same level of credit. In a double-digit code, such as "12", the first digit (1) indicates the score or level of response and the second digit (2) indicates the method or approach used by the student.

First phase of development

Typically, the following steps were taken in the first phase of the development of science cognitive items originating at a test development centre. The steps are described in a linear fashion, but in reality they were often negotiated in a cyclical fashion, with items often going through the various steps more than once.

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



Initial preparation

Test developers prepared units in the local language in a standard format, including stimulus, one or more items (questions), and a proposed coding guide for each item. Items were then subject 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). All items were subject to panelling and underwent local piloting. In addition, cognitive interviews were employed for a significant proportion of items,.

Local item panelling

Each unit first underwent extensive scrutiny at a meeting of members of the relevant 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. A combination of think-aloud methods, individual interviews and group interviews were used with students to ascertain the thought processes typically employed as students attempted the items.

Items were revised, often extensively, following their use with individuals and small groups of students. This stage was particularly useful in clarifying 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 item development, sets of units were piloted with several classes of 15-year-olds in schools in the country in which they were developed. 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, after being translated into English if their initial development had taken place in another language.

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 items or even 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 trialled 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 so 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 science cognitive items. A draft version of the guidelines was distributed to PISA 2003 NPMs in November 2003. The final version was distributed to PISA 2006 NPMs in February 2004.

The guidelines described the scope of the item development task for PISA 2006, the arrangements for national submissions of items and the item development timeline. In addition the guidelines contained a detailed discussion of item requirements and an overview of the full item development process for PISA 2006. Four complete sample units prepared at ACER were provided in an accompanying document.

The due date for national submission of items was 30 June 2004, as late as possible given field trial preparation deadlines. Items could be submitted in Chinese, Dutch, English, French, German, Italian, Japanese, Norwegian, Russian or Spanish, or any other language subject to prior consultation with the consortium. 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. 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.

A total of 155 units were processed from 21 countries, commencing in mid-March 2004. Countries submitting units were: Austria, Belgium, Canada, Chinese Taipei, the Czech Republic, Chile, Finland, France, Greece, Ireland, Italy, Korea, Mexico, Norway, New Zealand, Poland, Serbia, the Slovak Republic, Sweden, Switzerland and the United Kingdom. Most countries chose to submit their material in English, but submissions were received in five other languages (Dutch, French, German, Spanish and Swedish).

Some submitted units had already undergone significant development work, including pilot testing, prior to submission. Others were in a much less developed state and consisted in some cases of little more than a brief introduction and ideas for possible items. Often items were far too open-ended for use in PISA. Some countries established national committees to develop units and trialled their units with students. Other countries sub-contracted the development of units to an individual and submitted them without any internal review. The former approach yielded better quality units in general.



Each submitted unit was first reviewed by one of the test development centres to determine its general suitability for PISA 2006. Units initially deemed unsuitable were referred to another test development centre for a second and final opinion. About 25% of submitted units were deemed unsuitable for PISA 2006. The main reasons for assessing units as unsuitable were lack of context, inappropriate context, cultural bias, curriculum dependence, just school science and including content that was deemed to be too advanced.

The remaining 75% of submitted units were considered suitable in some form or other for use in PISA 2006. However, only a handful of these units were deemed not to need significant revision by consortium test developers. Various criteria were used to select those units to be further developed, including overall quality of the unit, amount of revision required and their framework coverage. Nevertheless, high importance was placed on including units from as wide a range of countries as possible. Some units were excluded because their content overlapped too much with existing units.

Units requiring further initial development were distributed among the test development centres. Typically, after local panelling and revision, they were fast-tracked into the second phase of item development as there was rarely time for cognitive interviews or pilot testing to be conducted locally. However, all these units underwent international pilot testing (as described above), along with the units that originated at test development centres and a handful of units that were developed from material supplied by individual members of the Science Expert Group.

A total of 40 units (150 items) arising from national submissions were included in the five bundles of items circulated to national centres for review. Feedback was provided to countries on their submitted units that were not used. This action, together with the provision of an item development workshop for national centre representatives early in a cycle, should improve the quality of national submissions in the future.

National review of items

In February 2004, NPMs were given a set of item review guidelines to assist them in reviewing cognitive items and providing feedback. A copy of a similar set of guidelines, prepared later for review of all items used in the field trial and including reference to attitudinal items was also available.

At the same time, NPMs were given a schedule for the distribution and review of bundles of draft items during the remainder of 2004. 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 science or technology. NPMs also were asked to identify any cultural concerns or other problems with the items, such as likely translation or marking difficulties, and to give each item an overall rating for retention in the item pool.

As items were developed to a sufficiently complete stage, they were despatched to national centres for review. Five bundles of items were distributed. The first bundle, including 65 cognitive items, was despatched in January 2004. National centres were provided with an Excel worksheet, already populated with unit names and item identification codes, in which to enter their ratings and other comments. Subsequent bundles were despatched in April (103 cognitive items), June (125 items), July (85 items) and August (114 items). In each case, about 4 weeks was scheduled for the submission of feedback.

For each bundle, a series of reports was generated summarising the feedback from NPMs. The feedback frequently resulted in further revision of 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 Science Expert Group (SEG) that took place in October 2003 and February, July and September 2004.

The July 2004 SEG meeting, held in Warsaw, was immediately preceded by a science forum, and all items that had been developed at that stage were reviewed in detail by forum participants. All PISA countries were invited to send national science experts to the forum. The forum also provided advice on issues that had arisen during framework and student questionnaire development.

Preparation of dual (English and French) source versions

Both English and French source versions of all 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.

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.

TEST DEVELOPMENT – ATTITUDINAL ITEMS

The assessment of the affective domain was a major focus of the first meeting of the PISA 2006 Science Expert Group held in October 2003. It was recommended that the assessment be restricted to three attitude scales, rather than the five scales proposed by the Science Framework Expansion Committee:

- Interest in science;
- Value placed on scientific enquiry – eventually renamed Support for scientific enquiry; and
- Responsibility towards resources and environments.

For convenience, the names of the scales will often be shortened to *interest*, *support* and *responsibility* in the remainder of this chapter.

At the first meeting of PISA 2006 test developers, held in Oslo in September 2003, staff from the IPN test development centre proposed that suitable units should contain items requiring students to indicate their level of agreement with three statements. This proposal was then put to the October SEG meeting which gave its support for future development of such Likert-style attitudinal items. Two examples from the released main study unit ACID RAIN are shown in Figure 2.1 and Figure 2.2. Like the interest item, the support item originally contained three parts, but one was dropped because it exhibited poor psychometric properties in the field trial.



Figure 2.1
Main study "Interest in Science" item

ACID RAIN – *QUESTION 10N (S485Q10N)*

How much interest do you have in the following information?

Tick only one box in each row.

	High Interest	Medium Interest	Low Interest	No Interest
d) Knowing which human activities contribute most to acid rain.	☐ 1	☐ 2	☐ 3	☐ 4
e) Learning about technologies that minimise the emission of gases that cause acid rain.	☐ 1	☐ 2	☐ 3	☐ 4
f) Understanding the methods used to repair buildings damaged by acid rain.	☐ 1	☐ 2	☐ 3	☐ 4

Figure 2.2
Main study "Support for Scientific Enquiry" item

ACID RAIN – *QUESTION 10S (S485Q10S)*

How much do you agree with the following statements?

Tick only one box in each row.

	Strongly agree	Agree	Disagree	Strongly disagree
g) Preservation of ancient ruins should be based on scientific evidence concerning the causes of damage.	☐ 1	☐ 2	☐ 3	☐ 4
h) Statements about the causes of acid rain should be based on scientific research.	☐ 1	☐ 2	☐ 3	☐ 4

A unipolar response format, rather than a conventional bipolar format, was used with interest items to reduce the influence of social desirability on student responses. It was recognised that there was an even greater risk of this occurring with support items but no satisfactory alternative could be found that would work in all PISA languages with the great variety of statement types employed. A four-point response scale was used with all Likert-style attitudinal items because it does not allow students to opt for a neutral response.

At the second meeting of the SEG, held in Athens in February 2004, test developers proposed a second type of attitudinal item illustrated in Figure 2.3. In this item-type, four ordered opinions about an issue, representing different levels of commitment to a sustainable environment, are given, and students have to choose the one that best matches their opinion. Like all attitudinal items, items of this type were placed at the end of the unit so that students were familiar with the context prior to being asked their opinion. Originally, the responses in match-the-opinion items were listed in random order, but this was changed to counter criticism that the items were too cognitive in nature.

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



Figure 2.3
Field trial "Match-the-opinion" Responsibility Item

ACID RAIN – *QUESTION 10M (S485Q10M)*

The burning of fossil fuels (coal, oil and gas) contributes to the amount of acid rain. Four opinions about this issue are given below.

Circle the letter beside the one response that is most like your own opinion. There is no "correct" or "incorrect" response.

A. I think acid rain is not enough of a problem to change our use of fossil fuels.

B. Action to achieve lower acid rain levels would be good, but not if this affects the lifestyle I enjoy.

C. To help reduce acid rain, I would reduce my dependence on energy produced from fossil fuels if everyone else did too.

D. I will lower my use of energy produced from fossil fuels so as to help reduce acid rain.

Likert-style items are very efficient in that they minimise demands on student response time. However, concern is sometimes expressed about possible cultural variation in responses to the graded adjectives used, and it has been suggested that batteries of Likert-style items may lead to patterns in the way that students respond. It was felt that match-the-opinion items would avoid these potential drawbacks with the options corresponding to points spread along an underlying scale. However, for several reasons – primarily their experimental nature and the cost of their development – it was decided to restrict development of match-the-opinion items to the *responsibility for resources and environments* scale.

Several changes to the three scale definitions took place in the first half of 2004. A pilot study conducted by IPN with embedded Likert-style items early in the year distinguished two scales within the original responsibility scale definition – personal responsibility and shared responsibility. The SEG meeting in Athens decided that the scale should focus on personal responsibility, so references to shared responsibility were removed from the definition. Another outcome of this pilot and two further pilots conducted in June was that the focus of the interest scale was gradually narrowed to *interest in learning about science*, as statements addressing broader aspects of interest in science tended not to scale on the same dimension.

Finally, it became apparent that the scope of the responsibility scale had to be broadened if possible as not enough units had contexts that made them suitable for items addressing *responsibility for resources and environments*. The SEG meeting held in Warsaw in July thus recommended expansion of the scale definition to include personal responsibility for achieving a healthy population, and rename it *responsibility for sustainable development*, subject to confirmation from the field trial that the items formed a single scale.

In June 2004 the PGB determined that 17% of science testing time in the field trial should be devoted to attitudinal items. This weighting, endorsed by the science forum held soon after in July, was considerably higher than had been recommended by the consortium and meant that development of attitudinal items had to be accelerated significantly.

Development of Likert-style items was shared by the ACER and IPN test development centres. On average, two such items, each comprising three statement parts, were developed for each of the 113 units that were



circulated to national centres for review. Interest items were developed for all but three units, support items for two-thirds of the units and responsibility items for 40% of them. In addition, match-the-opinion responsibility items were developed for 25 units at ACER. More items were produced for the interest scale than for the other two scales because feedback from pilot studies and NPM meetings indicated that it was the most likely scale to survive through to the main study.

All the items were subject to at least two rounds of panelling but time constraints meant that only about one-third were piloted with classes of students. The items included in units selected for the field trial were panelled again before being distributed to NPMs for review and, at the same time, submitted for translation into French and for professional editing. Feedback from these processes resulted in most items being revised and many items being discarded. In particular, feedback from the French expert identified many potential translation issues, especially with the support statements as the expression for the word support in French, and presumably some other languages, does not refer to an opinion only, but to taking some action as well.

FIELD TRIAL

A total of 113 science units (492 cognitive items) were circulated to national centres for review. After consideration of country feedback, 103 units (377 cognitive items) were retained as the pool of units considered by the SEG for inclusion in the field trial. Thirty-eight of these units (37%) originated in national submissions.

All units retained to this stage were subjected to an editorial check using the services of a professional editor. This helped uncover any remaining typographical errors, grammatical inconsistencies and layout irregularities, and provided a final check that the reading level of the material was appropriate for 15-year-olds.

Field trial selection

The new cognitive items to be used in the 2005 field trial were selected from the item pool at the meeting of the SEG held in Bratislava in mid-September 2004. The selection process took two-and-a-half days to complete. Each SEG member first chose ten units to be included in the field trial, with 67 of the 103 units receiving at least one vote. The SEG then reviewed these units in detail, item-by-item. This resulted in 14 units being omitted from the initial selection and some items being omitted from the remaining units. Next, all the units not included in the selection were reviewed item-by-item, resulting in a further 28 units being added to the selection. Throughout this process, SEG members made numerous suggestions of ways to improve the final wording of items.

At this stage, 81 units remained in the item pool, about 25% more items than required. The characteristics of the items, including framework classifications and estimated difficulties, were then examined and a final selection of 62 new units (237 cognitive items) was made to match framework requirements as far as possible. The ratings assigned to items by countries were taken into account at each step of the selection process, and at no time were SEG members informed of the origin of any item. The SEG selection was presented to a meeting of National Project Managers in the week after the SEG meeting, together with nine units (25 items) from 2003 that had been kept secure for use as link material.

Subsequently, one new unit was dropped from the item pool as a result of NPM concerns about the appropriateness of its context in some cultures, and another unit was replaced because of lingering doubts about the veracity of the science content. In addition, a small number of items had to be dropped because of space and layout constraints when the consortium test developers assembled the units into clusters and booklets. The final field trial item pool included a total of 247 science cognitive items, comprising 25 link items and 222 new items. These figures have been adjusted for the late substitution of one unit (DANGEROUS WAVES) for sensitivity reasons following the South-East Asian tsunami in December 2004.



Included in the pool were several units specifically designed to target students' major misconceptions about fundamental scientific concepts.

Attitudinal items for all nine link units and all but one of the new units in the field trial selection were circulated to national centres for review, a total of 144 items. After consideration of country and French expert feedback, 124 items remained and 105 of these were included in the final pool. Sixty of the 70 science field trial units contained at least one attitudinal item and 37 contained more than one attitudinal item. More information about the distribution of the attitudinal items is given in Table 2.3.

Table 2.3
Science field trial all items

	Attitudinal items				Total attitudinal items	Cognitive items	Grand total
	Interest	Support	Responsibility	Match-the-opinion			
Link items	6	3	3	0	12	25	37
New items	38	23	20	12	93	222	315
Total items	**44**	**26**	**23**	**12**	**105**	**247**	**352**

Field trial design

The 70 new science units were initially allocated to 18 clusters, S1 to S18. Next, two versions of six of the clusters were formed, differing in the attitudinal items that they contained. Clusters S1, S3, S11 and S13 contained only Likert-style attitudinal items that were replaced in clusters S1M, S3M, S11M and S13M by match-the-opinion items developed for the same units. This enabled the performance of the two types of attitudinal items to be compared.

Clusters S16A and S17A comprised the nine 2003 link units, including their newly prepared (Likert-style) attitudinal items, whereas the attitudinal items were replaced in clusters S16 and S17 by an extra unit of cognitive items of equivalent time demand. This enabled an investigation of any effect that embedding attitudinal items in a unit might have on students' performance on cognitive items.

The field trial design was correspondingly complicated and is shown in Table 2.4. Each cluster was designed to take up 30 minutes of testing time and appeared at least once in the first half of a booklet and at least once in the second half. Booklets 1 to 4 were identical to booklets 11 to 14 except for the types of attitudinal items they contained.

Table 2.4
Allocation of item clusters to test booklets for field trial

Booklet	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	S1	S11	S10	S18
2	S3	S13	S12	S2
3	S2	S12	S11	S1
4	S4	S14	S13	S3
5	S5	S15	S14	S4
6	S6	S16	S15	S5
7	S7	S17	S16	S6
8	S8	S16A	S17	S7
9	S9	S17A	S16A	S8
10	S18	S10	S17A	S9
11	S1M	S11M	S10	S18
12	S3M	S13M	S12	S2
13	S2	S12	S11M	S1M
14	S4	S14	S13M	S3M
15	M1	M2	R2	R1

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



R1 and R2 were the same two reading clusters as in the PISA 2003 main study, comprising a total of 31 items (8 units), although the units in R2 were ordered differently than in 2003. M1 and M2 were two mathematics clusters formed from 2003 main study items, comprising a total of 26 items (17 units). The reading and mathematics clusters only appeared in booklet 15. Countries that participated in PISA 2003 did not have to do this booklet. Half of these countries were assigned booklets 1 to 12 and the other half were assigned booklets 3 to 14. All countries new to PISA did booklet 15 and in addition were assigned either booklets 1 to 12 or 3 to 14.

Despatch of field trial instruments

Final English and French source versions of field trial units were distributed to national centres in two despatches, in October (link units) and November (new science units). Clusters and booklets were distributed in December 2004 in both Word and PDF formats. All material could also be downloaded from the PISA website from the time of despatch.

National centres then 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 despatch of items to be included in the field trial, various documents and materials were prepared to assist in the training of response coders. International coder training sessions for science, reading and mathematics were scheduled for February 2005. 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 separated different kinds of possible responses, which did not all necessarily receive different scores. A separate training workshop document was also produced for each subject area, once again in both English and French. These 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.

Countries sent representatives to the training sessions, which were conducted in Marbella, Spain. 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 despatch of coding guides and training materials two weeks later. Following the international training sessions, national centres conducted their own coder training activities using their verified translations of the consolidated coding guides.

Field trial coder queries

The consortium provided a coder query service to support the coding of scripts in each country. When there was any uncertainty, national centres were able to submit queries by e-mail to the query service, and they 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 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

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



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.

Field trial outcomes

Extensive analyses were conducted on the field trial cognitive item response data. These analyses have been reported elsewhere, but included the standard *ConQuest*® item analysis (item fit, item discrimination, item difficulty, distracter 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. On the basis of these critical measurement statistics, about 40 new items were removed from the pool of items that would be considered for the main study. The omissions included many of the items focussing on misconceptions and a few complete units. The statements in each complex multiple-choice item were also analysed separately and this led to some statements being dropped though the item itself was retained. Analyses also indicated that one of the nine PISA 2003 units should not be included in the main study.

Analyses of the responses to the attitudinal items, also reported elsewhere, showed that the presence of embedded attitudinal items in the main study test would have little, if any, effect on test performance. Each statement-part of an attitudinal item was considered a separate partial-credit item in these analyses. The analyses showed that the sets of interest and support items formed single scales, as did the match-the-opinion *responsibility for resources and environments* items. All but one of the 12 match-the-opinion items had sound psychometric properties.

Unfortunately, the analyses showed that Likert-style items designed to measure *responsibility for sustainable development* did not always load on one dimension and so could not be recommended for inclusion in the main study. Some of these items tended to load on the same dimension as items designed to measure support. Others loaded on a dimension representing concern for personal health and safety, together with some interest items that were consequently not considered for inclusion in the main study.

In accordance with the findings about responsibility items, the framework was revised following the field trial by removing reference to personal responsibility for achieving a healthy population from the responsibility scale definition and reinstating its original name, *responsibility for resources and environments*.

Timing study

A timing study was conducted to gather data on the average time taken to respond to items in the field trial, and the results were used to estimate the number of items that should be included in main study clusters. The timing information from clusters S16, S16A, S17 and S17A was used to estimate average time for embedded Likert-style attitudinal items. The estimated average time to complete a Likert-style attitudinal item was 0.75 minutes. The timing information from clusters S1 and S1M was used to estimate average time for embedded match-the-opinion attitudinal items. The estimated average time to complete a match-the-opinion item was 1.25 minutes.

Only the time taken to complete the first block (cluster) in booklets 1 to 14 was used to estimate average time for science cognitive items. Previous PISA timing studies have shown that there are far more missing responses as well as more guessing in the latter part of a test than in the earlier part. The estimated average time to complete each cognitive item in the first block of the test was 1.68 minutes.

It was concluded that main study science clusters should contain 17 cognitive items, less an allowance for embedded attitudinal items given approximately by the following formulas: about two Likert-style items (each containing 2-3 statements) per one cognitive item and five match-the-opinion items per four cognitive items.



National review of field trial items

A further round of national item review was carried out, this time informed by the experience at national centres of how the items worked in the field trial in each country. A set of review guidelines, was designed 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, both cognitive and attitudinal, to indicate its priority for inclusion in the main study. Almost all countries completed this review of all field trial items.

A comprehensive field trial review report also was prepared by all National Project Managers. 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 STUDY

A science attitudes forum was held in Warsaw on 30–31 August 2005. Its main purpose was to consider the results of the field trial analyses and hence provide advice on whether attitudinal items should be embedded in science units in the main study. About 75% of national experts were in favour of including interest items and about 25% were in favour of embedding support items as well. Consortium and SEG advice to the PGB was that match-the-opinion items to assess Responsibility towards resources and environments also should be included provided that this did not adversely affect the selection of cognitive items.

Main study science items

The Science Expert Group met in October 2005 in Melbourne to review all available material and recommend which science items should be included in the main study. Before the selection process began, advice was received from the concurrent PGB meeting in Reykjavik about the inclusion of embedded attitudinal items. The PGB advised that only embedded interest (*interest in [learning about] science*) and support (*support for scientific enquiry*) items should be used. The experimental nature of match-the-opinion items and the small number available acted against their inclusion.

Based on the results of the field trial timing study, and making allowance for the inclusion of embedded interest and support items, it was estimated that the main study selection needed to contain about 105 science cognitive items. This meant that about 83 new items had to be selected, as there were 22 items in the eight remaining units available for linking purposes with PISA 2003.

As a first step in the selection process, each SEG member nominated eight new units that they thought should be included in the selection because of their relevance to the assessment of scientific literacy. In refining its selection, the SEG took into account all available information, including the field trial data, national reviews and ratings, information from the translation process, information from the national field trial reviews and the requirements of the framework. Attitudinal items were ignored until the final step of the process, when it was confirmed that the selected units contained sufficient interest and support items to enable robust scales to be constructed.

The selection had to satisfy the following conditions:

- The psychometric properties of all selected items had to be satisfactory;
- Items that generated coding problems 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 had to be preferred, and items with lower ratings had to be avoided.



In addition, the combined set of new and link items had to satisfy these additional conditions as far as possible:

- The major framework categories (competencies and knowledge) had to be populated as specified in the scientific literacy framework;
- There had to be an appropriate distribution of item difficulties;
- The proportion of items that required manual coding could not exceed 40%.

The final SEG selection contained 30 new units (92 cognitive items). This was slightly more items than needed and subsequently six of the items, including one complete unit, were dropped, while retaining the required balance of framework categories. The selection contained a few misconception items with less-than-desirable psychometric properties because of the importance that the SEG placed on their inclusion.

The average NPM priority rating of selected items was 3.91 and the average rating for the remaining items was 3.69. Eleven of the 29 units in the final selection originated from the national submissions of eight countries. Overall, the 29 units were developed in 12 countries in eight different languages, with eight units being originally developed in English.

Nine of the 29 new units included both interest and support items, nine included an interest item only, five included a support item only and the remaining six units had no embedded attitudinal item. Link units were retained exactly as they appeared in 2003, without embedded attitudinal items, so that the complete main study science item pool contained 37 units (eight link units and 29 new units), comprising 108 cognitive items and 32 attitudinal items (18 interest items and 14 support items).

The SEG identified 18 units not included in the field trial that would be suitable for release as sample PISA science units once minor revisions were incorporated. Sixteen of these units, comprising a total of 62 items, were included as an annex to *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD, 2006). The other two units were retained for possible use in a future PISA survey.

The main study item pool was presented to a meeting of National Project Managers in Mildura, Australia in October 2005. Distributions of the science items, with respect to the major framework categories, are summarised in Table 2.5, Table 2.6 and Table 2.7.

Note that the scientific competency and knowledge dimensions as defined in the framework do not give rise to independent item classifications. In particular, by virtue of its definition, items classified as assessing the competency *explaining scientific phenomena* would also be classified as *knowledge of science* items.

Table 2.5
Science main study items (item format by competency)

	Scientific Competency			
Item format	**Identifying scientific issues**	**Explaining scientific phenomena**	**Using scientific evidence**	**Total**
Multiple-choice	9	22	7	38 (35%)
Complex multiple-choice	10	11	8	29 (27%)
Closed-constructed response	0	4	1	5 (5%)
Open-constructed response	5	16	15	36 (33%)
Total	**24 (22%)**	**53 (49%)**	**31 (29%)**	**108**

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



Table 2.6
Science main study items (item format by knowledge type)

Item format	Knowledge of science	Knowledge about science	Total
Multiple-choice	24	14	38 (35%)
Complex multiple-choice	15	14	29 (27%)
Closed-constructed response	4	1	5 (5%)
Open-constructed response	19	17	36 (33%)
Total	**62 (57%)**	**46 (43%)**	**108**

Table 2.7
Science main study items (knowledge category by competency)

	Scientific Competency			
Item scale	Identifying scientific issues	Explaining scientific phenomena	Using scientific evidence	Total
Physical systems		15	2	17 (13%)
Living systems		24	1	25 (23%)
Earth & space systems		12	0	12 (11%)
Technology systems		2	6	8 (7%)
Scientific enquiry	24		1	25 (23%)
Scientific explanations	0		21	21 (19%)
Total	**24 (22%)**	**53 (49%)**	**31 (29%)**	**108**

This can be seen in Table 2.7, which also shows that all items classified as assessing the competency *identifying scientific issues* are classified as *knowledge about science* items. This latter characteristic is due to a decision taken during test development to minimise the *knowledge of science* content in such items so that the *identifying scientific issues* and *explaining scientific phenomena* scales were kept as independent as possible. This was thought important given the PGB and SEG preference to use competency scales for the reporting of science achievement in PISA 2006.

It follows from the classification dependencies that the relative weighting of the two knowledge components in the item set will also largely determine the relative weightings of the three competencies. The percentage of score points to be assigned to the *knowledge of science* component of the assessment was determined by the PGB prior to the field trial, in June 2004, to be about 60%. This decision had a far reaching consequence in terms of the overall gender differences in the PISA 2006 science outcomes as males generally outperformed females on *knowledge of science* items and girls generally outperformed boys for *knowledge about science* items.

Main study reading items

The two PISA 2003 clusters containing a total of eight units (31 items) were used again. Unlike in the field trial, the order of the items was the same as in 2003. Distributions of the reading items, with respect to the major framework categories, are summarised in Table 2.8, Table 2.9 and Table 2.10.

Table 2.8
Reading main study items (item format by aspect)

	Process (Aspect)			
Item format	Retrieving information	Interpreting texts	Reflection and evaluation	Total
Multiple-choice	0	9	0	9 (29%)
Complex multiple-choice	1	0	0	1 (3%)
Closed-constructed response	6	1	0	7 (23%)
Open-constructed response	3	4	7	14 (45%)
Total	**10 (32%)**	**14 (45%)**	**7 (23%)**	**31**

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



Table 2.9
Reading main study items (item format by text format)

Item format	Continuous texts	Non-continuous texts	Total
Multiple-choice	8	1	9 (29%)
Complex multiple-choice	1	0	1 (3%)
Closed-constructed response	0	7	7 (23%)
Open-constructed response	9	5	14 (45%)
Total	**18 (58%)**	**13 (42%)**	**31**

Table 2.10
Reading main study items (text type by aspect)

	Process (Aspect)			
Text type	Retrieving information	Interpreting texts	Reflection and evaluation	Total
Narrative	0	1	2	3 (10%)
Expository	0	9	3	12 (39%)
Descriptive	1	1	1	3 (10%)
Charts and graphs	1	1	0	2(6%)
Tables	3	1	0	4 (13%)
Maps	1	0	0	1 (3%)
Forms	4	1	1	6 (19%)
Total	**10 (32%)**	**14 (45%)**	**7 (23%)**	**31**

Main study mathematics items

Four clusters containing a total of 31 units (48 items) were selected from the PISA 2003 main study when mathematics had been the major assessment domain. Initially, it was expected that mathematics and reading would each contribute three clusters to the PISA 2006 main study item pool. However when the Reading Expert Group formed its recommendation to retain the two intact reading clusters from 2003, this created the opportunity for mathematics to contribute an additional cluster to fill the gap. Sufficient suitable material from the 2003 main survey that had not been released was available, so four clusters were formed. This selection of items occurred after decisions had been taken regarding the quite substantial number of items for public release from PISA 2003. This had two consequences: first, it was not possible to retain intact clusters from the PISA 2003 assessment, as some items in each cluster had already been released; and second, the number of items required to fill the available space was virtually equal to the number of items available, and therefore the balance across framework characteristics was not as optimal as it might have been.

Distributions of the mathematics items, with respect to the major framework categories, are summarised in Table 2.11, Table 2.12 and Table 2.13.

Table 2.11
Mathematics main study items (item format by competency cluster)

	Competency Cluster			
Item format	Reproduction	Connections	Reflection	Total
Multiple-choice	5	3	4	12 (25%)
Complex multiple-choice	0	7	2	9 (19%)
Closed-constructed response	2	2	2	6 (13%)
Open-constructed response	4	12	5	21 (44%)
Total	**11 (23%)**	**24 (50%)**	**13 (27%)**	**48**

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



Table 2.12
Mathematics main study items (item format by content category)

Item format	Space and shape	Quantity	Change and relationships	Uncertainty	Total
Multiple-choice	3	3	1	5	12 (25%)
Complex multiple-choice	2	2	2	3	9 (19%)
Closed-constructed response	2	2	2	0	6 (13%)
Open-constructed response	4	6	8	3	21 (44%)
Total	**11 (23%)**	**13 (27%)**	**13 (27%)**	**11 (23%)**	**48**

Table 2.13
Mathematics main study items (content category by competency cluster)

	Competency Cluster			
Content category	Reproduction	Connections	Reflection	Total
Space and shape	2	7	2	11 (23%)
Quantity	4	7	2	13 (27%)
Change and relationships	3	5	5	13 (27%)
Uncertainty	2	5	4	11 (23%)
Total	**11 (23%)**	**24 (50%)**	**13 (27%)**	**48**

Despatch of main study instruments

After finalising the main study 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 addition of further sample student responses to the coding guides. French translations of all selected items were then updated. Clusters of items were formed as described previously, and booklets were formed in accordance with the main study rotation design, shown previously in Table 2.1. Clusters and booklets were prepared in both English and French.

English and French versions of all items, item clusters and test booklets were made available to national centres in three despatches, in August (link units), November (new science units) and December 2005 (clusters and booklets).

Main study coder training

International coder training sessions for science, reading and mathematics were held in February 2006. Consolidated coding guides were prepared, in both English and French, containing all the items that required manual coding. These were despatched to national centres on 30 January 2006. 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.

Coder training sessions were conducted in Arrecife in the Canary Islands, Spain in February 2006. All but three countries had representatives at the training meetings. Arrangements were put in place to ensure appropriate training of representatives from those countries not in attendance. As 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 were prepared and despatched early in March.

Main study coder query service

The coder query service operated for the main study across the three 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



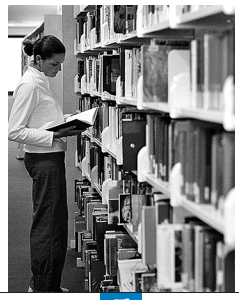
advice. The consortium was thereby able to provide consistent coding advice across countries. Reports of queries and the consortium responses were made available to all national centres via the consortium web site, and were regularly updated as new queries were received.

Review of main study item analyses

On receipt of data from the main study testing, extensive analysis of item responses were carried out to identify any items that were not capable of generating useful student achievement data. Such items were removed from the international dataset, or in some cases from particular national datasets where an isolated problem occurred. Two science items were removed from the international data set. In addition, three other items that focussed on misconceptions were retained in the database, although they did not form part of the scale.[1]

Note

1. The variables are: *S421Q02*, *S456Q01* and *S456Q02*.



3

The development of the PISA context questionnaires

Overview 50
The conceptual structure 51
- A conceptual framework for PISA 2006 51

Research areas in PISA 2006 55
The development of the context questionnaires 57
The coverage of the questionnaire material 58
- Student questionnaire 58
- School questionnaire 59
- International options 59
- National questionnaire material 60

The implementation of the context questionnaires 60

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



OVERVIEW

In its Call for Tender for PISA 2006, the PISA Governing Board (PGB) established the main policy issues it sought to address in the third cycle of PISA. In particular, the PGB required PISA 2006 to collect a set of basic demographic data as a core component that replicated key questions from the previous cycles. In addition, PISA 2006 needed to address issues related to important aspects of students' attitudes regarding science, information about students' experience with science in and out of school, motivation for, interest in and concern about science, and engagement with science-related activities.

Since the impact of out-of-school factors was considered of particular interest in a cycle where science was the major domain, the PGB recommended the inclusion of a parent questionnaire as an optional instrument, in order to collect additional information on issues such as science-related parental expectations and attitudes, as well as possible family investment in activities aimed at developing students' interest and learning in scientific areas.

The PISA 2006 Project consortium undertook the operationalisation of these goals with the assistance of a variety of experts. In particular, a Questionnaire Expert Group (QEG) was established, consisting of experts from a variety of research backgrounds and countries (see Appendix 8). The consortium and the QEG worked together to develop the contextual framework for PISA 2006 and the contextual instruments. Other experts were consulted where appropriate, especially some members of the Science Expert Group.

An initial step was the development of an organising conceptual structure which allowed the mapping of the PGB's priority policy issues to the design of PISA 2006. One important objective of the conceptual structure was to facilitate the development and choice of research areas that combine policy relevance effectively with the strengths of the PISA design. To aid this, a set of criteria established by the INES (International Indicators of Educational Systems) Network A were used:

- First, the research area must be of enduring policy relevance and interest. That is, a research area should have policy relevance, capture policy-makers' attention, address their needs for data about the performance of their educational systems, be timely, and focus on what improves or explains the outcomes of education. Further, a research area should be of interest to the public, since it is this public to which educators and policy-makers are accountable;
- Second, research areas must provide an internationally comparative perspective and promise significant added value to what can be accomplished through national evaluation and analysis. This implies that research areas need to be both relevant (*i.e.* of importance) and valid (*i.e.* of similar meaning) across countries;
- Third, there must be some consistency in the approach of each research area with PISA 2000 and PISA 2003;
- Fourth, it must be technically feasible and appropriate to address the issues within the context of the PISA design. That is, the collection of data about a subject must be technically feasible in terms of methodological rigour and the time and costs (including opportunity costs) associated with data collection.

The resulting research areas are listed below and described in more detail later in the chapter:

- Student's engagement in science
- Science attainment and the labour market
- Teaching and learning science
- Scientific literacy and environment
- Organisation of educational systems



THE CONCEPTUAL STRUCTURE

A conceptual framework for PISA 2006

Both the basic criteria for developing a conceptual framework presented above, and more comprehensive reviews of educational models (Scheerens and Bosker, 1997) reveal the complexity of variables and relationships that potentially influence student outcomes. The field is at the crossroads between a number of sociological, psychological, and cognitive theories, which all contribute important components to the overall picture.

Developing a new single, encompassing educational model for PISA would add little value to the many models already available in the literature. Rather than imposing unnecessary theoretical constraints on the thematic analyses that will be conducted using the study database, the primary role of the PISA conceptual structure for questionnaire development was to map the many components of existing models, to ensure that none of the essential dimensions are omitted from the data collection. These components were then checked against the general framework used for the OECD education indicators (INES) and the PGB priorities for PISA 2006.

This mapping also facilitated discussions around the feasibility and appropriateness of implementation within the constraints of the PISA design. In particular, the following aspects were considered, both in terms of restrictions and of potentialities related to the study design:

- PISA measures knowledge and skills for life and so does not have a strong curricular focus. This limits the extent to which the study is able to explore relationships between differences in achievement and differences in the implemented curricula. On the other hand, consideration was given to the out-of-school factors with a potential of enhancing cognitive and affective learning outcomes;
- PISA students are randomly sampled within schools, not from intact classrooms or courses and therefore come from different learning environments with different teachers and, possibly, different levels of instruction. Consequently, classroom-level information could only be collected either at the individual student level or at the school level;
- PISA uses an age-based definition of the target population. This is particularly appropriate for a yield-oriented study, and provides a basis for in-depth exploration of important policy issues, such as the effects of a number of structural characteristics of educational systems (*e.g.* the use of comprehensive vs. tracked study programmes, or the use of grade repetition). On the other hand, the inclusion in the study of an increasing number of non-OECD countries (where the enrolment rate for the 15-year-olds age group is maybe less than 100%) requires that retention be taken into account in the analysis of between-countries differences;
- The cross-sectional design in PISA does not allow any direct analysis of school effects over time. However, the cyclic nature of the study will permit not only the investigation of change in the criterion measures, but also in the effects of rates of change in the predictor variables.

Many conceptual models to explain learning outcomes distinguish different levels that relate both to the entities from which data might be collected and to the multi-level structure of national education systems (Scheerens 1990). Four levels can be distinguished:

- The education system as a whole (setting the context for teaching and learning);
- The educational institutions (schools but also other providers of education);
- The instructional setting and the learning environment within the institutions (classrooms, courses);
- The individual participants in learning activities (students).

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



A second dimension commonly found in many conceptual models groups the indicators at each of the above levels further into the following categories:

- Antecedents are those factors that affect policies and the way instruction is organised, delivered and received. It should be noted that they are usually specific for a given level of the education system and that antecedents at a lower level of the system may well be policy levers at a higher level (*e.g.* for teachers and students in a school, teacher qualifications are a given constraint while, at the level of the education system professional development of teachers is a key policy lever);
- Processes group information on the policy levers or circumstances that shape the outputs and outcomes at each level;
- Indicators on observed outcomes of education systems, as well as indicators related to the impact of knowledge and skills for individuals, societies and economies, are grouped under outcomes.

The four levels and the three aspects can be visualised as a two-dimensional grid with 12 potential variable types (Figure 3.1)). This basic conceptualisation has been adapted from the conceptual framework for the Second IEA Study of Mathematics (Travers and Westbury, 1989; Travers, Garden and Rosier, 1989) and also provided a conceptual basis for the planning of context questionnaires for the first two PISA surveys (Harvey-Beavis, 2002; OECD, 2005). As noted earlier, data on the instructional settings can only be collected at the individual or institutional level. However, conceptually they are still related to the level of the instructional settings (classroom, courses).

Figure 3.1 shows the basic components of this two-dimensional grid. It consists of four levels and variables at each level are classified as antecedents, processes or outcomes:

- At the system-level, the macroeconomic, social, cultural and political context sets constraints for the educational policies in a country. Outcomes at the system-level are not only aggregated learning outcomes but also equity-related outcomes;
- At the level of the educational institution, characteristics of the educational provider and its community context are antecedents for the policies and practices at the institutional level as well as the school climate for learning. Outcomes at this level are aggregates of individual learning outcomes and also differences in learning outcomes between sub-groups of students;

Figure 3.1
Conceptual grid of variable types

Antecedents	Processes	Outcomes
Level of the educational system		
Macro-economic, social, cultural and political context	Policies and organisation of education	Outcomes at the system level
Characteristics of educational institutions	Institutional policies and practice	Outcomes at the institutional level
Level of instructional units		
Characteristics of instructional units	Learning environment	Outcomes at the level of instructional units
Level of individual learners		
Student background and characteristics	Learning at the individual level	Individual learning outcomes

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



- At the level of the instructional units, characteristics of teachers and the classrooms/courses are antecedents for the instructional settings and the learning environment; learning outcomes are aggregated individual outcomes;
- At the student level, characteristics (like gender, age, grade) and background (like social status, parental involvement, language spoken at home) are antecedents for the individual learning process and learning outcomes (both cognitive and affective).

It should be noted that learning outcome variables consist not only of cognitive achievement but also of other potential learning outcomes. These include self-related cognitions (self-concept, self-efficacy), long-term interest in a subject or domain, educational expectations and aspirations as well as social outcomes like well-being and life skills.

While this mapping is useful for planning the coverage of the PISA questionnaires it is also important to supplement it with recognition of the dynamic elements of the educational system. System-level variables are important when interpreting relationships between variables at the lower levels and contradictory findings across countries are often due to differences in the structure of the educational systems.

From the existing conceptual frameworks and subsequent research one can derive hypotheses about (at least some of) the relationships between the elements in this two-dimensional grid. Typically, existing conceptual models assume antecedents to influence processes, which in turn produce learning outcomes, and conditions on higher levels are usually supposed to impact on those at lower levels (Scheerens, 1990).

Some models (Walberg 1984 and 1986; Creemers 1994) also expect that outcome variables have an effect on the learning process and, thus, allow for a non-recursive relationship between learning process and learning outcomes. Positive or negative experiences with subject-matter learning can influence process variables such as habits and attitudes towards the learning of a subject, increase or decrease the amount of time spent on homework, and so on. Another example is long-term interest in a subject or domain, which can be the outcome of learning but also affects the students' commitment to learning.

It also needs to be recognised that vertical or horizontal relationships might not be the only explanations for differences in learning outcomes. Antecedents at the school level, for example, are often influenced by process variables at the system level like educational policies. Another example is the possibility that the socio-cultural context (antecedent at the system level) might have an influence on instructional practices (process at the classroom level), which in turn leads to differences in student outcomes.

An important corollary of the intricate relationships between the various cells in Figure 3.1 is that each one of the observed variables is likely to convey multiple information (*i.e.* both information on the dimension that the variable is intended to measure, and information on related antecedents or process variables). For example, the variables identifying the study programme or grade of the students not only contain direct information on their instructional setting and curriculum, but, in many cases, also indirect information on students' probable prior level of achievement, maybe of their home background, and possibly some of the characteristics of their teachers.

In view of the complexity of potential relationships between these variable types, explicit causal relationships were not included in this conceptual mapping. There are too many potential relationships between these components (including cross-level relationships) that might be relevant for PISA and which could not be integrated into one 'general' conceptual model.



Figure 3.2

The two-dimensional conceptual matrix with examples of variables collected or available from other sources

	Antecedents	Processes	Outcomes
The education system as a whole	**Cell 1:** Macro-economic and demographic context *For example:* ▪ Gross Domestic Product ▪ Distribution of wealth (Gini index) ▪ Percentage of immigrants	**Cell 5:** Policies and organisation of education *For example:* ▪ Organisation of education (school autonomy, programme structure) ▪ Teacher qualifications and training requirements ▪ School entry age, retention	**Cell 9:** Outcomes at the level of the education system *For example:* ▪ System level aggregates of: reading, mathematical and scientific literacy ▪ Habits in relation to content domains ▪ Attitudinal outcomes ▪ Life skills and learning strategies ▪ Equity related outcomes
Educational institutions	**Cell 2:** Characteristics of educational institutions *For example:* ▪ The involvement of parents ▪ Social intake ▪ Source of funding, location and size ▪ Type of educational provider (*e.g.* out-of-school, educational media programme)	**Cell 6:** Institutional policies and practice *For example:* ▪ Instructional support including both material and human resources ▪ Policies and practices, including assessment and admittance policies ▪ Activities to promote student learning	**Cell 10:** Learning outcomes at the institutional level *For example:* ▪ Institution level aggregates of: reading, mathematical and scientific literacy ▪ Habits in relation to content domains ▪ Affective outcomes (*e.g.* attitudes to mathematics) ▪ Life skills and learning strategies ▪ Differences in outcomes for students of various backgrounds
Instructional settings	**Cell 3:** Characteristics of instructional settings *For example:* ▪ Teacher qualifications ▪ Classroom size	**Cell 7:** Learning environment *For example:* ▪ Ability grouping ▪ Teaching styles ▪ Learning time	**Cell 11:** Learning outcomes at the level of instructional setting *For example:* ▪ Classroom motivation to learn ▪ Average classroom performance
Individual participants in education and learning	**Cell 4:** Individual background *For example:* ▪ Parental occupational status ▪ Parental educational level ▪ Educational resources at home ▪ Ethnicity and language ▪ Age and gender	**Cell 8:** Individual learning process *For example:* ▪ Engagement and attitudes to science ▪ Self-concept and self-efficacy when learning science ▪ Motivation to learn science	**Cell 12:** Individual outcomes *For example:* ▪ Reading, mathematical and scientific literacy ▪ Affective outcomes (*e.g.* attitudes to science)



Therefore, this conceptual mapping provides a point of reference in the conceptual framework for PISA 2006 rather than as a general 'PISA model'. More detailed models should be developed for particular research areas and for specific relationships. Relevant variables in these more specific models, however, could still be located within this conceptual two-dimensional matrix.

Figure 3.2 shows examples of variables that were collected or are available for each cell of the two-dimensional conceptual matrix that has guided the development of context questionnaire for PISA 2006.

RESEARCH AREAS IN PISA 2006

PISA's contributions to policy makers' and educators' needs were maximised by identifying possible policy-relevant research areas and choosing carefully from among the many possibilities so that the strengths of the PISA design were capitalised on.

The following research areas were developed following recommendations from the Questionnaire Expert Group:

- Student's engagement in science: In part, this research area parallels the research area on engagement in mathematics in PISA 2003. However it has been expanded to incorporate aspects of the affective dimension more comprehensively, but in a way that is not bound to a 'cognitive unit context'. It covers self-related cognitions, motivational preferences, emotional factors as well as behaviour-related variables (such as participation in science-related activities in and out of school);
- Teaching and learning of science: This research area addresses how instructional strategies are used to teach science at school and to what extent science instruction is different across types of education and schools;
- Scientific literacy and environment: It is of interest to policy-makers to what extent schools contribute to the awareness of and attitudes toward environmental problems and challenges among 15-year-old students. This is an area related to scientific literacy (OECD, 2006) and school instruction in this area can be regarded as a potential source of information;
- Organisation of educational systems: This research area explores the relationships between scientific literacy and structural characteristics of educational systems, such as general vs. specialised curricula, comprehensive vs. tracked study programmes, centralised vs. decentralised management of schools;
- Science attainment and the labour market: The role and value of science education and scientific literacy as a preparation for future occupation are discussed in this research area, both in terms of students' expectations and school practices concerning orientation and information for students about science-related careers.

The following two research areas had been also been developed but were not retained for the main survey after reviewing the field trial results and after the PGB decided on the priorities for the final data collection:

- Student performance and gender: This research area focused on student performance in all three major domains and comprised not only data from PISA 2006 but also from previous PISA cycles and previous international studies (IEA mathematics and science studies, IEA reading literacy studies);
- Parental investment and scientific literacy: This research area was concerned with the effects of parental involvement and parenting styles on students' science-related career expectations and scientific literacy.

Table 3.1 shows for each research area the main constructs and variables that were included in the PISA 2006 main data collection to explore each of the research areas.



Table 3.1
Themes and constructs/variables in PISA 2006

Research area	Constructs or variables
Student engagement in science	Science self-efficacy (StQ)
	Science self-concept (StQ)
	Interest in learning science (StQ)
	Enjoyment of learning science (StQ)
	Instrumental motivation to learn science (StQ)
	Future-oriented science motivation (StQ)
	General value of science (StQ)
	Students' personal value of science (StQ)
	Students' science-related activities (StQ)
	Parents' general value of science (PaQ)
	Parents' personal value of science (PaQ)
Teaching and learning of science	Interactive science teaching (StQ)
	Hands-on science teaching activities (StQ)
	Student investigation in science lessons (StQ)
	Science teaching with focus on applications (StQ)
	Time spent on learning science (StQ)
Scientific literacy and the environment	Students' awareness of environmental issues (StQ)
	Students' perception of environmental issues (StQ)
	Students' environmental optimism (StQ)
	Responsibility for sustainable development (StQ)
	School activities to promote environmental learning (ScQ)
	Parents' perception of environmental issues (PaQ)
	Parents' environmental optimism (PaQ)
Organisation of educational systems	School size, location and funding (ScQ)
	Grade range (ScQ)
	Class size (ScQ)
	Grade repetition at school (ScQ)
	Ability grouping (ScQ)
	Teacher-student ratio (ScQ)
	Computer availability at school (ScQ)
	School selectivity (ScQ)
	School responsibility for resource allocation (ScQ)
	School responsibility for curriculum & assessment (ScQ)
	School accountability policies (ScQ)
	Assessment practices (ScQ)
	Activities to promote engagement with science learning
	Teacher shortage (ScQ)
	Quality of educational resources (ScQ)
	Parents' perception of school quality (PaQ)
Science attainment and the labour market	School preparation for science career (StQ)
	School information on science careers (StQ)
	Expected occupation at 30 (StQ)
	Career preparation at school (ScQ)
	Student's science activities at age 10 (PaQ)
	Parents' views on importance of science (PaQ)
	Parents' view on student's science career motivation (PaQ)

Note: StQ = Student questionnaire; ScQ = School questionnaire; PaQ = Parent questionnaire.



THE DEVELOPMENT OF THE CONTEXT QUESTIONNAIRES

From the theoretical bases of each research area, as elaborated, a large number of constructs were defined and their measurement operationalised through obtaining or writing questionnaire items (often in item batteries to form scales).

Small scale trials were undertaken in a range of countries and languages. Firstly a pre-pilot with a small convenience sample was undertaken in Australia. It involved a think aloud process where students were asked to complete the questionnaire while verbalising their thought processes. The pre-pilot provided qualitative feedback on the understanding and appropriateness of the items. After refining the items in light of the pre-pilot results, a series of pilot studies was undertaken in Japan (Japanese), Germany (German), Canada (French) and Australia (English). The pilots consisted of collecting questionnaire data from small convenience samples in each country. After data collection, students were collectively interviewed about their understanding of each question, particularly probing for relevance and ambiguity. The pilot therefore yielded both quantitative and qualitative data, plus conducting group interviews on the questions.

After further refinement of the questions, data was gathered in 2005 from a full scale field trial of student, school and parent questionnaires in each of the 57 participating countries in over 40 languages. The field trial was able to facilitate the investigation of a large number of student questionnaire items through the use of a rotational design with four questionnaire forms that were randomly allocated to students.

In addition, the field trial was used for in-depth analysis of the following aspects:

- Two sets of items were trialled as dichotomous and Likert-type items in parallel forms to explore cross-cultural differences in responses to either item type. Results showed some tendencies to more extreme responses in some countries but on balance it seemed more appropriate to use Likert-type items in the PISA questionnaires (Walker, 2006; Walker, 2007);
- Two sets of items were trialled with different category headings: Nine items measuring control strategies for science learning were trialled in one version asking about frequencies and in another one asking about agreement. Seven items measuring student participation in activities to protect the environment were trialled both with categories reflecting frequencies and with categories reflecting both frequency and intent. The field trial data were analysed to decide on the more appropriate but neither set of items was included in the final main study questionnaire;
- Two different sets of items measuring science self-efficacy were trialled. One set of items included asked about student confidence in tasks related to general science understanding, the other set about student confidence in doing science subject-specific tasks. Both of the item sets had good scaling characteristics and it was decided to retain the items measuring self-confidence in general science tasks due to a better fit with the science literacy framework;
- Student and parent questionnaire data were used to explore the consistency of responses regarding parental education and occupation. Results showed relatively high consistency between student and parent reports on occupation but somewhat lower consistencies for data on educational levels (Schulz, 2006).

Empirical analyses included the examination of:

- The frequency of missing values by country;
- The magnitude and consistency of item-total score correlations for each scale, by country;



- The magnitude and the consistency of scale reliability (Cronbach's Alpha), by country;
- The magnitude and consistency of correlations with each scale and science achievement as determined in the PISA field trial science test, by country;
- Exploratory and confirmatory factor analyses to determine construct validity and reliability of each scale across the pooled sample;
- Multiple-group models to assess the parameter invariance of factor models across countries;
- Item Response Theory (IRT) analyses to determine item fit for the pooled sample;
- Item-by-country interaction of items across countries using IRT scaling.

In addition to the empirical analyses, the choice of items, item format and wording was informed by:

- Direction from the PISA Governing Board;
- Feedback from National Project Managers;
- Feedback from linguistic experts;
- Discussions with the Questionnaire Expert Group;
- Discussions with members of the Science Expert Group;
- Consultation with science forum nominees of the PISA Governing Board;
- Consultation with the OECD secretariat.

Finally, in October 2005 a large and comprehensive set of potential items and topics was provided to the PISA Governing Board. From this set, the PGB indicated priority areas for investigation.

THE COVERAGE OF THE QUESTIONNAIRE MATERIAL

Student questionnaire

The student questionnaire was administered after the literacy assessment and it took students about 30 minutes to complete the instrument. The core questions on home background were similar to those used in PISA 2003, however, for some questions the wording was modified to improve the quality of the data collection based on experiences in previous surveys. Appendix 5 lists the core questions with changes in wording from PISA 2003 to PISA 2006.

The questionnaire covered the following aspects:

- Student characteristics: Grade, study programme, age and gender;
- Family background: Occupation of parents, education of parents, home possessions, number of books at home, country of birth for student and parents, language spoken at home;
- Students' views on science: Enjoyment of science, confidence in solving science tasks, general and personal value of science, participation in science-related activities, sources of information on science and general interest in learning science;
- Students; views on the environment: Awareness of environmental issues, source of information on the environment, perception of the impact of environmental issues, optimism about environmental issues and sense of responsibility for sustainable development;



- Students' views of science-related careers: Usefulness of schooling as preparation for the science labour market, information about science-related careers, future-oriented motivations for science and expected occupation at 30;
- Students' reports on learning time: Mode and duration of students' learning time in different subject areas and duration of students' out-of-school lessons;
- Students' views on teaching and learning of science: Science course taking in current and previous year, nature of science teaching at school (interactive, hands-on activities, student investigations and use of applications), future-oriented motivations to learn science, importance of doing well in subject areas (science, mathematics and test language subjects) and academic self-concept in science.

School questionnaire

The school questionnaire was administered to the school principal and took about 20 minutes to be completed. It covered a variety of school-related aspects:

- Structure and organisation of the school: Enrolment, ownership, funding, grade levels, grade repetition, average test language class size, community size and tracking/ability grouping;
- Staffing and management: Number of teachers, availability of science teaching staff, responsibility for decision-making at school and influences of external bodies on school-level decisions;
- The school's resources: Number of computers at school and principals' views on quality and quantity of staffing and educational resources;
- Accountability and admission practices: Accountability to parents, parental pressure on school, use of achievement data, parental choice of local school(s) and school admittance policies;
- Teaching of science and the environmental issues: School activities to promote learning of science, environmental issues in school curriculum and school activities to promote learning of environmental issues; and
- Aspects of career guidance: Students' opportunities to participate in career information activities, student training through local businesses, influence of business on school curriculum and structure of career guidance at school.

International options

As in previous surveys, additional questionnaire material was developed, which was offered as international options to participating countries. In PISA 2006, two international options were available, the ICT Familiarity questionnaire and the parent questionnaire.

Information communication technology (ICT) familiarity questionnaire

The ICT familiarity questionnaire consisted of questions regarding the students' use of, familiarity with and attitudes towards information communication technology which was defined as the use of any equipment or software for processing or transmitting digital information that performs diverse general functions whose options can be specified or programmed by its user. The questionnaire was administered to students after the international student questionnaire (sometimes combined within the same booklet) and it took about five minutes to be completed. It covered the following ICT-related aspects:

- Use of ICT: Students' experience with computers at different locations and frequency of ICT use for different purposes;
- Affective responses to ICT: Confidence in carrying out ICT-related tasks.

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



Parent questionnaire

The parent questionnaire covered both parental social background and aspects related to some of the research areas. It took about ten minutes to complete and one questionnaire was administered per student. The questionnaire covered the following aspects:

- Parental reports related to school and science learning: The students' past science activities, parental perceptions of value and quality of the student's schooling, parental views on science-related careers and parental general and personal value of science;
- Parental views on the environment: Parental awareness of environmental views and environmental optimism;
- Annual spending on children's education;
- Parental background: Age, occupation (both parents), education (both parents) and household income.

National questionnaire material

National centres could add nationally specific items to any of the questionnaires. Insertion of national items into the international questionnaires had to be agreed upon with the international study centre during the review of adaptations. National student questionnaire options, which took no longer than ten minutes to be completed, could be administered after the international student questionnaire and international options. If the length of the additional material exceeded ten minutes, national centres were requested to administer their national questionnaire material in follow-up sessions.

THE IMPLEMENTATION OF THE CONTEXT QUESTIONNAIRES

In order to make questions understood by 15-year-old students, their parents and school principals in participating countries, it was necessary to adapt parts of the questionnaire material from the international source version to the national context without jeopardising the comparability of the collected data. This is particularly important for questions that relate to specific aspects of educational systems like educational levels, study programmes or certain school characteristics which differ in terminology across countries.

To achieve maximum comparability, a process was implemented during which each adaptation was reviewed and discussed by the international study centre and national study centres. To facilitate this process, national centres were asked to complete a questionnaire adaptation spreadsheet (QAS), where adaptations to the questionnaire material were documented.

Each adaptation had to be reviewed and agreed upon before the questionnaire material could be submitted for linguistic verification and the final optical check (see Chapter 5). The QAS also contained information about additional national questionnaire material and any deviation from the international questionnaire format.

Prior to the review of questionnaire adaptations, national centres were asked to complete three different tables describing necessary adaptations:

- Study programme tables (STP): These document the range of different study programmes that are available for 15-year-old students across participating countries. This information was not only used as a codebook to collect these data from school records but also assisted the review of questionnaire adaptations;
- Language tables (LNT): These document the language categories included in the question about language use at home; and
- Country tables (CNT): These document the country categories in the questions about the country of birth for students and parents.

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



Information on parental occupation and the students' expected occupation was collected through open-ended questions both in student and parent questionnaires. The responses were then coded according to the International Standard Classification of Occupations (ISCO) (International Labour Organisation, 1990). Once occupations had been coded into ISCO, the codes were re-coded into the International Socio- Economic Index of Occupational Status (*ISEI*) (Ganzeboom, de Graaf & Treiman, 1992), which provides a measure of the socio-economic status of occupations comparable across the countries participating in PISA.

The International Standard Classification of Education (ISCED) (OECD, 1999) was used as a typology to classify educational qualifications and study programmes. The ISCED classification was used to get comparable data across countries. Whereas this information was readily available for OECD member countries, for partner countries and economies extensive reviews of their educational systems in cooperation with national centres were necessary to map educational levels to the ISCED framework.



4

Sample design

Target population and overview of the sampling design 64
Population coverage, and school and student participation rate standards 65
- Coverage of the PISA international target population 65
- Accuracy and precision 66
- School response rates 66
- Student response rates 68

Main study school sample 68
- Definition of the national target population 68
- The sampling frame 69
- Stratification 70
- Assigning a measure of size to each school 74
- School sample selection 74
- PISA and TIMSS or PIRLS overlap control 76
- Student samples 82

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



TARGET POPULATION AND OVERVIEW OF THE SAMPLING DESIGN

The desired base PISA target population in each country consisted of 15-year-old students attending educational institutions located within the country, in grades 7 and higher. This meant that countries were to include (i) 15-year-olds enrolled full-time in educational institutions, (ii) 15-year-olds enrolled in educational institutions who attended on only a part-time basis, (iii) students in vocational training types of programmes, or any other related type of educational programmes, and (iv) students 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 persons schooled in the home, workplace or out of the country would occur and therefore these students 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 2006 and 31 August 2006, unless otherwise agreed, during which they would administer the assessment.

Further, testing was not permitted during the first six weeks of the school year because of a concern that student performance levels may be 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 2006, the target population could have been defined as all students born in 1990 who were attending a school or other educational institution.

Further, a variation of up to one month in this age definition was permitted. This was done to allow a country testing in March or in May to still define the national target population as all students born in 1990. If the testing was to take place at another time until the end of August, the birth date definition had to be adjusted.

In all but one country, 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 eligible schools – the school sampling frame – with probabilities that were proportional to a measure of size. This is referred to as systematic probability proportional to size (or PPS) sampling. The measure of size was a function of the estimated number of eligible 15-year-old students enrolled. Prior to sampling, schools in the sampling frame were assigned to mutually exclusive groups called explicit strata, formed 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 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 although with agreement countries could use alternative values. From each list of students that contained more than the *TCS*, the *TCS* students were selected with equal probability and for lists of fewer than the *TCS*, all students on the list were selected.

In one country, a three-stage design was used. In this case, geographical areas were sampled first (first-stage units) using probability proportional to size sampling, and then schools (second-stage units) were selected within sampled areas. Students were the third-stage sampling units in three-stage designs.



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.

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

Coverage of the PISA international target population

NPMs might find 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 internationally 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 Technical Standards.[1]

Exclusion can take place at the school level (the whole school is excluded) or at the within-school level. Areas deemed by the PGB to be part of a country (for the purpose of PISA), but which were not included for sampling, were designated as non-covered areas, and documented as such – although this occurred infrequently. 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 perform in the PISA testing situation. 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 perform in the PISA testing situation. Functionally disabled students who could respond were to be included in the testing;
- Students with insufficient assessment language experience are students who need to meet all of the following criteria: a) are not native speakers of the assessment language(s), b) have limited proficiency in the assessment language(s), and c) have received less than one year of instruction in the assessment language(s). Students with insufficient assessment language experience could be excluded;
- Not assessable for some other reason as agreed upon. A nationally-defined within-school exclusion category was permitted if agreed upon by the consortium. A specific subgroup of students (dyslexic, for example) 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.

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 be kept below 5%. Restrictions on the level of exclusions of various types were as follows:

- School-level exclusions for inaccessibility, feasibility or other reasons were required to cover fewer than 0.5% of the total number of students in the International Target Population. Schools on the school sampling frame which had only one or two 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 schools would not cause a breach of the 0.5% allowable limit, then such schools could be excluded in the field if at that time, they still only had 1 or 2 PISA eligible students;
- School-level exclusions for intellectually or functionally disabled students, or students with insufficient assessment language experience, were required 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 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 a largely unpredictable part of each country's 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 (or all schools if there were fewer than 150 schools in a participating country) had to be selected in each country. Within each participating school, a predetermined number of students, denoted as *TCS* (usually 35), 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, 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, 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 stratification variables to reduce the sampling variance.

For countries that had participated in PISA 2003 that had larger than anticipated sampling variances associated with their estimates, recommendations were made about sample design changes that would help to reduce the sampling variances for PISA 2006. These included modifications to stratification variables, and increases in the required sample size.

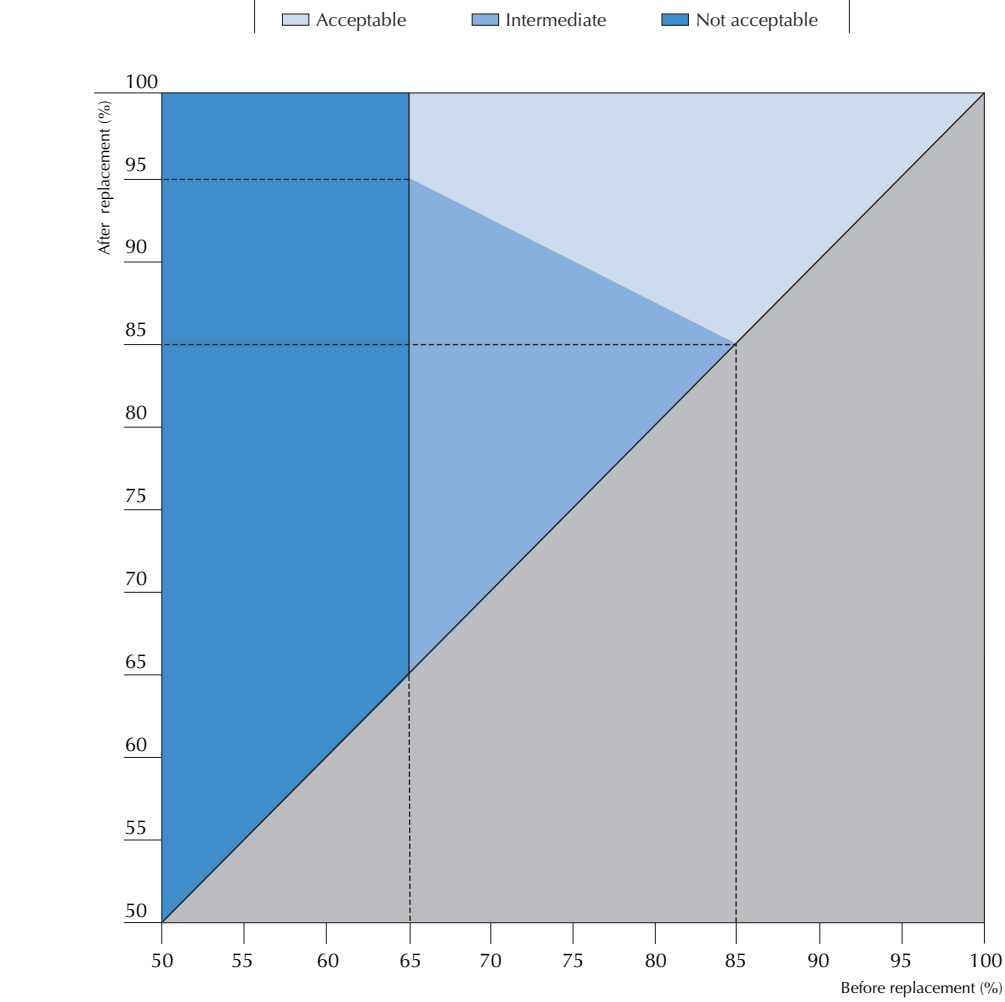
School response rates

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. 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 50%. 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 be introduced to the extent that the students who were absent were 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 is increased because of sample size attrition.

Figure 4.1
School response rate standard



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



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 an arbitrary one, 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 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 rapidly decrease.

These PISA standards applied to weighted school response rates. The procedures for calculating weighted response rates are presented in Chapter 8. Weighted response rates weight 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 in general selected with probability proportional to size, in most countries most schools contributed equal weights, or approximately so, as a consequence the weighted and unweighted school response rates were very similar. Exceptions could occur in countries that had explicit strata that were sampled at very different rates. Details as to how the PISA participants performed relative to these school response rate standards are included in Chapters 10 and 13.

Student response rates

A 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 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% 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 STUDY 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. Once these were approved, NPMs were alerted to avoid having the 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 eligible students, the deviations were described and enrolment data provided to measure how much 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 booklet, consisting of a subset of the PISA assessment items, deemed more suitable for students with special education needs.

Within participating schools, all eligible students (*i.e.* born within the defined time period and in grades 7 or higher) were to be listed. From this, either a sample of *TCS* students was randomly selected or all students



were selected if there were fewer than *TCS* 15-year-olds. 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. Country participant 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 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 absent sampled students.

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 by the School Sampling Preparation manual 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 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 very careful in constructing their 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 probability-proportional to size selection probabilities were based on it for both two and three-stage designs. The best *ENR* for PISA would have been the number of currently enrolled 15-year-old students. Current enrolment data, however, were rarely available at the time of 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 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;
- Total student enrolment, divided by the number of grades in the school.

The *School Sampling Preparation Manual* [3] noted that if reasonable estimates of *ENR* did not exist or if the available enrolment data were too out of date, schools might have to be selected with equal probabilities which might require an increased school sample size. No countries needed 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;
- Coded information about the school, such as region of country, school type and extent of urbanisation, which could be used as stratification variables.

As noted, a three-stage design and an area-level 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 (called the first stage of sampling) before actually sampling schools (the second stage of sampling). 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 about half of them needed to be sampled), NPMs were also instructed in the *School Sampling Preparation Manual* 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 consortium and received additional support in 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.

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 some 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 states, provinces, or other regions;
- Make sure that all parts of a population were included in the sample;
- Ensure adequate representation of specific groups of the target population in the sample.

There were two types of stratification possible: explicit and implicit. Explicit stratification consists of building separate school lists, or sampling frames, according to the set of explicit stratification variables under consideration. Implicit stratification consists essentially of sorting the schools within each explicit stratum by a set of implicit stratification variables. This type of stratification is a very simple 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). Guidelines were provided in the *School Sampling Preparation Manual* on how to go about choosing stratification variables.

Table 4.1 provides the explicit stratification variables used by each country, as well as the number of explicit strata, and the variables and their number of levels used for implicit stratification. As countries were requested to sort the sampling frame by school size, school size was also an implicit stratification variable, though it is not listed in Table 4.1. A variable used for stratification purposes is not necessarily included in the PISA data files.



Table 4.1 [Part 1/2]
Stratification variables

	Explicit stratification variables	Number of explicit strata	Implicit stratification variables
OECD			
Australia	State/Territory (8); Sector (3); School Size (3)	37	Geographic Zone (8); School Level for TAS and ACT Government Schools (3)
Austria	Programme (19); School Size (3)	20	Province-District (121)
Belgium			
Belgium (Flanders)	Form of Education (5); Public/Private (2); School Size (3);	11	Index of Over-aged Students
Belgium (French)	Special Education/Other (2); Public/Private School Types for Regular Schools (4)	5	Public/Private School Types for Special Education Schools (4); Index of Over-aged Students for Regular Schools
Belgium (German)	One Explicit Stratum (All of German Belgium)	1	None
Canada	Province (10); Language (3); School Size (4); Certainty Selections	44	Public/Private(2); Urban/Rural(2)
Czech Republic	Programmes (6); Region for Programmes 1 and 2 (14); School Size (4)	76	Region for Programmes 3, 4, 5, 6 (14)
Denmark	School Size (3)	3	School Type (5); Geo Area (5)
Finland	Region (6); Urban/Rural (2)	12	None
France	School Type (4); School Size (3)	6	None
Germany	School Category (3); State (16) for Normal Schools; School Size (3)	20	School Type for Normal Schools (5); State for Other Schools (16)
Greece	Region (10); Public/Private (2); Evening/Non-Evening (2); School Size (3)	16	School Type (3); Public/Private (2) when both in an explicit stratum
Hungary	School Type (4); School Size (3)	5	Region (7); National Grade 10 Math Score Categories (5) for Non-Primary Schools
Iceland	Region (9)	9	Urban/Rural (2); School Size (4)
Ireland	School Size (3)	3	School Type (3); School Gender Composition Categories (5)
Italy	Area (17); Programme (5); School Size (3); Certainty Selections	87	Public/Private (2)
Japan	Public/Private (2); School Type (2)	4	Levels of proportion of students taking University/College Entrance Exams(4)
Luxembourg	School Type (6)	6	None
Mexico	State (32); School Level (2); School Size (3); Certainty Selections	67	School Size (3); Public/Private (2); Urban/Rural (2); School Level (2); School Program (4 For Each School Level)
Netherlands	School Track (2)	2	School Type (6 for School Track A and 3 for School Track B)
New Zealand	Certainty/Non-Certainty (2)	2	Public/Private (2); Socio-Economic Status Category (3) and Urban/Rural (2) for Public Schools
Norway	School Type (2); Size (3)	4	None
Poland	Public Upper Secondary Lycea/Other Public (2); School Size (3) for Private Schools	5	Urbanisation (4)
Portugal	Region (7); School Type (4); School Size (3); Certainty Selections	27	Public/Private (2); Socio-Economic Status Category (4)
Korea	Urbanicity (3); School Program (3); School Size (2)	5	School Level (2)
Scotland	School S-Grade Attainment (5)	5	None
	Certainty Selections (1)	2	
Slovak Republic	Region (8); School Type (3); School Size (3)	26	Educational Programme (9); Language (2) in 4 of the Regions
Spain	Region (18); Public/Private (2); Teaching Modality for Basque (3); School Size (4); Certainty Selections	55	Postal Code for all
Sweden	School Size (2); Public/Private (2) for Lower Secondary schools; Urbanicity (5) for large Public Lower Secondary Schools; School Level (2)	10	Urbanisation (5) for Private Lower Secondary schools; Public/private (2) for Upper Secondary schools; Administrative Province (25) for Upper Secondary schools; Income Quartiles (4) for all except Private Lower Secondary schools
Switzerland	School has Grade 9 or not (2); Language (3); School Type (28) for Upper Secondary Schools; Public/Private (2); School Size (4); Certainty Selections	48	School Type (28); Canton (26)
Turkey	Regions (7); School Size (2); Certainty Selections	9	School Level (3); Public/Private (2); Urban/Rural (2)
United Kingdom	PRU/Non-PRU (2), Country (3), Certainty Selections (1)	10	School Type (6), GCSE Performance (6), Region (4), Local Authority, Education and Library Board Region (5)
		2	
		2	
United States	School Size (2)	2	Public/Private (2); Region (4); Urbanisation (3); Minority Status (2); Grade Span (4); Postal Code

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



Table 4.1 [Part 2/2]
Stratification variables

		Explicit stratification variables	Number of explicit strata	Implicit stratification variables
Partners	Argentina	Province (24); School Size (3)	27	Sector (2); School Type (5); School Level (5)
	Azerbaijan	Language (2); Public/Private(2); Education Department (3); School Type (4); School Size (3)	9	Urbanisation (4); Education Department (5)
	Brazil	State (27); School Size (3); Certainty Selections	30	School Type (3); HDI Category (2); School Size (3); Urban/Rural (2); Capital/Non-Capital (2)
	Bulgaria	Region (11); School Size (3)	13	School Type (3); Settlement Size (5); State/Municipal (2); Public/Private (2)
	Chile	School Administration (3); School Level (7); School Size (3);	17	Urban/Rural (2); Region (13)
	Colombia	School Size (3)	3	Urban/Rural (2); Public/Private(2)
	Croatia	Dominant Programme(6); Urbanicity (3); School Size (2); Primary Schools (1); Certainty Selections	16	County (21)
	Estonia	Language (2); School Size (3); Certainty Selections	6	Urbanisation (4); School Type (4); County (15)
	Hong Kong-China	School Type (4)	4	Student Academic Intake (4)
	Indonesia	Provinces (26); School Size (3)	28	School Type (5); Public/Private (2); National Achievement Score Categories (3)
	Israel	Inspection (5); School Size (3)	9	Location (9) for Public Schools, Except For Schools in Druz Migzar Sector; Group Size (5) for Regular Public Schools; Gender Composition (3) for Religious Public Schools; Migzar Sector (3) for Regular Public Arabic Schools; Cultivation Categories (4) for
	Jordan	School Type (4); School Size (3); Certainty Selections	6	Urbanisation (2); School Gender Composition (3); School form (2)
	Kyrgyzstan	Regions (9); Urbanicity (3); Language (3); School Size (3); Certainty Selections	43	School Type (5)
	Latvia	School Size (3); Certainty Selections	4	Urbanisation (4); School Type (4)
	Liechtenstein	One Explicit Stratum (All of Liechtenstein)	1	None
	Lithuania	School Type (4); School Size (3)	6	Urbanisation (3); Public/Private(2)
	Macao-China	School Type (3); Programme (2); Language (5)	10	Secondary Levels (3)
	Montenegro	Primary/Secondary (2); Region (3) for Secondary Schools	4	Region (3) for Primary Schools; Urban/Rural (2); School Type (3)
	Qatar	School Type (7); School Gender Composition Categories (3); School Level (3)	26	Qatari/Non-Qatari (2)
	Romania	School Program (3); School Size (3)	5	Language (3); Urbanisation (2)
	Russian Federation	Region PSU (45)	45	Urbanisation (9); School Type (4); School Sub-Type (16)
	Serbia	Region (8); School Size (2); Certainty Selections	10	Urban/Rural (2); School Type (7)
	Slovenia	School Programme (6); School Size (2); Certainty Selections	8	Urbanisation (4)
	Chinese Taipei	Region (6); School Type (7); School Size (3); Certainty Selections	17	Public/Private (2)
	Thailand	Department (6); School Type (3); School Size (3); Certainty Selections	12	Local Area (9)
	Tunisia	Public/Private (2); School Type (2); For General Public Schools: East/West (2) and School Level (3); School Size (2) for all; Certainty Selections	10	Category of Grade Repeating (3) for General Public Schools; East/West (2) for Private Schools and Vocational Schools; North/South (2) for all
	Uruguay	School Type (4); Programme (3 or 5 Depending on School Type); School Size (3); Certainty Selections	16	Area (4); Shift (4) for Public Secondary Schools; Shift (4) for Public Technical Schools



Treatment of small schools in stratification

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 in most countries). A very small school had an *ENR* less than one-half the *TCS* (17 or less in most countries)**.** A moderately small school had an *ENR* in the range of one-half the *TCS* to *TCS* (17 to 35 in most countries). 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 in-school sample size would fall short of expectations. A sample with many small schools could also be an administrative burden. To minimise these problems, procedures for stratifying and allocating school samples were devised for small schools in the sampling frame.

To determine what was needed – a single stratum of small schools (very small and moderately small combined), or a stratum of very small schools only, or two strata, one of very small schools and one of moderately small schools, or no small school strata – the *School Sampling Preparation Manual* stipulated that if:

- The percentage of students in very small schools was 1% or more and the percentage of students in moderately small schools was 4% or more, then an explicit stratum of moderately small schools and an explicit stratum for very small schools were required;
- Otherwise, if the percentage of students in very small schools was 1% or more, a stratum for very small schools was needed, but no stratum for moderately small schools;
- Otherwise, if the percentage of students in very small schools was less than 1%, and the percentage of students in moderately small schools was 4% or more, a combined stratum for small schools which included all very small and moderately small schools, was needed;
- Otherwise no small school strata were needed.

The small school strata were always sorted first by the explicit stratum to which they originally belonged, followed by the other defined implicit stratification variables.

When small schools were explicitly stratified, it was important to ensure that an adequate sample was selected without selecting too many small schools as this would lead to too few students in the assessment. In this case, the entire school sample would have to be increased to meet the target student sample size.

The sample had to be proportional to the number of students and not to the number of schools. Suppose that 10% of students attend moderately small schools, 10% very small schools and the remaining 80% attend large schools. In the sample of 5 250, 4 200 students would be expected to come from large schools (*i.e.* 120 schools with 35 students), 525 students from moderately small schools and 525 students from very small schools. If moderately small schools had an average of 25 students, then it would be necessary to include 21 moderately small schools in the sample. If the average size of very small schools was 10 students, then 52 very small schools would be needed in the sample and the school sample size would be equal to 193 schools rather than 150.

To balance the two objectives of selecting an adequate sample of explicitly stratified small schools, a procedure was recommended that assumes identifying strata of both very small and moderately small schools. The underlying idea is to under-sample by a factor of two the very small school stratum and to increase proportionally the sizes of the large school strata. When there was just a single small school stratum, the procedure was modified by ignoring the parts concerning very small schools. The formulae below assume a target school sample size of 150 and a target student sample size of 5 250.

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



- Step 1: From the complete sampling frame, find the proportions of total *ENR* that come from very small schools (*P*), moderately small schools (*Q*), and larger schools (those with *ENR* of at least *TCS*) (*R*). Thus, $P + Q + R = 1$.
- Step 2: Calculate the figure *L*, where $L = 1.0 + (P/2)$. Thus *L* is a positive number slightly more than 1.0.
- Step 3: The minimum sample size for larger schools is equal to 150 × R × L, rounded 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 (*VENR*). *MENR* is a number in the range of *TCS/2* to *TCS*, and *VENR* is a number no greater than *TCS/2*.
- Step 5: The number of schools that must be sampled from the stratum of 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 stratum of very small schools is given by: $(2\,625 \times P \times L)/(VENR)$.

To illustrate the steps, suppose that in participant country *X*, the *TCS* is equal to 35, with 0.1 of the total enrolment of 15-year-olds each in moderately small schools and in very small schools. Suppose that the average enrolment in moderately small schools is 25 students, and in very small schools it is 10 students. Thus $P = 0.1$, $Q = 0.1$, $R = 0.8$, $MENR = 25$ and $VENR = 10$.

From Step 2, $L = 1.05$, then (*Step 3*) the sample size of larger schools must be at least $150 \times (0.80 \times 1.05) = 126$. That is, at least 126 of the larger schools must be sampled. From Step 5, the number of moderately small schools required is $(5\,250 \times 0.1 \times 1.05)/25 = 22.1$ – *i.e.*, 22 schools. From Step 6, the number of very small schools required is $(2\,625 \times 0.1 \times 1.05)/10 = 27.6$ – *i.e.*, 28 schools.

This gives a total sample size of 126 + 22 + 28 = 176 schools, rather than just 150, or 193 as calculated above. Before considering school and student non-response, the larger schools will yield a sample of $126 \times 35 = 4\,410$ students. The moderately small schools will give an initial sample of approximately $22 \times 25 = 550$ students, and very small schools will give an initial sample size of approximately $28 \times 10 = 280$ students. The total initial sample size of students is therefore 4 410 + 550 + 280 = 5 240.

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 constructed as: *MOS* = max (*ENR, TCS*).

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

As sample schools were selected according to their size (PPS), setting the measure of size of small schools to 35 is equivalent to drawing a simple random sample of small schools.

School sample selection

Sorting the sampling frame

The *School Sampling Preparation Manual* indicated that, prior to selecting schools from the school sampling frame, 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 sorting variable, and so on, until all implicit stratification variables were exhausted. 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.

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 eligible students in each corresponding explicit stratum. There were two exceptions. If an explicit stratum of very small schools was required, students in them had smaller percentages in the sample than those in the population. To compensate for the resulting loss of sample, the large school strata 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 these cases, two schools were allocated for selection in the stratum to aid with variance estimation.

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: $I = S/D$;
- Recording the sampling interval, *I*, to four decimal places.

Next, a random number (drawn from a uniform distribution) had to be selected for each explicit stratum**.** The generated random number (*RN*) was to be a number between 0 and 1 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 *D* 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 simply 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 *D*) 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.

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



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

PISA and TIMSS or PIRLS overlap control

The main studies for PISA 2006 and the 2007 Trends in International Mathematics and Science Study (TIMSS) were to occur at approximately the same time in southern hemisphere countries and in northern hemisphere countries with late PISA testing. Furthermore, the PISA 2006 main study and the 2006 Progress in International Reading Literacy Study (PIRLS) were to occur at approximately the same time. Because of the potential for increased burden, an overlap control procedure was used for eight countries (Australia, Bulgaria, England, Hong Kong-China, Hungary, Scotland, Tunisia, and the USA) who wished for there to be a minimum incidence of the same schools being sampled for PISA and TIMSS (Australia, Bulgaria, England, Hong Kong-China, Scotland, Tunisia, and the USA) or a minimum of the same schools for PISA and PIRLS (Hungary). This overlap control procedure required that the same school identifiers be used on the TIMSS or PIRLS and PISA school frames for the schools in common.

The TIMSS and PIRLS samples were selected before the PISA samples. Thus, for countries requesting overlap control, the TIMSS and PIRLS International Study Center supplied the PISA consortium with their school frames, with the school IDs, the school probability of selection for each school, and an indicator showing which schools had been sampled for the relevant study.

Sample selections for PISA and the other 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 school burden. Each study's project manager had to decide if this was the path they wished to adopt. If they decided against this capping of probabilities, then it might have been possible for some large schools to be in both the PISA and the other study's samples. Among the countries choosing overlap control in the 2006 PISA, selection probabilities were capped at 0.5 only for Hong Kong-China. In the other countries, 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.

To control overlap, the sample selection of schools for PISA adopted a modification of the approach due to Keyfitz (1951), based on Bayes Theorem. To use TIMSS and PISA in an example of the overlap control approach, suppose that *PROBT* is the TIMSS probability of selection, and *PROBP* is the required PISA probability of selection. Then a conditional probability of selection into PISA, *CPROB* is determined as follows:

4.1

$$CPROB = \begin{cases} \max\left[0, \left(\dfrac{PROBT + PROBP - 1}{PROBT}\right)\right] & \text{if the school was TIMSS selected} \\ \min\left[1, \dfrac{PROBP}{(1 - PROBT)}\right] & \text{if the school was not TIMSS selected} \\ PROBP & \text{if the school was not a TIMSS eligible school} \end{cases}$$

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

CMOS=*CPROB* × stratum sampling interval (recorded to 4 decimal places)



The PISA school sample was then selected using the line numbers created as usual (see below), 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 sample size, but this was deemed acceptable.

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 determined 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.

Identifying replacement schools

Each sampled school in the main survey was assigned two replacement schools from the sampling frame, identified as follows. For each sampled school, the schools immediately preceding and following it in the explicit stratum 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 very close to each other in the sampling frame. There were times when it was only possible to assign a single replacement school, or even none, when two consecutive schools in the sampling frame were sampled.

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 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 eligible, non-excluded students as participants.

Monitoring school sampling

For PISA 2006, it was a strong recommendation that the consortium select the school samples. This was incorporated into the 2006 procedures to alleviate the weighting difficulties caused by receiving school frame files in many different formats. France and Japan selected their own school samples for reasons of confidentiality. Sample selection was replicated by the consortium to ensure quality. All other samples were selected by and checked in detail by the consortium. All countries were required to submit sampling forms 1 (time of testing and age definition), 2 (national desired target population), 3 (national defined target population), 4 (sampling frame description), 5 (excluded schools), 7 (stratification), and 11 (school sampling frame). The consortium completed and returned the others (forms 6, 8, 9, 10, and the base form 12) for countries for which they did the sampling. Otherwise, the country also submitted these other forms for approval. Table 4.2 provides a summary of the information required on each form and the timetables (which depended on national assessment periods).

Table 4.2
Schedule of school sampling activities

Activity	Submit to Consortium	Due Date
Specify time of testing and age definition of population to be tested	Sampling form 1 – time of testing and age definition	Submit three months before the school sample is to be selected
Define national desired target population	Sampling form 2 – national desired target population	Submit three months before the school sample is to be selected
Define national defined target population	Sampling form 3 – national defined target population	Submit three months before the school sample is to be selected
Create and describe sampling frame	Sampling form 4 – sampling frame Description	Submit two months before the school sample is to be selected
Decide on schools to be excluded from sampling frame	Sampling form 5 – excluded schools	Submit two months before the school sample is to be selected
Decide how to treat small schools	Sampling form 6 – Treatment of Small schools	The Consortium will complete and return this form to the NPM about one month before the school sample is to be selected.
Decide on explicit and implicit stratification variables	Sampling form 7 – stratification	Submit two months before the school sample is to be selected
Describe population within strata	Sampling form 8 – population counts by strata	The Consortium will complete and return this form to the NPM when the school sample is sent to the NPM.
Allocate sample over explicit strata	Sampling form 9 – sample allocation by explicit strata	The Consortium will complete and return this form to the NPM about one month before the school sample is to be selected.
Select the school sample	Sampling form 10 – school sample Selection	The Consortium will complete and return this form to the NPM when the school sample is sent to the NPM.
Identify sampled schools, replacement schools and assign PISA school IDs	Sampling form 11 – school sampling frame	Submit two months before the school sample is to be selected. The Consortium will return this form to the NPM with sampled schools and their replacement schools identified and with PISA IDs assigned when the school sample is selected.
Create a school tracking form	Sampling form 12 – school tracking form	Submit within one month of the end of the data collection period

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



Once received from each country, each form 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 consortium. In cases where approval could not be granted, countries were asked to make revisions to their sample design and sampling forms.

Checks that were performed in the monitoring of each form follow. All entries were observed in their own right but those below are additional matters explicitly examined.

Sampling form 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;
- Assessment dates were checked against recorded main study (MS) assessment dates on field trial (FT) sampling forms. Differences were queried;
- If assessment end dates were close to the end of the population birth date window, NPMs were alerted not to conduct any make-up sessions beyond the date when the population births dates were valid;
- Population birth dates were checked against those recorded for the MS on the FT sampling forms. Differences were queried.

Sampling form 2: National desired target population

- 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 population numbers compared to those from PISA 2003 were queried, as were seeming trends in population numbers (increasing or decreasing) 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 2% or if it was not noted for PISA 2003;
- Calculations were verified;
- For any countries using a three-stage design, a sampling form 2 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 form 2 was needed for each region;

Sampling form 3: National defined target population

- The population figure in the first question needed to correspond with the final population figure on sampling form 2;
- Reasons for excluding schools were checked for appropriateness;
- Exclusion types and extents were compared to those recorded for PISA 2003. Differences were queried;
- Use of the UH booklet was queried;
- Exclusions for language were checked against what was recorded for the MS on the FT sampling forms. Differences were queried;
- The number and percentage of students to be excluded at the school level and whether the percentage was less than the maximum percentage allowed for such exclusions were checked;



- Calculations were verified and the overall coverage figures were assessed;
- Reasonableness of assumptions about within-school exclusions was assessed by checking previous PISA coverage tables;
- The population figures on this form were compared against the summed sampling frame enrolment. Differences were queried;
- For any countries using a three-stage design, a sampling form 3 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 form 3 was needed for each region.

Sampling form 4: Sampling frame description

- Special attention was paid 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;
- Frame sampling units were compared against those used for PISA 2003. Differences were queried.

Sampling form 5: Excluded schools

- The number of schools and the total enrolment figures, as well as the reasons for exclusion, were checked to ensure correspondence with figures reported on sampling form 3 about school-level exclusions.

Sampling form 6: Treatment of small schools

- Calculations were verified, as was the decision about whether or not a moderately small schools stratum and/or a very small schools stratum were needed.

Sampling form 7: Stratification

- 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 country's choice of explicit stratification variables was assessed. If a country was known to have school tracking, and tracks or school programmes were not among the explicit stratifiers, a suggestion was made to include this type of variable;
- Identified stratification variables were compared against those noted for the MS on the FT sampling forms. Differences were queried;
- 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;
- The sampling frame was checked to ensure that the stratification variables were available for all schools. Different explicit strata were allowed to have different implicit stratifiers;
- Any indicated student sorting variables were compared to those used in PISA 2003. Differences were queried.

Sampling form 8: Population counts by strata

- Counts on sampling form 8 were compared to counts arising from the frame. Any differences were queried and corrected as appropriate.

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



Sampling form 9: Sample allocation by explicit strata

- All explicit strata had to be accounted for on sampling form 9;
- All explicit strata population entries were compared to those determined from the sampling frame;
- 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;
- The percentage of students in the sample for each explicit stratum had to be close to the percentage in the population for each stratum (very small schools strata were an exception since under-sampling was allowed);
- 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 form 10: School sample selection

- All calculations were verified;
- Particular attention was paid to the four decimal places that were required for both the sampling interval and the random number.

Sampling form 11: School sampling frame

- The frame number of sampling units was compared to the same for PISA 2003. Differences were queried;
- NPMs were queried about whether or not they had included schools with grades 7 or 8 that could potentially have PISA students at the time of assessment;
- NPMs were queried about whether or not they had included vocational or apprenticeship, schools with only part-time students, International or foreign schools or any other irregular schools that could contain PISA students at the time of the assessment;
- The frame was checked for proper sorting according to the implicit stratification scheme and enrolment values, and the proper assignment of the measure of size value, especially for moderately small and very small schools. The accumulation of the measure of size values was also checked for each explicit stratum. This final cumulated measure of size value for each stratum had to correspond to the 'Total Measure of Size' value on sampling form 10 for each explicit stratum. Additionally, each line selection number was checked against the frame cumulative measure of size figures to ensure that the correct schools were sampled. Finally, the assignment of replacement schools and PISA identification numbers were checked to ensure that all rules laid out in the *Sampling Manual* were adhered to. Any deviations were discussed with each country and either corrected or the deviations accepted.

Sampling form 12: School tracking form

- Sampling form 12 was checked to see that the PISA identification numbers on this form matched those on the sampling frame;
- Checks were made to ensure that all sampled and replacement schools were accounted for;
- Checks were also made to ensure that status entries were in the requested format.

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



Student samples

Student selection procedures in the main study were the same as those used in the field trial. Student sampling was generally undertaken using the consortium software, *KeyQuest*, at the national centres from lists of all 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 very important that the student sample be selected from accurate, complete lists, the lists needed to be prepared not too far in advance of the testing and had to list all eligible students. It was suggested that the lists be received one to two months before testing so that the NPM would have the time to select the student samples.

Twelve countries (Chile, the Czech Republic, Germany, Iceland, Japan, Korea, Liechtenstein, Mexico, Norway, Sweden, Switzerland and Uruguay) 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 steps. The Czech Republic, Korea, Mexico, Norway, Sweden, Switzerland (only in some explicit strata), and Uruguay used the standard method of direct student sampling described here. However, Mexico also sub-sampled schools in which to do the grade sampling from its large school sample. For Iceland and Japan, the sample constituted a de facto grade sample because nearly all of the PISA eligible 15-year-olds were in the grade sampled. Germany, Liechtenstein, and Switzerland (in a second set of explicit strata) supplemented the standard 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. 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.

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 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 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 1990 (or the appropriate 12-month age span agreed upon for the 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 suggested that schools retain a copy of the list in case the NPM had to call the school with questions;
- A computer list was to be up-to-date at the time of sampling rather than 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 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 sampling software, to select the student samples unless agreed upon. Three countries (Germany, Luxembourg, and Switzerland) did not use *KeyQuest* for all or for a part of the student sample



for reasons including extra student demographic data or due to an unusual, but approved, class sampling approach for a grade option.

Preparing instructions for excluding students

PISA was a timed assessment administered in the instructional language(s) of each 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 given 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 well documented and submitted to the consortium for review before testing.

Sending the student tracking form to the school co-ordinator and test administrator

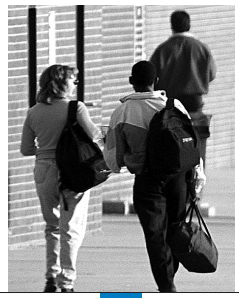
The school co-ordinator needed to know which students were sampled in order to notify them and their teachers (and parents), to update information and to identify the students to be excluded. The student tracking form was therefore sent about two weeks before the assessment session. It was recommended that a copy of the tracking form be made and kept at the national centre. Another recommendation was to have 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 show, 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 explaining the reason for exclusion. It was important that these criteria be followed strictly for the study to be comparable within and across countries. When in doubt, the student was included. The instructions for excluding students are provided in the PISA Technical Standards.

Notes

1. A student was deemed a participant 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* Chapter 17).

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



5

Translation and cultural appropriateness of the test and survey material

Introduction 86
Development of source versions 86
Double translation from two source languages 87
PISA translation and adaptation guidelines 88
Translation training session 89
Testing languages and translation/adaptation procedures 89
International verification of the national versions 91
- Vegasuite 93
- Documentation 93
- Verification of test units 93
- Verification of the booklet shell 94
- Final optical check 94
- Verification of questionnaires and manuals 94
- Final check of coding guides 95
- Verification outcomes 95

Translation and verification outcomes – national version quality 96
- Analyses at the country level 96
- Analyses at the item level 103
- Summary of items lost at the national level, due to translation, printing or layout errors 104

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



INTRODUCTION

Literature on empirical comparative research refers to translation issues as one of the most frequent problems in cross-cultural surveys. Translation errors are much more frequent than other problems, such as clearly identified discrepancies due to cultural biases or curricular differences. (Harkness, Van de Vijver and Mohler, 2003; Hambleton, Merenda and Spielberger, 2005).

If a survey is done merely to rank countries or students, this problem can be avoided somewhat since once the most unstable items have been identified and dropped, the few remaining problematic items are unlikely to affect the overall estimate of a country's mean in any significant way.

The aim of PISA, however, is to develop descriptive scales, and in this case translation errors are of greater concern. The interpretation of a scale can be severely biased by unstable item characteristics from one country to another. One of the important responsibilities of PISA is therefore to ensure that the instruments used in all participating countries to assess their students' literacy provide reliable and fully comparable information. In order to achieve this, PISA implemented strict verification procedures for translation/adaptation and verification procedures.

These procedures included:

- Development of two source versions of the instruments (in English and French);
- Double translation design;
- Preparation of detailed instructions for the translation of the instruments for the field trial and for their review for the main study;
- Preparation of translation/adaptation guidelines;
- Training of national staff in charge of the translation/adaptation of the instruments;
- Verification of the national versions by international verifiers.

DEVELOPMENT OF SOURCE VERSIONS

Part of the new test materials used in PISA 2006 was prepared by the consortium test development teams on the basis of the submissions received from the participating countries. Items were submitted by 21 different countries, either in their national language or in English. The other part of the material was prepared by the test development teams themselves in CITO, NIER, ILS, IPN and ACER. Then, all materials were circulated (in English) for comments and feedbacks to the Expert Groups and the NPMs.

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 cycles. The items developers used it as reference when developing and reviewing the items.

The French version was developed at this 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 translation issues do not become apparent until there is an attempt to translate the instruments. As in previous PISA cycles, the translation process proved to be very effective



in detecting residual errors overlooked by the test developers, and in anticipating potential translation problems. 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 National Project Managers with source material that was somewhat easier to translate or contained fewer potential translation problems than it would have had if only one source had been developed.

The final French source version was reviewed by a French domain expert, for appropriateness of the science 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 possible systematic translation errors in all or most of the national versions adapted from the French source version were conducted, using the main study item statistics from the five French-speaking countries participating in PISA 2006.

DOUBLE TRANSLATION FROM TWO SOURCE LANGUAGES

A back translation design has long been the most frequently used 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/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.

Other anticipated advantages of using two source languages in the PISA assessment included:

- 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 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) provides national reconcilers with accurate benchmarks in this respect, and that neither back translation nor double translation from a single language could provide.

Since PISA was the first major international survey using two different source languages, empirical evidence from the PISA 2000 field trial results was collected to explore the consequences of using alternative reference languages in the development phase of the various national versions of the survey materials. The outcomes of this study were reported in Chapter 5 of the *PISA 2000 Technical Report* (Adams and Wu, 2002; Grisay, 2003).

PISA 2003 main study data analyses were used to identify all items showing even minor weaknesses in the seven English-speaking countries or communities and the five French-speaking countries or communities that developed their national versions by just entering national adaptations in one of the source versions provided by the consortium (OECD 2005). Out of the 167 items used in the main study, 103 had no problems in any of the French and English versions and 29 had just occasional problems in one or two of the twelve countries. Thirteen items had weak statistics in both English and French versions but also appeared to have flaws in at least half of the participating countries. No items had weaknesses in all French versions and no flaws in any of the English versions. Some imbalance was observed for nine items. In fact the overall percentage of weak items was very similar in both the group of English testing countries and the group of French testing countries.

Empirical evidence on the quality of the national versions obtained was collected by analysing the proportion of weak items in each national data set, based again on the PISA 2003 main study item analyses, and using the same criteria for identifying weak items as for the source versions.

Among countries that used double translation from just one of the source versions, 12.5% of the items were considered weak, compared to 8.5% in countries that used both source versions in their translations, and 6.5% in countries whose versions were derived directly from either the English or French source version. This seems to indicate that double-translation from only one source language may be less effective than double translation from both languages, confirming a trend already observed in PISA 2000.

Due to these results, a double translation and reconciliation procedure using both source languages was again recommended in PISA 2006.

PISA TRANSLATION AND ADAPTATION GUIDELINES

The *PISA Translation and Adaptation Guidelines* as prepared in previous PISA studies were revised to include more detailed advice on translation and adaptation of science materials, and additional warnings about



common translation errors identified during the verification of the PISA 2003 materials and the development of the French source version. These guidelines were revised with a view to obtaining a document that would be relevant to any PISA cycle. The guidelines included:

- Instructions for national version(s): According to the PISA technical standards, students should be tested in the language of instruction used in their school. Therefore, the NPMs of multilingual countries were requested to develop as many versions of the test instruments as there were languages of instruction used in the schools included in their national sample. Cases of minority languages used in only a very limited number of schools could be discussed with the sampling referee to decide whether such schools could be excluded from the target population without affecting the overall quality of the data collection;
- Instructions on double or single translation: Double-translation was required for the tests, questionnaires and for the optional questionnaires, but not for the manuals and other logistic material;
- Instructions on recruitment and training: It was suggested, in particular, that translated material and national adaptations deemed necessary be submitted for review and approval to a national expert panel composed of domain specialists;
- Description of the PISA translation procedures: It was required that national versions be developed through double translation and reconciliation with the source material. It was recommended that one independent translator would use the English source version and that the second would use the French version. In countries where the NPM had difficulty appointing competent translators from French/English, double translation from English/French only was considered acceptable according the *PISA Technical Standards 5.1 and 5.2*.

Other sections of the *PISA Translation and Adaptations Guidelines* were intended for use by the national translators and reconcilers and included:

- Recommendations to avoid common translation traps. An extensive section giving detailed examples on problems frequently encountered when translating assessment materials, and advice on how to avoid them;
- Instructions on how to adapt the test material to the national context. This listed a variety of rules identifying acceptable/unacceptable national adaptations and including specific notes on translating mathematics and science material;
- Instructions on how to translate and adapt the questionnaires and manuals to the national context;
- The check list used for the verification of PISA material.

After completion of the field trial, an additional section of the Guidelines was circulated to NPMs, as part of their *Main Study NPM Manual*, together with the revised materials to be used in the main study. This section contained instructions on how to revise their national version(s).

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 September 2004 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.

TESTING LANGUAGES AND TRANSLATION/ADAPTATION PROCEDURES

NPMs had to identify the testing languages according to instructions given in the *Sampling Manual* 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 and in December 2004 the NPMs were asked to either confirm that the information given was accurate or to notify which changes were made.

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 PISA 2000 and 2003 that high quality translations and high levels of equivalence in the functioning of items were 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, having 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 should reduce cross-countries differences due to translation effects.

Table 5.1 lists countries that shared a common version of test items with national adaptations.

Table 5.1
Countries sharing a common version with national adaptations

Language	Countries	Collaboration
Arabic	Jordan and Qatar	Jordan developed a version in which Qatar introduced adaptations (Field trial only).
Chinese (c)	Hong Kong-China, Macao-China and Chinese Taipei	Commonly developed Chinese version: Two single translations produced by 2 countries and reconciliation by the third one
Dutch	Netherlands, Belgium	Belgium (Flemish Community) introduced adaptations in the verified Dutch version
English	Australia, Canada, Hong Kong-China, Ireland, Qatar, New Zealand, Scotland, Sweden, United Kingdom, USA	Adaptations introduced in the English source version
French	Belgium, Canada, France, Luxembourg, Switzerland	Adaptations introduced in the French source version
German	Austria, Belgium, Germany, Italy, Luxembourg, Switzerland	Adaptations introduced in a commonly developed German version
Hungarian	Hungary, Serbia, Slovak Republic, Romania	For their Hungarian versions, Serbia and the Slovak Republic introduced adaptations in the verified version from Hungary
Italian	Italy, Switzerland, Slovenia	Switzerland (Canton Ticino) and Slovenia introduced adaptations in the verified version from Italy
Russian	Russia, Azerbaijan, Estonia, Kyrgyzstan, Latvia, Lithuania	Adaptations introduced in the verified version from Russia or Kyrgyzstan[1]
Polish	Poland, Lithuania	For its Polish version, Lithuania introduced adaptations in the verified version from Poland
Slovene	Slovenia, Italy	Use of Slovene version in Italy
Portuguese	Portugal, Macao-China	Macao-China introduced adaptations in the verified version from Portugal
Spanish	Mexico, Argentina, Uruguay	Argentina and Uruguay introduced adaptations in the verified version from Mexico
Swedish	Sweden, Finland	For its Swedish version, Finland introduced adaptations in the verified version from Sweden

1. Kyrgyzstan first adapted the version from Russia, then in the Main Study, due to time constraints some countries adapted the verified version from Kyrgyzstan.

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



Additionally Chile and Colombia collaborated with each providing one translation (one from English and one from French) to the other. This however did not lead to a common version as each country performed the reconciliation separately.

Table 5.2 summarises the translation procedures as described in the country *Translation Plans.*

Table 5.2
PISA 2006 translation/adaptation procedures

Procedures	Number of national versions
Use one of the source versions with national adaptations	15
Use of a commonly developed version with national adaptations	7
Use of a borrowed verified version with or without national adaptations	19
Double translation from both source versions	16
Double translation from English or French source with cross-checks against the other source version	12
Double translation from English source only	15
Alternative procedures	3

A total of 87 national versions of the materials were used in the PISA 2006 main study, in 44 languages, The languages were: Arabic (4 versions), Azeri, Bahasa Indonesian, Basque, Bulgarian, Catalan, Chinese (3 versions), Croatian, Czech, Danish, Dutch (2 versions); Estonian, English (10 versions), Finnish, French (5 versions), Galician, German (6 versions), Greek, Hebrew, Italian (3 versions), Hungarian (3 versions), Icelandic, Irish, Japanese, Korean, Kyrgyz, Latvian, Lithuanian, Norwegian (Bokmål), Norwegian (Nynorsk), Polish (2 versions), Portuguese (3 versions), Romanian, Russian (5 versions), Serb Ekavian variant, Serb Yekavian variant, Slovak, Slovene (2 versions), Spanish (6 versions), Swedish (2 versions), Thai, Turkish, Uzbek and Valencian.

International verification (described in section below) occurred for 78 national versions out of the 87 used in the main study.

International verification was not implemented when:

- A testing language was used for minorities that make less than 5% of the target population as for Irish, Hungarian (Serbia and Romania), Polish (Lithuania), Valencian. In that case the verification is organised at the national level;
- When countries borrowed a version that had been verified at the national level without making any adaptations as for German (Belgium), English (Sweden), Portuguese (Macao-China), Slovene (Italy), Italian (Slovenia).

INTERNATIONAL VERIFICATION OF THE NATIONAL VERSIONS

As in PISA 2003, one of the most important quality control procedures implemented to ensure high quality standards in the translated assessment materials consisted in having an independent team of expert verifiers, appointed and trained by the consortium, verify each national version against the English and French source versions.

Two verification co-ordination centres were established. One was at ACER in Melbourne (for national adaptations used in the English-speaking countries). The second one was at cApStAn, which has been involved in preparing the French source versions of the PISA materials and verifying non-English national versions since PISA 2000.



The consortium undertook international verifications of all national versions in languages used in schools attended by more than 5% of the country's target population. For languages used in schools attended by 5% or less minorities, international-level verification was deemed unnecessary since the impact on the country results would be negligible, and verification of such languages was more feasible at national level.

For a few minority languages, national versions were only developed (and verified) in the main study phase. This was considered acceptable when a national centre had arranged with another PISA country to borrow its main study national version for their minority (*e.g.* adapting the Swedish version from Sweden for Swedish schools in Finland, the Russian version from the Russian Federation for Russian schools in Latvia), or when the minority language was considered to be a variant that differed only slightly from the main national language (*e.g.* Nynorsk in Norway).

English- or French-speaking countries or communities were allowed to only submit national adaptation forms for verification. This was also considered acceptable, since these countries used national versions that were identical to the source version except for the national adaptations.

The main criteria used to recruit translators to lead the verification 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;
- 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;
- Familiarity with the main domain assessed (in this case, science);
- A good level of computer literacy;
- As far as possible, experience as teachers and/or higher education degrees in psychology, sociology or education.

As a general rule, the same verifiers were used for homolingual versions (*i.e.* the various national versions from English, French, German, Italian and Dutch-speaking countries or communities). However, the Portuguese language differs significantly from Brazil to Portugal, and the Spanish language is not the same in Spain and in Latin American countries, so independent native translators had to be appointed for those countries.

In a few cases, both in the field trial and the main study verification exercises, the time constraints were too tight for a single person to meet the deadlines, and additional verifiers had to be appointed and trained.

Verifier training sessions were held prior to the verification of both the field trial and the main study materials. Attendees received copies of the PISA information brochure, *Translation Guidelines*, the English and French source versions of the material and a *Verification Check List* developed by the consortium. The training sessions focused on:

- Presenting verifiers with PISA objectives and structure;
- Familiarising them with the material to be verified;
- Reviewing and extensively discussing the *Translation Guidelines* and the *Verification Check List*;



- Conducting hands-on exercises on specially adapted target versions;
- Arranging for schedules and for dispatch logistics;
- Security requirements.

The verification procedures were improved and strengthened in a number of respects in PISA 2006, compared to previous rounds.

VegaSuite

- For the main study phase, cApStAn developed a web-based upload-download platform known as Vegasuite for file exchange and archiving, to facilitate and automate a number of processes as PISA verification grew in size. This development was well received by NPMs and verifiers.

Documentation

- Science textbooks selected and sent by the National Centres of the participating countries were distributed to verifiers. These textbooks, from the grades attended by most 15-year-olds in the respective countries, were used by verifiers as reference works because the NPMs deemed them representative of the level/register of scientific language familiar to 15-year-olds students in their country.

Verification of test units

- As in previous rounds, verifiers entered their suggested edits in MS Word files, using the track changes mode, to facilitate the revision of verified materials by the NPMs (who could directly accept or refuse the edits proposed). But for all issues deemed likely to affect equivalence between source version(s) and target version, verifiers were also instructed to insert a comment in English at the appropriate location in the test adaptation spreadsheet (TAS). This was to formalise the process by which a) the consortium verification referee is informed of such issues and can liaise as needed with the test developers; b) if there is disagreement with the National Centre (NC), 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 double-checked at final optical check (FOC) phase. In previous verification rounds, this process took place in a less structured way;
- Following the field trial verification, cApStAn analysed the comments made by verifiers in the TAS, leading to a classification using a relatively simple set of categories. The purpose was 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;
- For the main study phase, an innovation in the TAS was that verifiers used a scroll-down menu to categorize issues in one of 8 standardised verification intervention categories: added information, missing information, layout/visual issues, grammar/syntax, consistency, register/wording, adaptation, and mistranslation. a comments column allowed verifiers to explain their intervention with a back-translation or description of the problem;
- For the main study phase, the consortium's FT to MS revisions were listed in the TAS. For such revisions, the drop-down menu in the verifier intervention column was dichotomous: the verifier had the choice between OK (implemented) or NOT OK (overlooked). In case the change was partially implemented, the verifier would select OK (implemented) and comment on the issue in the verifier comment column. This procedure ensured that the verifier would check the correct implementation of every single FT to MS change.



- Another innovation for the main study phase: at the top of each TAS was a list of recurring terms or expressions that occur throughout the test material, such as Circle Yes or No. Verifiers were asked to keep track of across-unit consistency for these expressions and, at the end of the verification of a full set of units, to choose, in the verifier intervention column, from three options in a drop-down menu: "OK"; "Some inconsistencies"; or "Many inconsistencies".

Verification of the booklet shell

- This had not been a separate component in previous rounds. The booklet shell was dispatched together with a booklet adaptation spreadsheet (BAS) and verified following the same procedure as the test units. This proved very helpful for both the NCs' and the verifiers' work organisation, because it resulted in timely verification of sensitive issues. In previous rounds, the booklet shell was often verified on a rush basis when camera-ready instruments were submitted for final optical check (FOC).

Final optical check

- As in previous rounds, test booklets and questionnaire forms were checked page-by-page as regards correct item allocation, layout, page numbering, item numbering, graphic elements, item codes, footers, etc (classic FOC). As in previous rounds, this phase continues to prove essential in spotting residual flaws, some of which could not have been spotted during the item pool verification;

- An innovation in PISA 2006 was the systematic verification of whether key corrections resulting from the first verification phase were duly implemented. All TAS and BAS containing key corrections were thus also returned to each country with recommendations to intervene on any residual key correction that was overlooked or incorrectly implemented. A similarly annotated QAS was also returned in cases where corrections had been flagged by the consortium staff in charge of reviewing questionnaires, thus requesting follow-up at FOC stage. Note that in PISA 2000 and PISA 2003, National Centres were given the final responsibility for all proposed corrections and edits. Although the FOC brief previously included performing random checks to verify whether crucial corrections proposed during Item Pool verification were duly implemented, in practice this was made difficult by the uncertainty on whether the National Centre had accepted, rejected or overlooked corrections made by the verifier. With the systematic verification of key corrections labelled by the consortium, it was possible to have a quantitative and systematic record of implementation of crucial corrections;

Verification of questionnaires and manuals

- As in PISA 2003, NPMs were required to have their questionnaire adaptation spreadsheet (QAS) and manual adaptaton spreadsheet (MAS) approved by consortium staff before submitting them for verification along with their translated questionnaires and manuals;

- The procedure proved to be effective for questionnaires: the instructions to the verifiers were straightforward and the instruments submitted to their scrutiny had already been discussed extensively with consortium staff by the time they had to verify them. Verifiers were instructed to refrain from discussing agreed adaptations unless the back translation into English of the agreed adaptation inadequately conveyed its meaning, in which case the consortium might have unknowingly approved an inappropriate adaptation;

- A significant improvement in PISA 2006 was that the QAS contained entries for all parts of the questionnaires, including notes and instructions to respondents;



- In the case of manuals, verification continued to be challenging in PISA 2006 because of the greater freedom that countries had in adapting these instruments. Following cApStAn's recommendation after the field trial, it was decided to limit the verification of manuals for the main study to a number of key components. The usefulness and effectiveness of this process remains marginal.

Final check of coding guides

- As in PISA 2003, a verification step was added at the main study phase for the coding guides, to check on the correct implementation of late changes in the scoring instructions introduced by the consortium after the NPM coding seminar. Verifiers checked the correct implementation of such edits. These edits had been integrated into the post-FOC TAS of countries for which the verification was over and in the standard TAS of other countries;
- In line with the innovation for PISA 2006 concerning key corrections, the final check of coding guides included a check on the correct implementation of key corrections located in the scoring rubrics, which had been left pending at booklet FOC stage.

Verification outcomes

In previous cycles, the verification reports contained qualitative information about the national versions and illustrative examples of typical errors encountered by the verifiers. In the PISA 2006 main study, 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 in the next PISA cycle, could be instrumental in revising items for the main study but would also give valuable information on how to avoid such problems in further cycles.

It 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 for reading difficulty).

As in previous PISA data collections, the verification exercise proved to be an essential mechanism for ensuring quality even though the national versions were generally found to be of high quality in terms of psychometric equivalence. In virtually all versions, the verifiers identified errors that would have seriously affected the functioning of specific items – mistranslations, omissions, loan translations or awkward expressions, incorrect terminology, poor rendering of graphics or layout, errors in numerical data, grammar and spelling errors.

Link material raised a concern again – in a larger than expected number of countries, it proved to be somewhat difficult to retrieve the electronic files containing the final national version of the materials used in the PISA 2003 main study, from which the link items had to be drawn. The verification team performed a litmus check (convergence check on a sample of link units submitted by the countries versus PISA 2003 main study archive) to determine whether the link units submitted were those actually used in the PISA 2003 test booklets. In a number of cases, the verification team or the consortium had to assist by providing the correct national versions from their own central archives.

To prevent this type of problem in future studies, the central archive at ACER was improved to host copies of all final national versions of the materials used in PISA 2006.



TRANSLATION AND VERIFICATION OUTCOMES – NATIONAL VERSION QUALITY

Analyses at the country level

One way to analyse the quality of a national version consists of analysing the item-by-country interaction coefficient. As the cognitive data have been scaled with the Rasch model for each country and for many languages (see Chapter 9), the relative difficulty of an item for a language within a country can be denoted δ_{ijk}, with *i* denoting the item, *j* denoting the language and *k* denoting the country. Further, each item can also be characterised by its international relative difficulty, denoted $\delta_{i\bullet\bullet}$, computed on a student random sample of equal size from all OECD country samples.

As both the national and international item calibrations were centred at zero, the mean of the δ_{ijk}, for any language *j* within a country *k* is equal to zero. In other words:

5.1

$$\sum_{i=1}^{I}\delta_{ijk} = 0 \quad \text{for all } j \text{ and } k$$

The item-by-country interaction is defined as the difference between any δ_{ijk} and its corresponding international item difficulty $\delta_{i\bullet\bullet}$. Therefore, the sum (and consequently the arithmetic mean) of the item-by-country interaction for a particular language within a country is equal to zero. Indeed,

5.2

$$\sum_{i=1}^{I}\left(\delta_{ijk} - \delta_{i..}\right) = \sum_{i=1}^{I}\delta_{ijk} - \sum_{i=1}^{I}\delta_{i..} = 0$$

As summary indices of item-by-country interaction for each language in a country we use the mean absolute deviation;

5.3

$$MAD_{jk} = \frac{1}{I}\sum_{i=1}^{I}\left|\delta_{ijk} - \delta_{i\bullet\bullet}\right|$$

and the root mean squared error

5.4

$$RMSE_{jk} = \sqrt{\frac{1}{I}\sum_{i-1}^{I}\left(\delta_{ijk} - \delta_{i\bullet\bullet}\right)^2}$$

and a chi-square statistic equal to;

5.5

$$X^2 = \sum_{i=1}^{I}\frac{\left(\delta_{ijk} - \delta_{i\bullet\bullet}\right)^2}{\mathrm{var}\left(\delta_{ijk}\right)}$$

As the sets item-by-country interactions by language and country, have a mean of zero, the mean of the absolute values is equal to the mean deviation and the root mean squared error is equal to the standard deviation of the item-by-country interactions.

A few science items were deleted at the national level (*i.e. S447Q02*, *S447Q03*, *S465Q04*, *S495Q04*, *S519Q01*, *S131Q04T*, *S268Q02T*, *S437Q03*, *S466Q01*, *S519Q03*, and *S524Q07*). To ensure the comparability of the analyses reported below, these items were removed from the science item parameter database and the national and international parameter estimates of the 92 remaining science items were re-centred on zero for each language and country.



Table 5.3 [Part 1/2]
Mean deviation and root mean squared error of the item by country interactions for each version

		Language	Absolute Value Mean or Mean deviation	RMSE or STD	X^2
OECD	Australia	English	0.24	0.29	223.99
	Austria	German	0.25	0.32	148.33
	Belgium	Dutch	0.28	0.34	173.36
	Belgium	French	0.25	0.31	110.95
	Belgium	German	0.25	0.32	59.60
	Canada	English	0.24	0.30	248.14
	Canada	French	0.20	0.28	118.04
	Czech Republic	Czech	0.25	0.32	156.73
	Denmark	Danish	0.22	0.30	133.23
	Finland	Finnish	0.34	0.43	235.97
	Finland	Swedish	0.38	0.51	80.94
	France	French	0.34	0.42	274.92
	Germany	German	0.25	0.31	142.98
	Greece	Greek	0.30	0.38	213.42
	Hungary	Hungarian	0.32	0.41	233.67
	Iceland	Icelandic	0.30	0.37	167.13
	Ireland	English	0.29	0.39	206.61
	Italy	German	0.30	0.38	110.40
	Italy	Italian	0.24	0.29	253.40
	Japan	Japanese	0.40	0.51	405.92
	Luxembourg	French	0.25	0.32	67.43
	Luxembourg	German	0.26	0.32	128.64
	Mexico	Spanish	0.31	0.40	580.70
	Netherlands	Dutch	0.30	0.39	217.46
	New Zealand	English	0.27	0.33	163.26
	Norway	Norwegian	0.23	0.30	130.45
	Poland	Polish	0.25	0.32	162.04
	Portugal	Portuguese	0.29	0.36	194.93
	Korea	Korean	0.42	0.55	433.22
	Slovak Republic	Hungarian	0.38	0.48	65.40
	Slovak Republic	Slovak	0.27	0.33	157.42
	Spain	Basque	0.37	0.47	136.18
	Spain	Catalan	0.28	0.35	103.32
	Spain	Galician	0.27	0.34	59.07
	Spain	Spanish	0.23	0.28	202.13
	Sweden	Swedish	0.23	0.29	121.16
	Switzerland	French	0.22	0.29	104.20
	Switzerland	German	0.25	0.31	188.76
	Switzerland	Italian	0.30	0.38	65.26
	Turkey	Turkish	0.32	0.41	247.18
	United Kingdom	English	0.29	0.36	291.11
	United Kingdom	Welsh	0.38	0.48	87.40
	United States	English	0.26	0.31	154.83



Table 5.3 [Part 2/2]
Mean deviation and root mean squared error of the item by country interactions for each version

		Language	Absolute Value Mean or Mean deviation	RMSE or STD	X^2
Partners	Argentina	Spanish	0.27	0.35	157.96
	Azerbaijan	Azeri	0.72	0.96	1115.60
	Azerbaijan	Russian	0.58	0.79	236.88
	Brazil	Portuguese	0.32	0.43	365.22
	Bulgaria	Bulgarian	0.29	0.38	209.40
	Chile	Spanish	0.26	0.32	166.02
	Colombia	Spanish	0.32	0.40	213.79
	Croatia	Croatian	0.30	0.40	225.32
	Estonia	Estonian	0.37	0.48	285.39
	Estonia	Russian	0.35	0.44	139.65
	Hong Kong-China	Chinese	0.45	0.56	418.56
	Indonesia	Indonesian	0.48	0.64	829.06
	Israel	Arab	0.41	0.51	156.82
	Israel	Hebrew	0.36	0.45	265.56
	Jordan	Arab	0.41	0.54	495.76
	Kyrgyzstan	Kyrgyz	0.62	0.79	526.08
	Kyrgyzstan	Russian	0.38	0.49	188.29
	Kyrgyzstan	Uzbek	0.64	0.79	238.67
	Latvia	Latvian	0.32	0.42	220.49
	Latvia	Russian	0.34	0.42	148.36
	Liechtenstein	German	0.36	0.46	76.65
	Lithuania	Lithuanian	0.37	0.47	323.31
	Lithuania	Russian	0.42	0.52	79.04
	Macao-China	Chinese	0.39	0.51	345.97
	Macao-China	English	0.46	0.57	155.65
	Montenegro	Montenegrin	0.37	0.45	291.95
	Qatar	Arab	0.47	0.57	425.06
	Qatar	English	0.45	0.58	241.25
	Romania	Hungarian	0.49	0.67	98.69
	Romania	Romanian	0.33	0.42	263.34
	Russian Federation	Russian	0.34	0.42	281.31
	Serbia	Hungarian	0.46	0.59	69.03
	Serbia	Serbian	0.30	0.40	233.18
	Slovenia	Slovenian	0.31	0.39	250.28
	Chinese Taipei	Chinese	0.51	0.66	839.30
	Thailand	Thai	0.38	0.48	385.94
	Tunisia	Tunisian	0.39	0.50	360.92
	Uruguay	Spanish	0.25	0.33	159.98

Country interactions for each language version are shown in Table 5.3. The six national versions with the highest mean deviation are:

- The Azeri version from Azerbaijan;
- The Uzbek version from Kyrgyzstan;
- The Kyrgyz version from Kyrgyzstan;
- The Russian version from Azerbaijan;
- The Hungarian version from Romania;
- The Chinese version from Chinese Taipei.

In a large number of countries with more than one language, the mean deviations of the different national versions are very similar. For instance, in Belgium, the mean deviations are respectively equal to 0.28, 0.25 and 0.25 for the Flemish version, the French version and the German version. In Estonia, they are respectively equal to 0.35 and 0.37 for the Estonian version and the Russian version. In Qatar, the English version and the Arabic version have a mean deviation of 0.45 and 0.47 respectively.



However, the mean deviations are quite different in a few countries. In Azerbaijan and in Kyrgyzstan, the mean deviation of the Russian version is substantially lower than the other national versions. The Hungarian versions used in Serbia, Romania and in the Slovak Republic present a larger mean deviation than the other national versions.

These results seem to indicate two sources of variability: the country and the language. The following tables present the correlations between the national version item parameter estimates for a particular language as well as the correlations between these item parameter estimates and the international item parameter estimates. If a language effect was suspected, then the within language correlations would be higher than the correlations with the international item parameter estimates.

Table 5.4
Correlation between national item parameter estimates for Arabic versions

	Israel	Jordan	Qatar	International Item Parameter
Israel				0.82
Jordan	0.84			0.82
Qatar	0.84	0.82		0.81
Tunisia	0.83	0.77	0.84	0.83

Table 5.5
Correlation between national item parameter estimates for Chinese versions

	Hong Kong-China	Macao-China	International Item Parameter
Hong Kong-China			0.82
Macao-China	0.94		0.85
Chinese Taipei	0.81	0.88	0.75

Table 5.6.
Correlation between national item parameter estimates for Dutch versions

	Belgium	International Item Parameter
Belgium		0.93
Netherlands	0.94	0.92

Table 5.7
Correlation between national item parameter estimates for English versions

	Australia	Canada	Great Britain	Ireland	Macao-China	New Zealand	Qatar	International Item Parameter
Australia								0.95
Canada	0.96							0.95
Great Britain	0.94	0.93						0.93
Ireland	0.92	0.93	0.96					0.92
Macao-China	0.77	0.80	0.80	0.79				0.80
New Zealand	0.98	0.95	0.94	0.91	0.77			0.94
Qatar	0.76	0.74	0.77	0.73	0.71	0.74		0.78
United States	0.97	0.96	0.94	0.91	0.78	0.95	0.81	0.94

Table 5.8
Correlation between national item parameter estimates for French versions

	Belgium	Canada	Switzerland	France	International Item Parameter
Belgium					0.94
Canada	0.95				0.95
Switzerland	0.97	0.96			0.95
France	0.94	0.90	0.94		0.89
Luxembourg	0.94	0.93	0.95	0.90	0.95



Table 5.9
Correlation between national item parameter estimates for German versions

	Austria	Belgium	Switzerland	Germany	Italy	Liechtenstein	International Item Parameter
Austria							0.95
Belgium	0.96						0.95
Switzerland	0.97	0.96					0.95
Germany	0.98	0.96	0.97				0.95
Italy	0.97	0.95	0.96	0.95			0.93
Liechtenstein	0.93	0.92	0.97	0.94	0.92		0.92
Luxembourg	0.96	0.96	0.97	0.97	0.96	0.93	0.95

Table 5.10
Correlation between national item parameter estimates for Hungarian versions

	Hungary	Romania	Serbia	International Item Parameter
Hungary				0.92
Romania	0.83			0.79
Serbia	0.89	0.81		0.85
Slovak Republic	0.93	0.80	0.87	0.89

Table 5.11
Correlation between national item parameter estimates for Italian versions

	Italy	International Item Parameter
Italy		0.95
Switzerland	0.95	0.92

Table 5.12
Correlation between national item parameter estimates for Portuguese versions

	Brazil	International Item Parameter
Brazil		0.88
Portugal	0.87	0.94

Table 5.13
Correlation between national item parameter estimates for Russian versions

	Azerbaijan	Estonia	Kyrgyzstan	Lithuania	Latvia	International Item Parameter
Azerbaijan						0.65
Estonia	0.76					0.89
Kyrgyzstan	0.81	0.88				0.86
Lithuania	0.79	0.89	0.84			0.85
Latvia	0.76	0.95	0.89	0.89		0.89
Russia	0.80	0.96	0.92	0.90	0.95	0.89

Table 5.14
Correlation between national item parameter estimates for Spanish versions

	Argentina	Chile	Colombia	Spain	Mexico	International Item Parameter
Argentina						0.93
Chile	0.94					0.94
Colombia	0.92	0.91				0.90
Spain	0.93	0.92	0.90			0.96
Mexico	0.92	0.92	0.93	0.90		0.91
Uruguay	0.94	0.93	0.91	0.93	0.93	0.93

Table 5.15
Correlation between national item parameter estimates for Swedish versions

	Finland	International Item Parameter
Finland		0.90
Sweden	0.94	0.95

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



For the various Arabic-, Dutch-, German- and Spanish-language versions, the within-language correlations do not differ substantially from the correlations between the national and the international item parameter estimates.

The correlations within the Chinese-language versions are substantially higher than their respective correlations with the international item parameter estimates. This might reflect a language effect or a cultural effect, included a curriculum effect.

The correlations within English-language versions show an interesting pattern. First of all, the correlations between parameter estimates for the English-language versions from the two countries where English is a minority language (*i.e.* Qatar and Macao-China) are lower than the respective correlations for the countries where English is the majority language. Further, the English-speaking countries seem to form two groups: Great Britain and Ireland in the first group and the others in the second group. Within a group, the correlations between the national versions are higher than their correlations with the international items parameter estimates while between group, the correlations appears to be equal or lower than the correlations with the international item parameter estimates.

The correlation pattern of the French-language versions outlines an increase of the correlation for France. While the item parameter estimates for France correlate at 0.89 with the international item parameter estimates, they correlate at 0.94 with the item parameter estimates of the French-language version of Belgium and Switzerland.

The Hungarian-language versions from Romania, Serbia and the Slovak Republic better correlate with the national version of Hungary than with the international item parameter estimates. The same phenomenon is also observed with the Russian-language versions. For any country that tested some part of their population in the Russian language, the item parameter estimates correlate better with the item parameter of Russia than with the international item parameter estimates.

Table 5.16
Correlation between national item parameter estimates within countries

		Language 1	Language 2	Correlation
OECD	Belgium	Dutch	French	0.89
		Dutch	German	0.89
		French	German	0.90
	Canada	English	French	0.92
	Switzerland	French	German	0.91
		French	Italian	0.93
		German	Italian	0.92
	Spain	Basque	Catalan	0.87
		Basque	Galician	0.89
		Basque	Spanish	0.91
		Catalan	Galician	0.93
		Catalan	Spanish	0.94
		Galician	Spanish	0.95
	Finland	Finish	Swedish	0.86
	Slovak Republic	Slovak	Hungarian	0.87
	United Kingdom	English	Welsh	0.89
Partners	Azerbaijan	Russian	Azeri	0.77
	Estonia	Estonian	Russian	0.85
	Israel	Hebrew	Arabic	0.81
	Kyrgyzstan	Uzbek	Kyrgyz	0.90
		Kyrgyz	Russian	0.84
		Uzbek	Russian	0.82
	Lithuania	Russian	Lithuanian	0.78
	Latvia	Russian	Latvian	0.89
	Macao-China	English	Chinese	0.78
	Qatar	English	Arabic	0.91
	Romania	Romanian	Hungarian	0.83
	Serbia	Serbian	Hungarian	0.80

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



Among all these correlation matrices, it appears that the matrix for the English version is the most instructive. It seems that the cultural effects or the curriculum effect are more important than the language effects. To confirm this hypothesis, correlations have been computed between national versions within countries. If the hypothesis is correct, then the correlation between the national versions within a country should be higher than the correlation between national versions within languages.

Based on Table 5.16, a few observations can be made:

- Where a country has borrowed a version from another country or if countries have cooperated to produce a common version, the national item parameter estimates better correlates within the language than within the country. For instance, the Belgian-Flemish version shows a higher correlation with the Dutch version than with the Belgian-French version. This is also the case for the Swedish version in Finland;
- As the correlation between the national item parameter estimates of the two versions in Canada (English and French) is lower than most of the correlations for the English version and the French version, one cannot dismiss some effect of the language;
- The correlation between the Arabic-language Qatari version the three national versions in Kyrgyzstan seem to reflect a curriculum effect. While the English-language version and the Arabic-language version in Qatar correlate respectively at 0.78 and 0.80 with the international item parameter estimates, they correlate 0.91 with each other. Also, while the Kyrgyz-language version and the Uzbek-language version correlate respectively 0.73 and 0.69 with the international item parameter estimates, they correlate 0.90 with each other;
- On the other hand, for Macao-China, the correlation between different language versions is not higher than the correlation with the international item parameter estimates. This could reflect some translation or equivalence issues.

To further disentangle the effects, variance decomposition models of the absolute value of the item-by-country interaction have been performed.

Table 5.17 shows the results of a nested analysis of variance of the absolute value of the item by country interaction of the 92 science items, which includes those countries with multiple language versions and the multiple versions for each country are treated as nested within the country.

Table 5.17
Variance estimate

	Variance estimates	Variance estimates without Azerbaijan and Kyrgyzstan
Country	0.010	0.003
Version (Country)	0.003	0.002
Residual	0.090	0.069

The country variance estimate is substantially higher than the version-within-country variance estimate. However, as already mentioned, Azerbaijan and Kyrgyzstan national versions had high mean deviations and low correlation with the international item parameter estimates. Without these two countries, the country variance estimates and the version-within-country variance estimates are quite similar. In each case, the most important variance component is the residual. To better understand the meaning of this residual, the unit and the item effects were included in the decomposition of the item by country interactions.

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



Table 5.18
Variance estimates

Effect	Variance estimate
Test unit	0.00095
Item within unit	0.00279
Country	0.00317
Country by test unit	0.00132
Country by item within unit	0.00632
Version within country	0.00226
Version within country by test unit	0.00002
Version within country by item within unit	0.05783

Table 5.18 presents the variance decomposition with four main effects, (i) country, (ii) language version nested in country, (iii) test unit and (iv) item embedded nested unit. Science units with a single item and countries with only one national version were therefore removed from the database. It therefore remains 17 countries, 38 countries representing 23 languages, 87 items embedded in 31 units.

The first two variance estimates are a test effect. They both reflect that some units, on average, have more item-by-country interactions than others and more particularly that some items have on average larger item-by-country interactions than others. The next section of this chapter is devoted to analyses at the item and the unit levels.

The second set of variance estimates provided in Table 5.18 are cultural or curriculum effects. The country effect,in Table 5.3 confirms that some countries have on average, larger item-by-country interactions than others. The interaction between the country and the unit reflects that some units are relatively easier or more difficult for the different national versions within a country. Finally, the interaction between the country and the item, which is the largest effect after the residual effect, confirms that some items appear to be relatively easier or more difficult for the different versions within a country. As it is quite unlikely that a translation problem occurs for the same unit or for the same item in each national version within a country, and further has the same effect, these two interactions can therefore be considered as cultural effect or curriculum effect.

Finally, the last three effects show equivalence problems, translation problems or a cultural and/or curriculum, linguistic effect. Indeed, in countries like Belgium, there are no national curricula, as education is a responsibility of the linguistic communities.

About 75% of the variability of the item-by-country interaction is at the lowest level, *i.e.* the interaction between the item and the national version.

Analyses at the item level

On average across countries, a unit has an item-by-country interaction of 0.34. It ranges from 0.25 for unit *S447* to 0.44 for unit *S493*. None of the unit characteristics (*i.e.* application area, original language of the item) are related to the unit item-by-country interaction average.

The average item-by-country interaction at the item level ranges from 0.19 (*S498Q04*) to 0.53 (*S458Q01*). The item format and the item focus do not affect the item-by-country interaction. average but the assessed competency is significantly associated with the item-by-country interaction. Items designed for assessing *using scientific evidence* on average present a mean item-by-country interaction of 0,33, items for *identifying scientific issues* a mean of 0.33 and items for *explaining phenomena scientifically* a mean of 0.36.

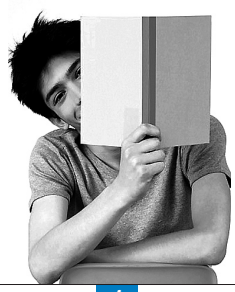


Summary of items lost at the national level, due to translation, printing or layout errors

In all cases when large DIF or other serious flaws were identified in specific items, the NPMs were asked to review their translation of the item and to provide the consortium with possible explanations.

As often happens in this kind of exercise, no obvious translation error was found in a majority of cases. However, some residual errors could be identified, that had been overlooked by both the NPMs and the verifier. Out of the 179 mathematics, reading and science items, 28 items were omitted in a total of 38 occurrences for the computation of national scores for the following reasons:

- Mistranslations or confusing translations: 20 items;
- Poor printing: 13 items;
- Layout issues: one item;
- Omission of key words: three items;
- Problematic item since PISA 2000: one item.



6

Field operations

Overview of roles and responsibilities 106
- National project managers 106
- School co-ordinators 107
- Test administrators 107
- School associates 108

The selection of the school sample 108
Preparation of test booklets, questionnaires and manuals 108
The selection of the student sample 109
Packaging and shipping materials 110
Receipt of materials at the national centre after testing 110
Coding of the tests and questionnaires 111
- Preparing for coding 111
- Logistics prior to coding 113
- Single coding design 115
- Multiple coding 117
- Managing the coding process 118
- Cross-national coding 120
- Questionnaire coding 120

Data entry, data checking and file submission 120
- Data entry 120
- Data checking 120
- Data submission 121
- After data were submitted 121

The main study review 121

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



OVERVIEW OF ROLES AND RESPONSIBILITIES

PISA was implemented in each country by a National Project Manager (NPM) who implemented the procedures prepared by the consortium. Each NPM typically had several assistants, working from a base location that is referred to throughout this report as a national centre (NC). For the school level operations the NPM coordinated activities with school level staff, referred to as school co-ordinators (SCs). Trained test administrators (TAs) 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 consortium, such as national and international options, oversampling for regional comparisons, additional analyses and reporting, *e.g.* by language group;
- Established procedures for the security 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 consortium 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 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 to administer the tests within schools;
- Nominated suitable persons to work on behalf of the consortium 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 consortium;
- 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;
- A *Data Management Manual*, which described all aspects of the use of *KeyQuest*, the data entry software prepared by the consortium for the data entry of responses from the tracking instruments, test booklets and questionnaires.

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



School co-ordinators

School co-ordinators (SCs) co-ordinated school-related activities with the national centre and the test administrators.

The SC:

- 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 consortium;
- Received, distributed and collected the school questionnaire;
- Received and distributed the parent questionnaire in the countries that implemented this international option;
- Informed school staff, students and parents of the nature of the test and the test date, 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;
- Assisted the test administrator with room arrangements for the test day.

On the test day, the SC was expected to ensure that the sampled students attended the test session(s). If necessary, the SC 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 consortium that described in detail the activities and responsibilities of the SC.

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 TA 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. Prior to the test date, TAs were trained by national centres. Training included a thorough review of the *Test Administrator's Manual,* prepared by the consortium, 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 SC;
- Contacting the SC 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 SC;
- Completing the student tracking form and the assessment session report form (a form designed to summarise session times, student attendance, 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 SC; 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 (SA). A *School Associate's Manual* was prepared by the consortium, combining the source material provided in the individual SC and TA manuals to describe in detail the activities and responsibilities of the SA.

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 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 consortium who selected the school sample. Having the consortium 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. It also relieved the burden of this task from national centres. NPMs worked very closely with the consortium 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.

While all countries were required to thoroughly document their school sampling plan, a small number of countries were permitted to select the school sample themselves. In these cases, the national centre was required to explain in detail the sampling methods used, to ensure that they were consistent with those used by the consortium. In these cases, the standard procedure the consortium 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 study are provided in Chapter 4.

PREPARATION OF TEST BOOKLETS, QUESTIONNAIRES AND MANUALS

As described in Chapter 2, thirteen different test booklets had to be assembled with clusters of test items arranged according to the test booklet design specified by the consortium. 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 in the source versions, including allocation of items to pages. A slightly smaller or larger font than in the source version was permitted if it was necessary to ensure the same page set-up as that of the source version.

NPMs were required to submit their cognitive material in units, along with a form documenting any proposed national adaptations for verification by the consortium. NPMs incorporated feedback from the verifier into their material and assembled the test booklets. These were submitted once more to the consortium, who performed a final optical check (FOC) 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 and Computer Technology Familiarity international option questionnaire component was being added to the core component. Forty countries chose to administer this component. The core component had to be presented first in the questionnaire booklet.

Sixteen countries also administered an optional parent questionnaire.

As with the test material, source versions of the questionnaire instruments in both French and English were provided to NPMs to be used to assist in the translation of this material.

NPMs were permitted to add questions of national interest as national options to the questionnaires. Proposals and text for these were submitted to the consortium for approval as part of the process of reviewing adaptations to the questionnaires. It was recommended 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 consortium for approval. Following approval of adaptations, the material was verified by the consortium. 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 (SC) and test administrator (TA) manuals (or SA manual for those countries that combined the roles of the SC and TA) were also required to be translated into the national languages. French and English source versions of each manual were provided by the consortium. NPMs were required to submit a form documenting all proposed national adaptations to the manuals to the consortium for approval. Following approval of the adaptations, the manuals were prepared and submitted to the consortium. A verification of key elements 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 TA manual was translated into the language of the test.

THE SELECTION OF THE STUDENT SAMPLE

Following the selection of the school sample by the consortium, 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 *student listing form*, and was used by NPMs to select the student sample.

NPMs were required in most cases to select the student sample using *KeyQuest*, the PISA student sampling and data entry software prepared by the consortium. *KeyQuest* generated the list of sampled students for each school, known as the *student tracking form* that served as the central administration document for the study and linked students, test booklets and student questionnaires.

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



Only in exceptional circumstances were NPMs permitted to select their student sample without using *KeyQuest*. Alternative sampling procedures required the approval of the consortium prior to implementation.

PACKAGING AND SHIPPING MATERIALS

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

- Test booklets and student questionnaires for the number of students sampled;
- Student tracking form;
- Two copies of the Assessment Session Report Form;
- Packing form;
- Return shipment form;
- Additional materials, *e.g.* rulers and calculators, as per local circumstances;
- Additional school and student questionnaires and a bundle of extra test booklets.

Of the thirteen 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 student tracking 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 student tracking 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; TAs open the bundle immediately prior to the session and label the booklets with the students' names and ID numbers from the student tracking form.

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 TA.

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, 45% of the cognitive items across the science, reading and mathematics 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 4500 students, more than 116 000 responses had to be evaluated. An average of 26 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 consortium and provided to NPMs in coding guides for each of science, reading and mathematics.

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 consortium 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;
- Single code booklets according to the international design;
- Multiple code a selected sub-sample of booklets once the single coding was completed;
- Submit a sub-sample of booklets for the International Coding Review (see Chapter 13).

Detailed instructions for each step were provided in the *Main Study NPM's Manual*. Key aspects of the process are included here.

International 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 study. At the training sessions consortium 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 two months.

The consortium 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 study, it was recommended to have 16 coders coding across the domains of science and mathematics, and an additional four coders to code reading. 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 consortium 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 science and mathematics and one back-up 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 consortium.

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 consortium 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.

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



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 consortium 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 science clusters, the four mathematics clusters and the two 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 consortium. 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 student tracking 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. student tracking 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 consortium, 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 each of booklets 1, 3, 5, 6, 8 and 10 for multiple coding. The first two clusters from each of these booklets were multiple coded, except booklet 5 where the first three clusters were multiple coded. This arrangement ensured that all clusters were included in the multiple coding.

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 was 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 language represented 20% or more of the target population, the 600 booklets to be set aside for multiple coding were allocated in proportion to the language group. Multiple coding was not required for languages representing less than 20% of the target population.

Booklets for single coding

Single coding was required for the booklets remaining after those for multiple coding had been set aside, as well as for the clusters in the latter part of the book from those set aside for multiple coding. 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 last coder in the multiple-coding process coded these items in the booklets set aside for multiple coding, as well as the items requiring single coding from the 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 the reading 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, but occasionally with three degrees of correctness indicated by codes of "1", "2" and "3".



For the mathematics and science clusters, a two-digit coding scheme was adopted for the items requiring constructed responses. The first digit represented the degree of correctness code, as in reading; the second indicated the content of the response or the type of solution method used by the student. Two-digit codes were originally proposed by Norway for the TIMSS and were adopted in PISA because of their potential for use in studies of student learning and thinking.

Coder identification numbers

Coder identification numbers were assigned according to a standard three-digit format specified by the consortium. 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, sixteen coders coding across the domains of science and mathematics were given identification numbers 501 to 516. Four coders who coded just reading were given identification numbers 201 to 204. 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.

Single coding design

Single coding of science and mathematics

In order to code by cluster, each coder needed to handle four of the thirteen booklet types at a time. For example, science cluster 1 occurred in booklets 1, 9, 10 and 12. Each of these appearances 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 16 coders and the minimum sample size of 4 500 students from 150 schools. With 150 schools and 16 coders, each coder had to code a cluster within a booklet from eight or nine schools ($150 / 16 \approx 9$). Figure 6.1 shows how booklets needed to be assigned to coders for the single coding. Further explanation of the information in this table is presented below.

According to this design, cluster S1 in school subset 1 (schools 1 to 9) was to be coded by coder 501. cluster S1 in subset 2 (schools 10 to 18) was to be coded by coder 502, and so on. For cluster S2, coder 501 was to code all from subset 2 (schools 10 to 18) and coder 502 was to code all from subset 3 (schools 19 to 27). Subset 1 of cluster M2 (schools 1 to 9) was to be coded by coder 509.

Figure 6.1

Design for the single coding of science and mathematics

		Batches															
Cluster	**Booklets**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
S1	**1, 9, 10, 12**	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516
S2	**1, 2, 8, 11**	516	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515
S3	**2, 3, 5, 9**	515	516	501	502	503	504	505	506	507	508	509	510	511	512	513	514
S4	**1, 3, 4, 6**	514	515	516	501	502	503	504	505	506	507	508	509	510	511	512	513
S5	**4, 5, 11, 12**	513	514	515	516	501	502	503	504	505	506	507	508	509	510	511	512
S6	**5, 6, 8, 10**	512	513	514	515	516	501	502	503	504	505	506	507	508	509	510	511
S7	**1, 5, 7, 13**	511	512	513	514	515	516	501	502	503	504	505	506	507	508	509	510
M1	**3, 8, 12, 13**	510	511	512	513	514	515	516	501	502	503	504	505	506	507	508	509
M2	**4, 7, 8, 9**	509	510	511	512	513	514	515	516	501	502	503	504	505	506	507	508
M3	**2, 4, 10, 13**	508	509	510	511	512	513	514	515	516	501	502	503	504	505	506	507
M4	**3, 7, 10, 11**	507	508	509	510	511	512	513	514	515	516	501	502	503	504	505	506

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



If booklets from all participating schools were available before the coding began, the following steps would be involved in implementing the design:

Step 1: Set aside booklets for multiple coding and then divide the remaining booklets into school subsets as above (subset 1: schools 1 to 9; subset 2: schools 10 to 18, etc. to achieve 16 subsets of schools).

Step 2: Assuming that coding begins with cluster S1: coder 501 takes booklets 1, 9, 10 and 12 for school subset 1; coder 502 takes booklets 1, 9, 10 and 12 for school subset 2; etc.; until coder 516 takes booklets 1, 9, 10 and 12 for school subset 16.

Step 3: Coders code all of the first cluster S1 item requiring coding in the booklets that they have.

Step 4: The second cluster S1 item is coded in all four booklet types, followed by the third cluster S1 item, etc., until all cluster S1 items are coded.

Step 5: For cluster S2, as per the row of the table in Figure 6.1 corresponding to S2 in the left-most column, each coder is allocated a subset of schools different from their subset for cluster S1. Coding proceeds item by item within the cluster.

Step 6: For the remaining clusters, the rows corresponding to S3, S4, etc. in the table are followed in succession.

Single coding of reading

A similar design was prepared for the single coding of reading (Figure 6.2). As the recommended number of coders for reading (4) was one quarter that recommended for coding science and mathematics, each coder was allocated 'four subsets worth' of schools. Also, as there were just two different clusters of reading, each of which appeared in four booklet types, each coder coded just one of the four appearances of a cluster. This ensured that a wider range of coders was used for each school subset. For the coding of cluster R1, for example, coder 201 coded this cluster in booklet 1 from school subsets 1-4 (*i.e.* schools 1-36), coder 202 coded this cluster from booklet 1 for school subsets 5-8, and so on. For the next appearance of cluster R1 (in booklet 6), coder 204 coded these from school subsets 1-4, coder 201 from school subsets 5-8, and so on.

As a result of this procedure, the booklets from each subset of schools were processed by fifteen different coders, one for each distinct cluster of science and mathematics, and four for each cluster of reading. Each student's booklet was coded by four different coders, one for each of the four clusters in the student's booklet. Spreading booklets among coders in this way minimised the effects of any systematic leniency or harshness in coding.

Figure 6.2
Design for the single coding of reading

		Batches			
Cluster	**Booklet**	**1-4**	**5-8**	**9-12**	**13-16**
R1	**2**	201	202	203	204
R1	**6**	204	201	202	203
R1	**7**	203	204	201	202
R1	**12**	202	203	204	201
R2	**13**	201	202	203	204
R2	**11**	204	201	202	203
R2	**9**	203	204	201	202
R2	**6**	202	203	204	201



In practice, most countries would not have had completed test booklets back from all their sampled schools before coding needed to begin. NPMs were encouraged to organise the coding in two waves, so that it could begin after materials were received back from one-half of their schools. Schools would not have been able to be assigned to school sets for coding exactly in their school identification order, but rather by identification order combined with when their materials were received and processed at the national centre.

Booklet UH

Countries using the shorter, special purpose booklet UH 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 booklet UH, as for the main coding.

Multiple coding

For PISA 2006, four coders independently coded all short response and open-constructed response items from a selection of clusters from a sample of booklets. 100 of each of Booklets 1, 3, 5, 6, 8 and 10 (a total of 600 booklets) were selected for this multiple coding activity. 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 (science, reading or mathematics), 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 four 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 science and mathematics

The specified multiple coding design for science and mathematics, shown in Table 6.1, assumed 16 coders with identification numbers 501 to 516. The importance of following the design exactly as specified was stressed, as it provided for links between clusters and coders. Table 6.1 shows 16 coders grouped into four groups of four, with Group 1 comprising the first four coders (501-504), Group 2 the next four (505-508), etc. The design involved two steps, with the booklets divided into two sets - booklets 1, 3, 8 and 10 made up one set, and booklet 5 the second set. The coders assigned to the second step consisted of one coder from each of the groups formed at the first step. The four codings were to be carried out by rotating the booklets to the four coders assigned to each group.

In this scenario, with all 16 coders working, booklets 1, 3, 8 and 10 were to be coded at the same time in the first step. The 100 booklet 1's, for example, were to be divided into four bundles of 25 and rotated among coders 501, 502, 503 and 504, so that each coder eventually would have coded clusters S1 and S2 from all of the 100 booklets. At the fourth rotation, after each coder had finished the multiple coding of clusters S1

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



and S2 from the 25 booklets in their pile, they would then single code any science or maths clusters from the second half of the booklet. The same pattern was to be followed for booklets 3, 8 and 10.

After booklets 1, 3, 8 and 10 had been put through the multiple-coding process, one coder from each of the four coding groups was selected to complete the multiple-coding of booklet 5. That is, coders 501, 506, 511 and 516 were assigned to code booklet 5,

Allocating booklets to coders for multiple coding was quite complex and the coding supervisor had to monitor the flow of booklets throughout the process.

Table 6.1
Design for the multiple coding of science and mathematics

Booklet	Coder IDs	Clusters for multiple coding	Clusters for single coding
1	501. 502. 503. 504	S1. S2	S4. S7
3	505. 506. 507. 508	S3. S4	M4. M1
8	509. 510. 511. 512	M1. M2	S2. S6
10	513. 514. 515. 516	M3. M4	S6. S1
5	501. 506. 511. 516	S5. S6. S7	S3
6	Any coders available from 501 – 516		S4. S6

Multiple coding of reading

The multiple-coding design for reading shown in Table 6.2 assumed four coders, with identification numbers 201 to 204.

If different coders were used for science or mathematics, a different multiple-coding design was necessary. The NPM would negotiate a suitable proposal with the consortium. The minimum allowable number of coders coding a domain was four; in this case each booklet had to be coded by each coder.

Table 6.2
Design for the multiple coding of reading

Booklet	Coder IDs	Clusters for multiple coding	Clusters for single coding
6	201. 202. 203. 204	R1. R2	none

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 (booklets 1 to 13), 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 science, mathematics 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



reading and the coding of science and mathematics be done at least partly at different times (for example, reading 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. In science and mathematics, this reliability was to be assessed on the first digit of the two-digit codes. 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 coding

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 and student's occupational expectation. Four-digit International Standard Classification of Occupations (ISCO88) codes (International Labour Organisation, 1988) were assigned to these three 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. A national classification was preferred because relationships between occupational status and achievement could then be compared within a country using both international and national measures of occupational status.

The PISA website gave a clear summary of ISCO codes and occupational titles for countries to translate if they had neither a national occupational classification scheme nor access to a full translation of ISCO.

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 consortium provided participating countries with the data entry software *KeyQuest, which* 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 SC and TA on the student tracking 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 *KeyQuest* Manual provided generic technical details of the functionality of the *KeyQuest* software. A separate *Data Entry Manual* provided complete instructions specific to the main study regarding data entry, data management and validity checks.

Data Checking

NPMs were responsible for ensuring that many checks of the quality of their country's data were made before the data files were submitted to the consortium. These checks were explained in detail in the Data Entry Manual, and could be simply applied using the *KeyQuest* software. The checking procedures required that the list of sampled schools and the student tracking 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 submission

Files to be submitted included:

- Data for the test booklets and context questionnaires;
- Data for the international option instrument(s), if used;
- Data for the multiple-coding study;
- Session report data;
- Data cleaning reports;
- The list of sampled schools;
- Student tracking forms.

Hard or electronic copies of the last two items were also required.

After data were submitted

NPMs were required to designate a data manager who would work actively with the consortium'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 STUDY REVIEW

NPMs were required to complete a structured review of their main study operations. The review was an opportunity to provide feedback to the consortium 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 study review was submitted to the consortium four weeks after the submission of the national database.



7

Quality Assurance

PISA quality control 124
- Comprehensive operational manuals 124
- National level implementation planning document 124

PISA quality monitoring 124
- Field trial and main study review 124
- Final optical check 126
- National centre quality monitor (NCQM) visits 126
- PISA quality monitor (PQM) visits 126
- Test administration 127
- Delivery 128

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



It is essential that users of the PISA data have confidence that the data collected through the PISA survey are fit for use for the intended purposes. To ensure this, the various data collection activities have been undertaken in accordance with strict quality assurance procedures. The quality assurance that provides this confidence in the fitness for use of the PISA 2006 data consists of two components. The first is to carefully develop and document procedures that result in data of the desired quality; the second is to monitor and record the implementation of the documented procedures. Should it happen that the documented procedures are not fully implemented, it is necessary to understand to what extent they were not and the likely implications for the data.

PISA QUALITY CONTROL

PISA quality standards are established through comprehensive operational manuals and agreed national level implementation planning documents. These materials state the project goals, and how to achieve those goals according to clearly defined procedures on an agreed timeline. Each stage of the process is then monitored to ensure that implementation of the programme has proceeded as planned.

Comprehensive operational manuals

PISA field operational manuals describe the project implementation procedures in great detail and clearly identify connections to the *PISA Technical Standards* at various stages.

For the PISA 2006 field trial and main study, the *PISA National Project Manager's Manual,* the *PISA Test Administrator's Manual,* the *PISA School Coordinator's Manual,* the *PISA School Sampling Preparation Manual,* and the *PISA Data Management Manual* were produced.

National level implementation planning document

National level planning documents are developed from the operational manuals and allow participants to record their specific project information and any approved variations to standard procedures.

Through a negotiation process, the consortium and each NPM reach an agreement on all the planning documents submitted by the national centre. For PISA 2006 these documents included sampling forms, the translation plan, the preferred verification schedule, the print quality agreement, a form covering participation in international and national options, and adaptation forms related to each of the manuals, the questionnaires and the cognitive test instruments.

To increase the transparency of this negotiation process, all planning documents submitted by the national centre are posted on the PISA website, with file status showing as 'submitted', 'in progress' or 'agreed'. Each national centre's key project information is also displayed on the profile page of the PISA website.

PISA QUALITY MONITORING

While the aim of quality control is to establish procedures and guide implementation, quality monitoring activities are set to observe and record any deviations from those agreed procedures during the implementation of the study.

Field trial and main study review

After the implementation of the field trial and the main study, NPMs were given the opportunity to review and provide feedback to the consortium on all aspects of the field operations.



The field trial and main study 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 consortium;
- Review the usefulness of the two modes used to deliver materials to the national centre – email and the PISA website or and the newly developed web pages that show the national centre's profile and milestone documents;
- Implementation of national and international options;
- Confirm if national centre had executed any national and international options as agreed;
- 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;
- Sampling plan;
- Confirm if the PISA field trial test was implemented as agreed in the sampling plan;
- Printing;
- Review the print quality agreement process;
- Link item revision;
- Confirm if the revision of the link items proposed in the test adaptation spreadsheet (TAS) had been implemented as agreed;
- Security arrangements;
- Review security arrangements to confirm if they had been implemented;
- Archiving of materials;
- Confirm if the national centre had archived the test materials in accordance with the technical standards;
- Test administration;
- Review TA training processes and test administration procedures;
- Special education need (SEN) codes;
- Review the use of SEN codes;
- Coding;
- Review coder training procedures, coding procedures, coding designs and the time required for coding;
- 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 consortium for a final optical check (FOC). The FOC is undertaken by the consortium's verifiers and involves the page-by-page inspection of test booklets and questionnaire forms with regard to correct item allocation, layout, page numbering, item numbering, graphic elements, item codes, footers and so on (see Chapter 5).

Any errors found during the FOC are recorded and forwarded to National Centres for correction.

National Centre Quality Monitor (NCQM) visits

Fifteen consortium representatives, national centre quality monitors (NCQMs), visited all 57 participating national centres in the month preceding the country's main study testing period. During the visit, the NCQM conducted a half-day training session for PISA quality monitors which included the selection of a list of schools to visit and a face-to-face interview with the NPM or a representative. Any potential problems identified by the NCQM were forwarded to the relevant consortium expert for appropriate action. In some cases the school list was not ready at the time of the visit, so the selection of schools to visit was carried out through e-mail and phone calls afterwards.

The NCQMs have comprehensive knowledge about and extensive experience with 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 suitably aware of country-specific information during the interview with NPMs.

The NCQM interview schedule is a list of questions that was prepared for the consortium representatives to lead the interview in a structured way, so that the outcomes of the NCQM site visit could be recorded systematically. This interview schedule covers the following areas:

- General organisation of PISA in each country;
- Test administration;
- Security and confidentiality;
- Selection of school sample for the main study;
- Selection of student sample for the main study;
- Student tracking form;
- Translation and verification;
- National and international options;
- Assembly of assessment booklets;
- Coding;
- Adequacy of the manuals;
- Data entry;
- PISA quality monitors.

PISA quality monitor (PQM) visits

PQMs are individuals employed by the consortium and located in participating countries. They visit a sample of schools to record the implementation of the documented field operations in the main study. They typically visit 15 schools in each country.



All PQMs are nominated by the NPMs through a formal process of submission of nominations to the consortium. Based upon the NPM nominations, which are accompanied by resumes, the consortium selects PQMs who are totally independent from the national centre, knowledgeable in testing procedures or with a background in education and research, and able to speak English or French. Where the resume does not match the selection criteria, further information or an alternate nomination is sought.

Typically, two PQMs were engaged for each country. Each PQM visited seven or eight schools. The *PQM Manual*, other operational manuals and copies of data collection sheets were posted to all PQMs upon receipt of their signed confidentiality agreement. The PQMs were also given access to a designated PQM web page from which they could download materials and information. During the NCQM visit, all PQMs were trained in person. The NCQM and PQMs collaborated to develop a schedule of school visits to ensure that a range of schools was covered and that the schedule of visits was both economically and practically feasible. The consortium paid the expenses and fees of each PQM.

The majority of school visits were unannounced. However, in some countries it is not possible to gain access to a school without arrangement beforehand.

A PQM data collection sheet was developed for PQMs to record their observations systematically during each school visit. The data collection sheet covers the following areas:

- Preparation for the assessment;
- Test session activities;
- General questions concerning the assessment;
- Interview with the school co-ordinator.

A general observation sheet was also developed for PQMs to record general impressions of the implementation of PISA at the national level. The general observation sheet records information on:

- Security of materials;
- Contribution of test administrators;
- Contribution of school coordinators;
- Support from the national centre;
- Attitude and response of students to test sessions;
- Attitude and response of students to the questionnaire;
- Suggestions for improvement.

Test administration

Test administrators record all key test session information using a test session report. This report provides detailed data on test administration, including:

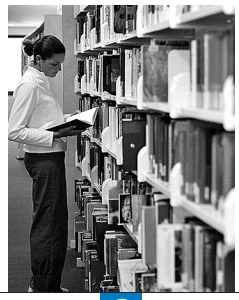
- Session date and timing;
- The position of the test administrator;
- Conduct of the students;
- Testing environment.



Delivery

All quality assurance data collected throughout the cycle are entered and collated in a central database. Comprehensive reports are then generated for the Technical Advisory Group (TAG) to consider during the data adjudication process (see Chapter 14).

The TAG experts use the consolidated quality-monitoring reports from the central 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.



8

Survey Weighting and the Calculation of Sampling Variance

Survey weighting 130
The school base weight 131
- The school weight trimming factor 132
- The student base weight 132
- School non-response adjustment 132
- Grade non-response adjustment 134
- Student non-response adjustment 135
- Trimming student weights 136
- Comparing the PISA 2006 student non-response adjustment strategy with the strategy used for PISA 2003 136
- The comparison 138

Calculating sampling variance 139
- The balanced repeated replication variance estimator 139
- Reflecting weighting adjustments 141
- Formation of variance strata 141
- Countries where all students were selected for PISA 141

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



Survey weights were required to analyse PISA data, to calculate appropriate estimates of sampling error and to make valid estimates and inferences. The consortium calculated survey weights for all assessed, ineligible and excluded students, and provided variables in the data that permit users to make approximately unbiased estimates of standard errors, to conduct significance tests and to create confidence intervals appropriately, given the sample design for PISA in each individual 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 correct 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 very large, the selection probability was based on the assumption that only a sample of its students would be selected for PISA. But if the school turned out to be quite small, all students would have to be included and would have, overall, a higher probability of selection in the sample than planned, making these inclusion probabilities higher than those of most other students in the sample. Conversely, if a school thought to be small turned out to be large, the students included in the sample would have had 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 eligible population in a school (such as those 15-year-olds 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 eligible and not excluded, but did not participate in the assessment will be under-represented in the data unless weighting adjustments are made;
- Trimming 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 sampling weights can lead to unstable estimates – large sampling errors – but cannot be well estimated. Trimming weights introduces a small bias into estimates but greatly reduces standard errors (Kish, 1992).

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: 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). See Cochran, (1977), Lohr (1999) and Särndal, Swensson and Wretman (1992) for the underlying statistical theory for the analysis of survey data.



The weight, W_{ij}, for student j in school i consists of two base weights – the school and the within-school – 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} is the school base weight, is given as the reciprocal of the probability of inclusion of school i into the sample;

w_{2ij} is 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 the fact that, in some countries, in some schools only 15-year-old students who were enrolled in the modal grade for 15-year-olds 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 trimming factor, used to reduce unexpectedly large values of w_{1i}; and

t_{2ij}, is a student 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 probability proportional- to-size school sampling method used in PISA, this is given as:

8.2

$$w_{1i} = \begin{cases} int(g/i) \,/\, mos(i) & \text{if } mos\ (i) < int\ (g/i) \\ 1 & otherwise \end{cases}$$

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-olds in the school, if it was greater than the predetermined Target Cluster Size (*TCS*) (35 in most countries).

If the enrolment of 15-year-olds was less than the *TCS*, then *mos* (*i*) = *TCS*.

The term *int*(*g*/*i*) 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 100 15-year-olds 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, with a school sample size of 150, then int(1/*i*) = 150000/150 = 1000, for school *i* (and others in the sample),

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



giving $w_{1i} = 1000/100 = 10.0$. Roughly speaking, the school can be thought of as representing about 10 schools from 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 $w_{1i} = 1$.

The school 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 see if the school weights required trimming. The school trimming factor t_{1i}, is the ratio of the trimmed to the untrimmed school base weight, and is equal to 1.0000 for most schools and therefore most students, and never exceeds this value.

The school-level trimming adjustment was applied to schools that turned out to be much larger than was believed at the time of sampling – where 15-year-old enrolment exceeded $3 \times \max(TCS, mos(i))$. For example, if $TCS = 35$, then a school flagged for trimming had more than 105 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 $3 \times \max(TCS, mos(i))$ in the school base weight formula.

The student base weight

The term w_{2ij} is referred to as the student base weight. With the PISA procedure for sampling students, w_{2ij} did not vary across students (j) within a particular school i. This is given as:

8.3

$$w_{2ij} = enr(i) \Big/ sam(i)$$

where $enr(i)$ is the actual enrolment of 15-year-olds in the school (and so, in general, is somewhat different from the estimated $mos(i)$), and $sam(i)$ is the sample size within school i. It follows that if all students from the school were selected, then $w_{2ij} = 1$ for all eligible students in the school. For all other cases $w_{2ij} > 1$.

In the case of the international 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 (the Czech Republic, Korea, Mexico, Norway, Sweden, certain explicit strata in Switzerland and Uruguay) have the same within school student weights for the extra grade students as those for PISA students from the same school.

Additional weight components were needed for the grade students in Chile, Germany, Liechtenstein, Mexico and certain strata in Switzerland. For these first four 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).) For Mexico, the extra weight component at the school level accounted for the sub-sampling of schools in which the grade sample would be implemented. In these five countries, the extra weight component resulted in the necessity of a second weighting stream for the extra grade students.

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.

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



Table 8.1
Non-response classes

		Implicit stratification variables used to create school non-response cells (within explicit stratum) and number of original and final cells	Number of original cells	Number of final cells
OECD	Australia	Geographic Zone (8); School Level for TAS and ACT Government Schools (3)	88	78
	Austria	Province-District (121)	114	27
	Belgium	Flanders -- Index of Overaged students; French Community -- Public/Private School Types for Special Education Schools (4); Index of Overaged students for Regular Schools; German Community – None	154	38
	Canada	Public/Private(2); Urban/Rural(2)	107	51
	Czech Republic	Region for Programmes 3. 4. 5. 6 (14)	151	38
	Denmark	School Type (5); Geo Area (5)	37	16
	Finland	None	33	13
	France	None	20	11
	Germany	School Type for Normal Schools (5); State for Other Schools (16)	63	27
	Greece	School Type (3); Public/Private (2) when both in an explicit stratum	40	18
	Hungary	Region (7); National Grade 10 math Score Categories (5) for Non-Primary Schools	105	24
	Iceland	Urban/Rural (2); School Size (4)	33	24
	Ireland	School Type (3); School Gender Composition Categories (5)	31	13
	Italy	Public/Private (2)	107	54
	Japan	Levels of proportion of students taking University/College Entrance Exams(4)	16	12
	Korea	School Level (2)	12	8
	Luxembourg	None	16	4
	Mexico	School Size (3); Public/Private (2); Urban/Rural (2); School Level (2); School Program (4 For Each School Level)	343	107
	Netherlands	School Type (6 for School Track A and 3 for School Track B)	9	5
	New Zealand	Public/Private (2); Socio-Economic Status Category (3) and Urban/Rural (2) for Public Schools	7	6
	Norway	None	12	8
	Poland	Urbanicity (4)	11	8
	Portugal	Public/Private (2); Socio-Economic Status Category (4)	50	15
	Slovak Republic	Programme (9); Language (2) in 4 of the Regions	60	16
	Spain	2 or 3 digits of Postal Code for adjudicated regions	323	84
	Sweden	Urbanicity (5) for Private Lower Secondary schools; Public/private (2) for Upper Secondary schools; Administrative Province (25) for Upper Secondary schools; Income Quartiles (4) for all except Private Lower Secondary schools	55	23
	Switzerland	School Type (28); Canton (26)	186	52
	Turkey	School Level (3); Public/Private (2); Urban/Rural (2)	39	12
	United Kingdom	England: School Type (6). GCSE Performance (6). Region (4). Local Authority Northern Ireland: School Type (3). Education and Library Board Region (5) Scotland: None Wales: School Type (2). Region (3). Local Authority	252	65
	United States	Public/Private (2); Region (4); Urbanicity (3); Minority Status (2); Grade Span (4); Postal Code	79	15
Partners	Argentina	Sector (2); School Type (5); School Level (5)	96	25
	Azerbaijan	Urbanicity (4); Education Department (5)	108	9
	Brazil	School Type (3); HDI Category (2); School Size (3); Urban/Rural (2); Capital/Non-Capital (2)	355	124
	Bulgaria	School Type (3); Settlement Size (5); State/Municipal (2); Public/Private (2)	94	79
	Chile	Urban/Rural (2); Region (13)	114	29
	Columbia	Urban/Rural (2); Public/Private(2)	4	3
	Croatia	County (21)	110	21
	Estonia	Urbanicity (4); School Type (4); County (15)	67	18
	Hong Kong-China	Student Academic Intake (4)	10	7
	Indonesia	School Type (5); Public/Private (2); National Achievement Score Categories (3)	225	62
	Israel	Location (9) for Public Schools. Except For Schools in Druz Migzar Sector; Group Size (5) for Regular Public Schools; Gender Composition (3) for Religious Public Schools; Migzar Sector (3) for Regular Public Arabic Schools; Cultivation Categories (4) for Public Jewish Schools; Cultivation (Continuous Measure) in All	61	31
	Jordan	Urbanicity (2); School Gender Composition (3); School form (2)	28	16
	Kyrgyzstan	School Type (5)	60	18
	Latvia	Urbanicity (4); School Type (4)	15	8
	Liechtenstein	Urbanicity (3); Public/Private(2)	2	2
	Lithuania	None	12	8
	Macao-China	Secondary Levels (3)	14	3
	Montenegro	Region (3) for Primary Schools; Urban/Rural (2); School Type (3)	14	10
	Qatar	Qatari/Non-Qatari (2)	26	18
	Romania	Language (3); Urbanicity (2)	13	6
	Russian Federation	Urbanicity (9); School Type (4); School Sub-Type (16)	194	94
	Serbia	Urban/Rural (2); School Type (7)	77	19
	Slovenia	Urbanicity (4)	24	18
	Chinese Taipei	Public/Private (2)	60	30
	Thailand	Local Area (9)	57	22
	Tunisia	Category of Grade Repeating (3) for General Public Schools; East/West (2) for Private Schools and Vocational Schools; North/South (2) for all	39	13
	Uruguay	Area (4); Shift (4) for Public Secondary Schools; Shift (4) for Public Technical Schools	65	40



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 stratum, based on their enrolment size. It was desirable to ensure that each group had at least six participating schools, as small groups can lead to unstable weight adjustments, which in turn would inflate the sampling variances. However, 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. Adjustments greater than 2.0 were flagged for review, as they can cause increased variability in the weights and lead to an increase in sampling variances. In either of these situations, cells were generally collapsed over the last implicit stratification variable(s) until the violations no longer existed. In countries with very high overall levels of school non-response after school replacement, the requirement for school non-response adjustment factors all to 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 $\Gamma(i)$ the schools within the group (originals and replacements) that participated, while the sum in the numerator is over $\Omega(i)$, those same schools, plus the original sample schools that refused and were not replaced. The numerator estimates the population of 15-year-olds in the group, while the denominator gives the size of the population of 15-year-olds 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 eligible students enrolled, no adjustment was necessary since this was neither non-response nor under-coverage.

Table 8.1 shows the number of school non-response classes that were formed for each country, and the variables that were used to create the cells.

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-olds in the modal grade for 15-year-olds, rather than all 15-year-olds. Since the modal grade generally includes the majority of the population to be covered, such schools may be accepted as participants. 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-olds, 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 2006, 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

$$f_{1i}^{A} = \begin{cases} \sum_{k \in C(i)} w_{1k} enra(k) \\ 1 \end{cases} \Big/ \sum_{k \in B(i)} w_{1k} enra(k)$$

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.

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 1.0000 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.

Student non-response adjustment

Within each final school non-response adjustment cell, explicit stratum and high/low grade, gender, and school combination, 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 school non-response adjustment cell and explicit stratum-grade-gender-school combination; and,

$X(i)$ is all assessed students in the final school non-response adjustment cell and explicit stratum-grade-gender-school combination plus all others who should have been assessed (*i.e.* who were absent, but not excluded or ineligible).

The high and low grade categories in each country were defined so that each contain a substantial proportion of the PISA population in each explicit stratum of larger schools.

The definition was then applied to all schools of the same explicit stratum characteristics but regardless of school size. 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 of small cells (*i.e.* final school non-response adjustment cell and explicit stratum/grade/gender/school category combinations) sizes (fewer than 15 respondents), it was necessary to collapse cells together, then apply the more complex formula shown above. 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 and explicit stratum. Note that the calculation of the high / low grades, the use of gender, and the order of cell collapsing represent differences from the student non-response adjustment strategy used for PISA 2003.



Some schools in some countries had very low student response levels. In these cases it was determined that the small sample of assessed students was potentially too biased as a representation of the school to be included in the PISA data. For any school where the student response rate was below 25%, the school was therefore 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 2.0000 and 4.0000, 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 (the Czech Republic, Korea, Mexico, Norway, Sweden, certain explicit strata in Switzerland and Uruguay), 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, Germany, Liechtenstein, Mexico and certain strata in Switzerland since a separate weighting stream was needed for the grade students.

As noted above, a few beneficial changes were introduced to the 2006 strategy for student non-response adjustments: namely the calculation of the high/low grade categories within explicit strata rather than over the whole set of schools, the use of gender in forming the student non-response cells, and the removal of the school as the basis for the final cell formation. As a result of this latter change, the final weights for students within schools could vary.

Trimming student weights

This final trimming check was used to detect student records 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 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 very large weights, which adds considerably to 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 1.0000 for the great majority of students. The final weight variable on the data file was called *w_fstuwt*, which is the final student weight that incorporates any student-level trimming. As in PISA 2000 and PISA 2003, little trimming was required at either the school or the student levels.

Comparing the PISA 2006 student non-response adjustment strategy with the strategy used for PISA 2003

The student non-response adjustment section of this chapter noted that changes had been made to the 2006 student non-response adjustment strategy. While the changes were thought to be beneficial because they used more student information in the adjustments, an assessment of the impact of the change was nevertheless conducted. This section is devoted to that investigation, which compares the 2006 student non-response adjustment strategy to the 2003 non-response adjustment strategy for countries that also participated in PISA 2003.



Recall that the final student weight consists of:

- The school base weight;
- The school weight trimming factor;
- The school non response adjustment;
- The student base weight;
- The student non response adjustment;,
- The student weight trimming factor;

as well as potentially the grade non-response adjustment factor if needed (but this was not needed for 2006).

The student non-response adjustment is designed to reduce the potential bias introduced by the non-participating students. As described in the *PISA 2000 Technical Report* (Adams & Wu, 2002), the student non- response adjustment factor was computed in PISA 2000 as follows:

- Within each participating school, the student non-response adjustment was equal to the sum of all sampled student initial weights divided by the sum of the participating student initial weights;
- If the school sample had fewer than 15 participating students, the school was collapsed with the next school within the same school non-response cell;
- If the adjustment factor was greater than 2, the school was collapsed with the next school within the same school non-response cell.

Secondary analyses of the PISA 2000 student tracking forms have revealed substantial differential non-response rates in some countries (Monseur, 2005) countries. For instance, the response rates of Portuguese students attending grade 7 to grade 10 were respectively equal to 0.76, 0.80, 0.87 and 0.88. As grade highly correlates with performance, these differential response rates introduced a bias.

In 2003, it was therefore decided to include the grade attended by the student in the computation of the student non-response adjustment. Concretely:

- Grades were grouped into two categories: lower grades and higher grades so that each had a substantial proportion of students;
- Within each participating school and high/low grade combination, the student non-response adjustment was calculated;
- If the combination of school and high/low grade cells had fewer than 10 participating students or the corresponding adjustment factor was greater than 2, the two initial student non-response cells were collapsed within that school;
- If this collapsing within a particular school did not allow satisfying the two criteria (*i.e.* a minimum of 10 students and an adjustment factor lower than 2), further collapsing was done as in PISA 2000.

This procedure, however, had a limited impact as in most countries, the within school sample size was equal to 35 students. Therefore, the requirement of 10 participating students per student non-response cell was not reached in a large number of schools, so that the two non-response cells (lower versus higher grade cells) were collapsed.



Previous analyses (Monseur, 2005) had also shown a differential participation rate for boys and girls. For instance, in Portugal, the student response rate for boys was 82.6% and for girls 87.8%. As gender also correlated with performance, particularly in reading literacy, the student non-response adjustment developed for PISA 2006 aimed to better compensate for differential grade and gender response rates.

As described above the student non-response adjustment was computed in PISA 2006 as follows:

For each school, four student non response cells were created:

- Higher grades/girls;
- Higher grades/boys;
- Lower grades/girls;,
- Lower grade/boys,

where the high/low grades were derived within each explicit stratum.

In single sex schools or in schools with students enrolled in only one grade, only two student non-response cells were created.

The major change between the previous procedures and the PISA 2006 procedure is not the addition of the gender variable for creating the non- response cell but the ordering of the collapsing.

In PISA 2003, non-response cells were first collapsed within school, then, if required, schools were collapsed.

In 2006, a non-response cell from a school was first collapsed with a non-response cell sharing the same gender and grade but from another school. However, these two schools had to be in the same school non-response cell and explicit stratum. If further collapsing was required, usually non-response cells were collapsed across gender and then across grade.

As this modification in the computation of the student non-response adjustment might have an impact on population estimates, in particular on performance estimates, it was decided to compute the 2006 data the student non response adjustment according to (i) the new algorithm and (ii) the PISA 2003 algorithm for only the countries that participated in the 2003 and 2006 surveys. Comparing population estimates for the two sets of weights allows measuring the impact of the weighting modification.

The comparison

Three sets of weights have been used in the subsequent analyses:

- The initial student weight that consists of:
  a. The school initial base weight;
  b. The school trimming factor;
  c. The school non-response adjustment factor;
  d. The student initial within school weight;
- The final student weight based on the 2003 non response adjustment method; and,
- The final student weight based on the 2006 non response adjustment method.

For the second and third sets of weights, only participating students were included in the analyses while absent and participating students were included for the first set of weights.

Each set of weights can be used with the gender and grade data collected through the student tracking form.



In several countries, the difference between the males' response rate and the females' response rate was greater than 2%. Even with these response rate differences, the 2006 method weighted estimates were equal or very close to the estimates computed using the initial student weights and data from the student tracking form, while the 2003 method weighted estimates differed to a greater extent. The 2006 adjustment method appears more efficient in reducing a potential bias due to the differential participation rates between males and females (as expected since gender was not used in the PISA 2003 strategy).

Regarding grade, there were also several countries where the difference between the initial weighted estimate and the 2003 method adjusted estimate was at least 1%. In almost all cases, the 2006 method adjusted weights reduced the differences when compared to the initial weighted estimates.

Finally, looking at the three literacy scales, in all countries, the differences in mean performance between the two sets of weight results was always less than two PISA scale points and for most countries the difference was less than one scale point.

Country comparisons are not provided since all differences were small.

In summary, the new method of student non-response adjustment used in 2006 does not appear to have generated any spurious changes in achievement means of any consequence.

CALCULATING SAMPLING VARIANCE

To estimate the sampling variances of PISA parameter estimates, a replication methodology was employed. This reflected 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, although computationally the calculation of the two components can be carried out in a single program, such as *WesVar 5.1* (Westat, 2007).

The balanced repeated replication variance estimator

The approach used for calculating sampling variances for PISA is known as balanced repeated replication (BRR), or balanced half- samples; the particular variant known as Fay's method was used. This method is very 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 BRR over the jackknife is that the jackknife method 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 very unstable, and despite empirical evidence that it can behave well in a PISA-like design, theory is lacking. In contrast BRR does not have this theoretical flaw. The standard BRR procedure can become unstable when used to analyse sparse population subgroups, but Fay's modification overcomes this difficulty, and is well justified in the literature (Judkins, 1990).

The BRR approach was implemented as follows, for a country where the student sample was selected from a sample of schools, rather than all schools:

- Schools were paired on the basis of the explicit and implicit stratification and frame ordering used in sampling. The pairs were originally sampled schools, with each participating replacement taking the place of the original school that it replaced. 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 (the Primary Sampling Unit, PSU) 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 are 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 is 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 sampling weights from one of the two PSUs in each stratum by 1.5, and the weights from the remaining PSUs by 0.5. The determination as to which PSUs 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 & 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 any 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, to the extent possible, from different implicit sampling strata also;
- 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. In some countries for part of the sample (and for the entire samples for 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 (the Czech Republic, Korea, Mexico, Norway, Sweden, certain explicit strata in Switzerland and Uruguay) 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, Germany, Liechtenstein, certain strata in Switzerland, and Mexico, since a separate weighting stream was needed in these cases;
- The variance estimator is then:

8.7

$$V_{BRR}(X^*) = 0.05 \sum_{t=1}^{80} \left\{ (X_t^* - X^*)^2 \right\}$$

The properties of BRR 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 glosses over 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 consortium 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 big component of variance in any PISA country, but the procedure gives a more accurate estimate of sampling variance.

Countries where all students were selected for PISA

In Iceland, Liechtenstein, Luxembourg, and Qatar, all eligible students were selected for PISA. It might be considered surprising that the PISA data should reflect any sampling variance in these countries, but students have been assigned to variance strata and variance units, and the BRR formula does give a positive



estimate of sampling variance for two reasons. First, in each country there was some student non-response, and, in the case of Iceland and Qatar, some school non-response. Not all 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 12 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.

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



9

Scaling PISA Cognitive Data

The mixed coefficients multinomial logit model 144
- The population model 145
- Combined model 146

Application to PISA 146
- National calibrations 146
- National reports 147
- International calibration 153
- Student score generation 153

Booklet effects 155
Analysis of data with plausible values 156
Developing common scales for the purposes of trends 157
- Linking PISA 2003 and PISA 2006 for reading and mathematics 158
- Uncertainty in the link 158

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



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 & Wilson, 1997).

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, ξ, while the student outcome levels (the latent variable), θ, 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 $\xi^T = (\xi_1, \xi_2,\ldots, \xi_p)$, 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. D, 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 $\mathbf{A}^T = (\mathbf{a}_{11}, \mathbf{a}_{12},\ldots, \mathbf{a}_{1K_1}, \mathbf{a}_{21}, \ldots, \mathbf{a}_{2K_2},\ldots, \mathbf{a}_{IK_i})$, 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 $\theta = (\theta_1, \theta_2,\ldots, \theta_D)'$, 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 other responses may also be scored zero.)

The probability of a response in category j of item i is modelled as

9.2

$$\Pr(X_{ij} = 1; A, B, \xi \mid \theta) = \frac{\exp(b_{ij}\theta + a'_{ij}\xi)}{\sum_{k=1}^{K_i} \exp(b_{ik}\theta + a'_{ik}\xi)}.$$

For a response vector, we have:

9.3

$$f(x; \xi \mid \theta) = \psi(\theta, \xi) \exp[x'(B\theta + A\xi)]$$

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



with

9.4

$$\psi(\theta, \xi) = \left\{ \sum_{z \in \Omega} \exp\left[\mathbf{z}^T (\mathbf{B}\theta + \mathbf{A}\xi) \right] \right\}^{-1}$$

where Ω 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, θ. The complete definition of the model, therefore, requires the specification of a density, $f_\theta(\theta, \alpha)$ for the latent variable, θ. Let α symbolise a set of parameters that characterise the distribution of θ. 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

$$f_\theta(\theta; \alpha) \equiv f_\theta(\theta; \mu, \sigma^2) = (2\pi\sigma)^{-1/2} \exp\left[-\frac{(\theta - \mu)^2}{2\sigma^2} \right]$$

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, $\mathbf{Y}_n^T\beta$, 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

$$\theta_n = \mathrm{Y}_n^T\beta + E_n$$

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

$$f_\theta(\theta_n; \mathbf{Y}_n, \mathrm{b}, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (\theta_n - \mathbf{Y}_n^T\beta)^{\mathrm{T}} (\theta_n - \mathbf{Y}_n^T\beta) \right]$$

a normal distribution with mean $\mathbf{Y}_n^T\beta$ and variance σ^2. If is used as the population model then the parameters to be estimated are β, σ^2 and ξ.

The generalisation needs to be taken one step further to apply it to the vector-valued θ rather than the scalar-valued θ. The extension results in the multivariate population model:

9.9

$$f_\theta(\theta_n; \mathbf{W}_n, \gamma, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2} (\theta_n - \gamma\mathbf{W}_n)^T \Sigma^{-1} (\theta_n - \gamma\mathbf{W}_n) \right]$$

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



where γ is a $u \times d$ matrix of regression coefficients, Σ 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 [9.3] and the population model [9.9] are combined to obtain the unconditional, or marginal, item response model:

9.10

$$f_x(x; \xi, \gamma, \Sigma) = \int_\theta f_x(x; \xi \mid \theta) f_\theta(\theta; \gamma, \Sigma) d\theta .$$

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 γ, Σ and ξ.

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

$$h_\theta(\theta_n; W_n, \xi, \gamma, \Sigma \mid x_n) = \frac{f_x(x_n; \xi \mid \theta_n)\ f_\theta(\theta_n; W_n, \gamma, \Sigma)}{f_x(x_n; W_n, \xi, \gamma, \Sigma)}$$

$$= \frac{f_x(x_n; \xi \mid \theta_n)\ f_\theta(\theta_n; W_n, \gamma, \Sigma)}{\int_{\theta_n} f_x(x_n; \xi \mid \theta_n)\ f_\theta(\theta_n; W_n, \gamma, \Sigma)} .$$

APPLICATION TO PISA

In PISA, this model 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 is used in conjunction with the population model , but conditioning variables are not used. That is, it is assumed that students have been sampled from a multivariate normal distribution.

Two five-dimensional scaling models were used in the PISA 2006 main study. The first model, made up of one reading, one science, one mathematics and two attitudinal dimensions, was used for reporting overall scores for reading, science, mathematics and two attitudinal scales. A second model, made up of one reading, one mathematics and three science dimensions, was used to generate scores for the three science scales.

The design matrix was chosen so that the partial credit model was used for items with multiple score categories and the simple logistic model was fit to the dichotomously scored items.

National calibrations

National calibrations were performed separately, country by country, using unweighted data. 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 regarded as not-administered in particular countries if it has poor psychometric characteristics in those countries but functions well in the vast majority of others. If an item is identified as behaving differently in different countries, the second option will have the same impact on inter-country comparisons.

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 (Infit Mean Square)

For each item parameter, the *ConQuest*® fit mean square statistic 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. An infit mean square greater than one is often associated with a low discrimination index, and an infit mean square less than one is often associated with a high discrimination index.

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 0.25. 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 be lower than the point-biserial correlation of 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.

National reports

After national scaling, four reports were returned to each participating country to assist in reviewing their data with the consortium.



Report 1: 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 (*see* Figure 9.1).

The first column in the table, *Label,* shows each of the possible response categories for the item. For this particular multiple-choice item, relevant categories were 1, 2, 3, 4 (the multiple-choice response categories), 8 (invalid, usually double responses) and 9 (missing).

The second column indicates the score assigned to the different categories. For this item, score 1 was allocated for the category 2 (the correct response for this multiple-choice item). Categories 1, 3, 4, 8 and 9 each received a score of 0. In this report non-reached values were treated as not administered, because this report provides information at the item calibration stage. Therefore, non-reached values are not included in this table.

The columns *Count* and *% of tot* show the number and percentage of students who responded to each category. For example, in this country, 138 students, or 38.87%, responded to *S423Q01* correctly and received score 1.

The next three columns, *Pt Bis*, *t*, and *(p)*, represent the point-biserial correlation between success on the item and a total score, the *t*-statistics associated with the point-biserial correlation and *p*-value for the *t*-statistics, respectively.

The two last columns, *PV1Avg:1* and *PV1 SD:1,* show the average ability of students responding in each category and the associated standard deviation. The average ability is calculated by domain. In this example the average ability of those students who responded correctly (category 2) is 0.12, while the average ability of those students who responded incorrectly (categories 1, 3 and 4) are –0.30, 0.07 and –0.41, respectively. Average ability of those students who selected distracter 3 for this item (0.07) is similar to the average ability of the students who selected the correct response 2. This suggests close checking of distracter three.

Figure 9.1
Example of item statistics in Report 1

```
Item:70 (S423Q01)
Cases for this item    355   Discrimination  0.13
Item Threshold(s):     0.49  Weighted MNSQ   1.17
Item Delta(s):       0.49
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t   (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0                   0       0.00     NA       NA (.000)    NA       NA
   1       0.00       65      18.31   -0.16    -3.02(.003) -0.30     0.78
   2       1.00      138      38.87    0.13     2.54(.011)  0.12     0.89
   3       0.00      115      32.39    0.09     1.76(.080)  0.07     0.83
   4       0.00       26       7.32   -0.08    -1.44(.152) -0.41     0.83
   5                   0       0.00     NA       NA (.000)    NA       NA
   6                   0       0.00     NA       NA (.000)    NA       NA
   8       0.00        4       1.13   -0.06    -1.19(.233) -0.62     0.79
   9       0.00        7       1.97   -0.15    -2.87(.004) -0.76     0.58
===============================================================================
```

Report 2: Summary of descriptive statistics by item

Report 2 provided descriptive statistics and comparisons of national and international parameters by item. An example of this report for the item *S478Q01* is shown in Figure 9.2.

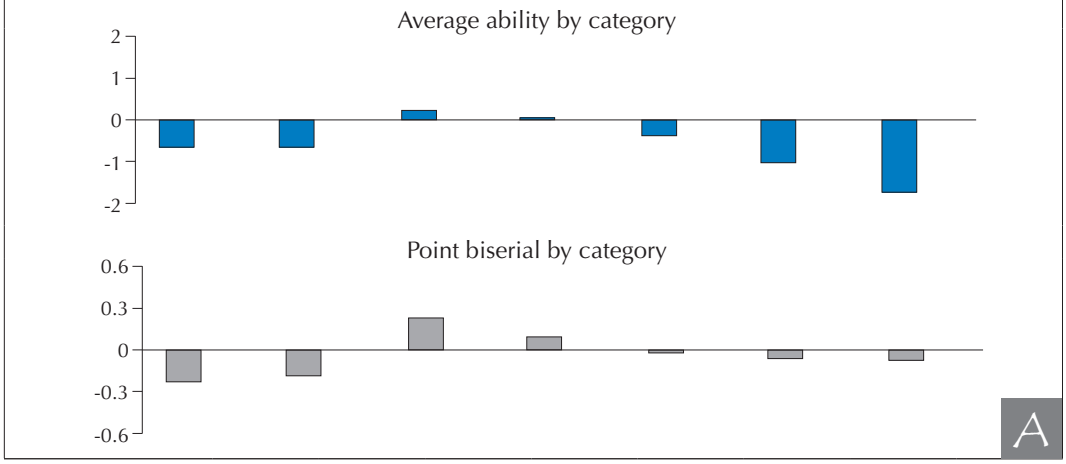


Figure 9.2
Example of item statistics in Report 2

PISA 2006 Main Study: item details, Science – *S478Q01*

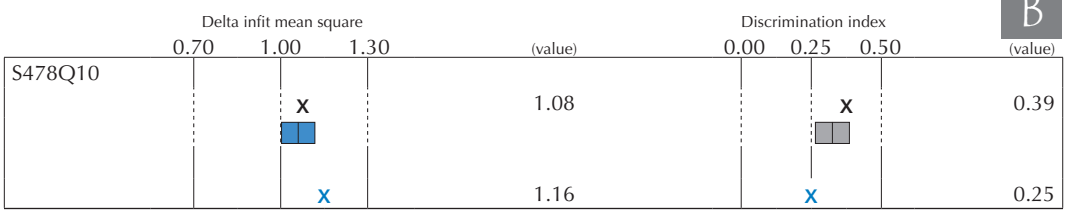
Response Frequencies

Category	1	2	3	4	8	9	r	Total
Number of students	256	145	544	467	25	9	5	1 448
Percentage	18	10	38	32	2	1	0	

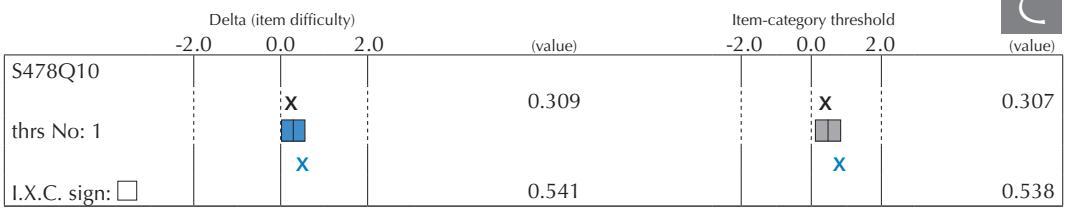


ID: S478Q10	Discrimination: 0.25
Name: Antibiotics Q1	Key: 3

B



C



D

	Item by country interactions			Discrimination				PISA 2003 link items	
	Number of valid response	Easier than expected	Harder than expected	Non-key PB is positive	Key PB is negative	Low discrimination	Ability not ordered	Link items	Requires checking
S478Q10	1 443	☐	☐	☑	☐	☐	☐	☐	☐

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



In this example, the graph marked with the letter A displays the statistics from Report 1 in a graphical form. The table above graph A shows the number and percentage of students in each response category, as shown in the columns *Label, Count* and *% of tot* in Report 1. The categories (1, 2, 3, 4, 8, 9 and *r*) are shown under each of the bar charts. An additional category, *r*, is included to indicate the number of students who did not reach the item.

The graph marked with A in Figure 9.4 facilitates the process of identifying the following anomalies:

- A non-key category has positive point-biserial or a point-biserial higher than the key category;
- A key category has a negative point-biserial;
- In the case of partial-credit items, checks can be made on whether the average ability (and the point-biserial) increases with score points.

For example, category 4 was circled by 461 students (32%) and has positive point biserial.

The initial national scaling provides the following item statistics for each country and for each item:

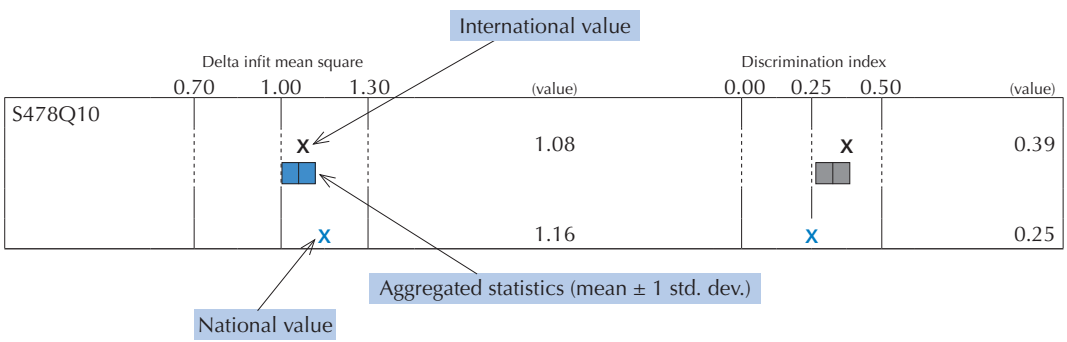
- Delta infit mean square;
- Discrimination index;
- Difficulty estimate (delta); and
- Thresholds.

Graph B (*see* Figure 9.3) and Graph C (*see* Figure 9.4) of Report 2 present the above statistics for each item in three different forms.

- National value, calculated for country;
- Average of national values across all countries (vertical line within the shaded box);
- International value calculated for all countries scaled together.

Graph B presents a comparison of the delta infit mean square statistic and the discrimination index.

Figure 9.3
Example of item statistics shown in Graph B



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



Graph C presents a comparison of the item difficulty parameters and the thresholds.

Substantial differences between the national value and the international value or the national value and the mean show that the item is behaving differently in that country in comparison with all other countries. This may be an indication of a mistranslation or some other problem.

Figure 9.4
Example of item statistics shown in Graph C

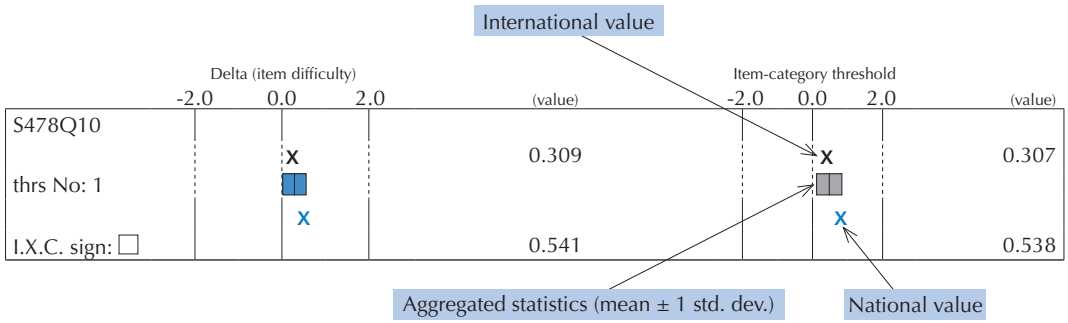


Table D (see Figure 9.5) indicates if an item is a dodgy item for the national dataset, *i.e.* an item that was flagged for one of the following reasons:

- The item difficulty is significantly lower than the average of all available countries;
- The item difficulty is significantly higher than the average of all available countries;
- 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 10 students);
- The key category point-biserial correlation is lower than –0.05;
- The item discrimination is lower than 0.2;
- The category abilities for partial credit items are not ordered;
- Link item difficulty is different from the PISA 2003 main study national item difficulty. ("Link item" box indicates if an item is a link item. "Requires checking" box is ticked when the link item performed differently in Pisa2006 main study. Only relevant to the countries that participated in both PISA cycles).

In this example item *S478Q01* was flagged as having a positive point-biserial for a non-key category.

Figure 9.5
Example of item statistics shown in Table D

	Item by country interactions			Discrimination				PISA 2003 link items	
	Number of valid response	Easier than expected	Harder than expected	Non-key PB is positive	Key PB is negative	Low discrimination	Ability not ordered	Link items	Requires checking
S478Q10	1 443	☐	☐	☑	☐	☐	☐	☐	☐

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



Report 3a: national summary of dodgy items

Report 3a summarises the dodgy items for each country as listed in report 2 section D (*see* Figure 9.6).

Figure 9.6

Example of summary of dodgy items for a country in Report 3a

PISA 2006 Main Study, Report 3a: Science dodgy items

	Item by country interactions			Discrimination				PISA 2003 link items	
	Number of valid responses	Easier than expected	Harder than expected	Non-key PB is positive	Key PB is negative	Low discrimination	Ability not ordered	Link items	Requires checking
S456Q02	1 437	☐	☐	☐	☐	☑	☐	☐	☐
S476Q01	1 482	☑	☐	☐	☐	☐	☐	☐	☐
S477Q04	1 442	☑	☐	☐	☐	☐	☐	☐	☐
S478Q01	1 443	☐	☐	☑	☐	☐	☐	☐	☐
S493Q01	1 452	☑	☐	☐	☐	☐	☐	☐	☐
S495Q01	1 442	☑	☐	☐	☐	☐	☐	☐	☐
S495Q02	1 440	☐	☑	☐	☐	☐	☐	☐	☐
S508Q02	1 435	☐	☑	☐	☐	☐	☐	☐	☐
S510Q04	1 459	☑	☐	☐	☐	☐	☐	☐	☐
S519Q01	1 438	☐	☑	☐	☐	☐	☐	☐	☐
S524Q06	1 427	☐	☑	☐	☐	☑	☐	☐	☐

Report 3b: international summary of dodgy items

Report 3b (see Figure 9.7) provided a summary of dodgy items for all countries included in the analysis. If an item showed poor psychometric properties in a country but also in most of the other countries then it could most likely be explained by reasons other than mistranslation and misprint. Note that item *S478Q01* that has been used as an example in Report 1 and Report 2 was problematic in many countries. It was easier than expected in two countries, harder in three countries, had positive point-biserial for a non-key category in 27 countries and a poor discrimination in 15 out of 58 countries.

Figure 9.7

Example of summary of dodgy items in Report 3b

PISA 2006 Main Study, Report 3: Summary of Science dodgy items – Number of countries: 58

	Item by country interactions		Discrimination				Fit	
	Easier than expected	Harder than expected	Non-key PB is positive	Key PB is negative	Low discrimination	Ability not ordered	Small, high dicr. item	Large, low discr. item
S476Q02	1	8	0	0	0	0	0	0
S476Q03	3	2	2	1	1	0	0	1
S477Q01	1	0	0	0	11	0	0	0
S477Q02	1	1	0	0	0	0	0	0
S477Q03	0	1	0	0	0	0	3	0
S477Q04	4	4	0	0	2	0	0	0
S478Q01	2	3	27	0	15	0	0	10
S478Q02	1	1	0	0	1	0	0	0
S478Q03	1	0	0	0	3	0	0	4
S478Q04	1	1	0	0	0	0	1	0
S485Q02	3	4	0	0	0	0	0	2
S485Q03	0	0	0	0	0	0	0	0
S485Q04	3	0	28	0	25	0	0	3
S485Q05	3	1	0	0	0	0	0	2
S485Q08	8	8	0	0	0	0	1	0
S493Q01	7	3	0	0	6	0	0	0
S493Q03	7	6	0	0	19	0	0	4
S493Q04	10	2	0	0	2	0	1	0



International calibration

International item parameters were set by applying the conditional item response model (9) in conjunction with the multivariate population model (15), without using conditioning variables, to a sub-sample of students. This subsample of students, referred to as the international calibration sample, consisted of 15 000 students comprising 500 students drawn at random from each of the 30 participating OECD countries[1].

The allocation of each PISA item to one of the five PISA 2006 scales is given in Appendix 1.

Student score generation

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.

Plausible values

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, , 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.

M vector-valued random deviates, $\{\varphi_{mn}\}_{m=1}^{M}$, from the multivariate normal distribution, $f_{\theta}(\theta_n; W_n, \gamma, \Sigma)$, for each case n.[2] These vectors are used to approximate the integral in the denominator of , using the Monte-Carlo integration

9.12

$$\int_{\theta} f_x(x; \xi \mid \theta) f_{\theta}(\theta, \gamma, \Sigma) d\theta \approx \frac{1}{M} \sum_{m=1}^{M} f_x(x; \xi \mid \varphi_{mn}) \equiv \Im .$$

At the same time, the values

9.13

$$p_{mn} = f_x(x; \xi \mid \varphi_{mn})\, f_{\theta}(\varphi_{mn}; W_n, \gamma, \Sigma)$$

are calculated, so that we obtain the set of pairs $\left(\varphi_{mn}, p_{mn} \big/ \Im\right)_{m=1}^{M}$, which can be used as an approximation of the posterior density [9.11]; and the probability that φ_{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; and for each random draw, the vector, φ_{ni_0}, that satisfies the condition

9.15

$$\sum_{s=1}^{i_0-1} q_{sn} < \eta_i \leq \sum_{s=1}^{i_0} q_{sn}$$

is selected as a plausible vector.



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 TIMSS (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 ID were represented by deviation contrast codes and were used as direct regressors. Each booklet was represented by one variable, except for reference booklet 11. Booklet 11 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 2003 is that with deviation contrast coding the sum of each column is zero (except for the UH booklet), whereas for simple contrast coding the sum is one. The contrast coding scheme is given in Table 0.1. In addition to the deviation contrast codes, regression coefficients between reading or mathematics and the booklet contrasts that represent booklets without mathematics 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 booklet is used as independent variable. This way, the imputation of abilities for students that did not respond to any mathematics 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 simple contrast codes for schools with the largest school as reference school. In PISA 2003 school mean performance in the major domain was used as regressor instead of contrast codes to simplify the model. The intra-class correlation was generally slightly higher in PISA 2006 than in PISA 2003, which is likely to be caused by using school dummy coding instead of school performance means. As expected, using school means slightly underestimates the between school variance.

All other categorical variables from the student, ICT and parent questionnaire were dummy coded. These dummy variables and all numeric variables (the questionnaire indices) were analysed in a principle component analysis. The details of recoding the variables before the principle component analysis are listed in Appendix 2. 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.

Table 9.1
Deviation contrast coding scheme

	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	UH
Booklet 1	1	0	0	0	0	0	0	0	0	0	0	0	0
Booklet 2	0	1	0	0	0	0	0	0	0	0	0	0	0
Booklet 3	0	0	1	0	0	0	0	0	0	0	0	0	0
Booklet 4	0	0	0	1	0	0	0	0	0	0	0	0	0
Booklet 5	0	0	0	0	1	0	0	0	0	0	0	0	0
Booklet 6	0	0	0	0	0	1	0	0	0	0	0	0	0
Booklet 7	0	0	0	0	0	0	1	0	0	0	0	0	0
Booklet 8	0	0	0	0	0	0	0	1	0	0	0	0	0
Booklet 9	0	0	0	0	0	0	0	0	1	0	0	0	0
Booklet 10	0	0	0	0	0	0	0	0	0	1	0	0	0
Booklet 11	–1	–1	–1	–1	–1	–1	–1	–1	–1	–1	–1	–1	0
Booklet 12	0	0	0	0	0	0	0	0	0	0	1	0	0
Booklet 13	0	0	0	0	0	0	0	0	0	0	0	1	0
UH	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**



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.

Two models were run, each with five dimensions. The first model included mathematics, reading, science, interest and support. The second model included mathematics, reading and the three science scales. For each domain plausible values were drawn using the method described in the *PISA 2003 Technical Report* (OECD, 2005).

BOOKLET EFFECTS

As with PISA 2003, the PISA 2006 test design was balanced, 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 PISA 2000 and PISA 2003.

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[3].

The calibration model given above was used to estimate the international item parameters. It was estimated using the international calibration sample of 15 000 students, and not-reached items in the estimation were treated as not administered.

The booklet parameters obtained from this analysis were not used to correct for the booklet effect. Instead, a set of booklet parameters was obtained by scaling the entire data set of OECD countries using booklet as a conditioning variable and a senate weight. The students who responded to the UH booklet were excluded from the estimation. The booklet parameter estimates obtained are reported in Chapter 12. The booklet effects are the amount that must be added to the proficiencies of students who responded to each booklet.

To correct the student scores for the booklet effects, two alternatives were considered:

- To correct all students' scores using one set of the internationally estimated booklet parameters; or
- To correct 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.

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 (17). As such, plausible values contain random error variance components and are not 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).

The PISA student file contains 45 plausible values, five for each of the eight PISA 2006 scales. *PV1MATH* to *PV5MATH* are for mathematical literacy; *PV1SCIE* to *PV5SCIE* for *scientific literacy*, *PV1READ* to *PV5READ* for *reading literacy, PV1INTR* to *PV5INTR* for *interest in science* and *PV1SUPP* to *PV5SUPP* for *support for scientific inquiry*. For the three scientific literacy scales, *explaining phenomena scientifically*, *identifying scientific issues, using scientific evidence,* the plausible values variables are *PV1SCIE1* to *PV5SCIE1*, *PV1SCIE2* to *PV5SCIE2, and PV1SCIE3* to *PV5SCIE3,* respectively.

If an analysis were to be undertaken with one of these eight scales, 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 $r(\theta, \mathbf{Y})$ is a statistic that depends upon the latent variable and some other observed characteristic of each student. That is: $(\theta, Y) = (\theta_1, y_1, \theta_2, y_2, \ldots, \theta_N, y_N)$ 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 $h_\theta(\theta_n; y_n, \xi, \gamma, \Sigma \mid x_n)$. If $h_\theta(\theta; Y, \xi, \gamma, \Sigma \mid X)$ is the joint marginal posterior for $n = 1, \ldots N$ then we can compute:

9.16

$$r^*(X, Y) = E[r^*(\theta, Y)|X, Y]$$

$$= \int_\theta r(\theta, Y) h_\theta(\theta; Y, \xi, \gamma, \Sigma \,|X) d\theta$$

.

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



The integral in can be computed using the Monte-Carlo method. If M random vectors $(\theta_1, \theta_2, \ldots, \theta_M)$ are drawn from $h_\theta(\theta; Y, \xi, \gamma, \Sigma \mid X)$ is approximated by:

9.17

$$r^*(X,Y) \approx \frac{1}{M}\sum_{m=1}^{M} r(\theta_m, Y)$$

$$= \frac{1}{M}\sum_{m=1}^{M} \hat{r}_m$$

where $\hat{r}_m$ is the estimate of r computed using the m-th set of plausible values.

From **[9.16]** we can see that the final estimate of r is the average of the estimates computed using each plausible value in turn. If U_m is the sampling variance for $\hat{r}_m$ then the sampling variance of r^* is:

9.18

$$V = U^* + (1+M^{-1})B_M,$$

where $U^* = \frac{1}{M}\sum_{m=1}^{M} U_m$ and $B_M = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{r}_m - r^*)^2$.

An α-% confidence interval for r^* is $r^* \pm t_\nu\left(\frac{(1-\alpha)}{2}\right)V^{1/2}$ where $t_\upsilon(s)$ is the s- percentile of the t-distribution with ν degrees of freedom. $\upsilon = \left[\frac{f_M^2}{M-1} + \frac{(1-f_M)^2}{d}\right]^{-1}$, $f_M = (1 + M^{-1})B_M/V$ 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.

DEVELOPING COMMON SCALES FOR THE PURPOSES OF TRENDS

The reporting scales that were developed for each 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 27 OECD countries that participated in PISA 2000 that had acceptable response rates (Wu & Adams, 2002).[4]

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 – that is, the mean score in 2000 of the sampled students in the 27 OECD countries that participated in PISA 2000.[5]

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.[6]

For PISA 2006 the decision was made to report the reading on these previously developed scales. That is the reading reporting scales used for PISA2000, 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.[7]

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.

Further details on the various PISA reporting scales are given in Chapter 12.

Linking PISA 2003 and PISA 2006 for reading and mathematics

The linking of PISA 2006 reading and mathematics to the existing scales was undertaken using standard common item equating methods.

The steps involved in linking the PISA 2003 and PISA 2006 reading and mathematics scales were as follows:

Step 1: Item parameter estimates for reading and mathematics where obtained from the PISA 2006 calibration sample.

Step 2: The above item parameters estimates where transformed through the addition of constant, so that the mean of the item parameter estimates for the common items was the same in 2006 as it was in 2003.

Step 3: The 2006 student abilities where estimated with item parameters anchored at their 2006 values.

Step 4: The above estimated students abilities where transformed with the shift estimated in step 2.

Note that this is a much simpler procedure than the 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 for both PISA 2003 and 2006.

Uncertainty in the link

In each case the transformation that equates the 2006 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 in2000 and let $\hat{\delta}_i^{2003}$ be the estimated difficulty of link i in2003, where the mean of the two sets difficulty estimates for all of the link items for a domain is set at zero. We now define the value:

$c_i = \hat{\delta}_i^{2003} - \hat{\delta}_i^{2000}$.



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:

$$error_{2000,2003} = \sqrt{\frac{1}{L}\sum c_i^2}.$$

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 a 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

$$c_{ij} = \hat{\delta}_{ij}^{2006} - \hat{\delta}_{ij}^{2003}.$$

The size (total number of score points) of unit j is m_j so that:

$$\sum_{j=1}^{K} m_j = L \quad \text{and} \quad \bar{m} = \frac{1}{L}\sum_{j=1}^{K} m_j.$$

Further let:

$$c_{\bullet j} = \frac{1}{m_j}\sum_{i=1}^{m_j} c_{ij} \quad \text{and} \quad \bar{c} = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{m_j} c_{ij}$$

and then the link error, taking into account the clustering is as follows:

$$error_{2006,2003} = \sqrt{\frac{\sum_{j=1}^{K} m_j^2 (c_{\bullet j} - \bar{c})^2}{K(K-1)\bar{m}^2}}.$$

The link standard errors are reported in chapter 12.

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 2003 and PISA 2006) in reading resulted in an over-estimation of student scores by two points on the PISA 2003 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 2006 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 2006 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 (eg males and females) in PISA 2006 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 (eg a country) between PISA 2003 and PISA 2006 will be influenced because each student's score in PISA 2003 will be influenced by the error; and finally;
- A change in the difference in performance between two groups from PISA 2003 to PISA 2006 will not be influenced. This is because neither of the components of this comparison, which are differences in scores in 2006 and 2003 respectively, is influenced by a common error that is added to all student scores in PISA 2006.

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.

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 2006 of the performance of Canada in mathematics. The mean performance of Canada in 2003 was 532 with a standard error of 1.8, while in 2006 the mean was 527 with a standard error of 2.0. The standardised difference in the Canadian mean is -1.82, which is computed as follows: $-1.82 = (527 - 532)/\sqrt{2.0^2 + 1.8^2 + 1.4^2}$, and is not statistically significant.



Notes

1. 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.

2. The value *M* should be large. For PISA we have used 2000.

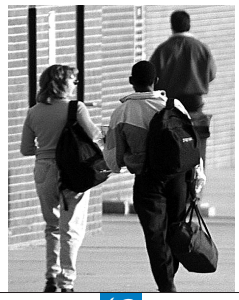
3. Note that because the design was balanced the inclusion of the booklet term in the item response model did not have an appreciable effect on the item parameter estimates.

4. Using senate weights.

5. Again using senate weights.

6. Again using senate weights.

7. Again using senate weights.



10

Data Management Procedures

Introduction 164
KeyQuest 167
Data management at the national centre 167
- National modifications to the database 167
- Student sampling with *KeyQuest* 167
- Data entry quality control 167

Data cleaning at ACER 171
- Recoding of national adaptations 171
- Data cleaning organisation 171
- Cleaning reports 171
- General recodings 171

Final review of the data 172
- Review of the test and questionnaire data 172
- Review of the sampling data 172

Next steps in preparing the international database 172



INTRODUCTION

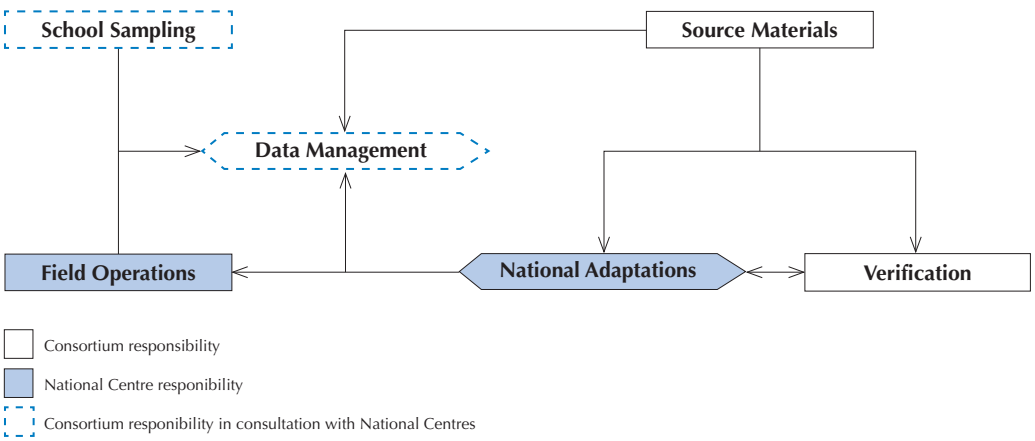
The PISA assessment establishes certain data collection requirements that are common to all PISA participants. Test instruments include the same test items in all participating countries, and data collection procedures are applied in a common and consistent way amongst all participants to help ensure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

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

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

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

Figure 10.1
Data management in relation to other parts of PISA



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



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

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

Figure 10.1 illustrates the sequence of major data management tasks in PISA, and shows something of the division of responsibilities between national centres, the consortium, and those tasks that involve negotiation between the two. This section briefly introduces each of the tasks. More details are provided in the following sections.

First, ACER provides the data management software *KeyQuest* to all national centres. *KeyQuest* is generic software that can be configured to meet a variety of data entry requirements. In addition to its generic features, the latest version of *KeyQuest* was pre-configured specifically for PISA 2006.

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

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

As soon as data are submitted to ACER, additional checks are applied. During the process of data cleaning, ACER sends cleaning reports containing the results of the checking procedures to national centres, and asks national centres to clarify any inconsistencies in their database. The national data sets are then continuously updated according to the information provided by the national centres. The cleaning reports are described in more detail below.

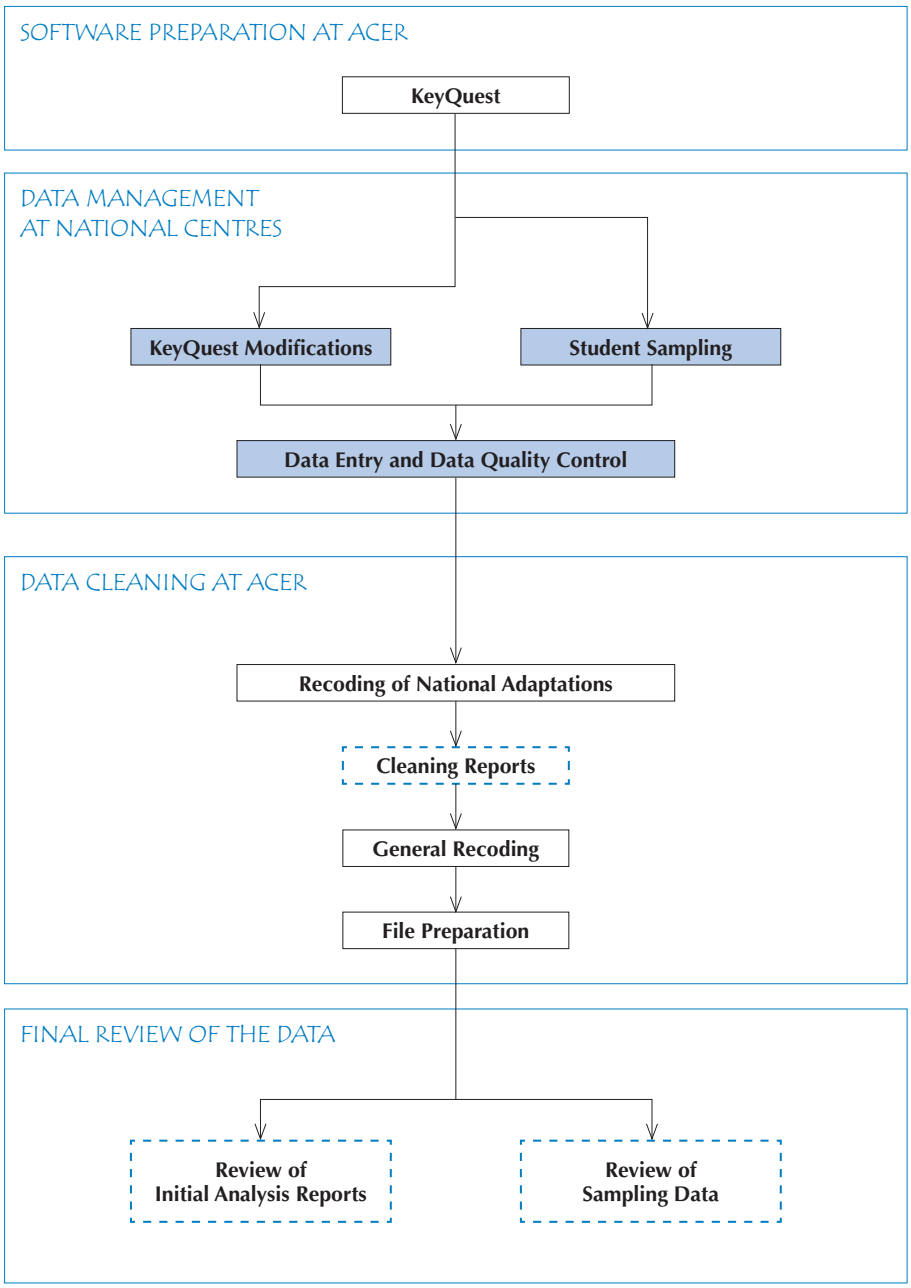
Once ACER has received all cleaning reports from the national centres and has introduced into the database all corrections recommended in these reports, a number of general rules are applied to the small number of unresolved inconsistencies in the PISA database.

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

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



Figure 10.2
Major data management stages in PISA



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



KEYQUEST

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

KeyQuest was preconfigured with all the PISA 2006 standard instruments: cognitive test booklets, background and contextual questionnaires, and student tracking instruments that are derived following implementation of the school sampling procedures. However, it also allows for instrument modifications such as addition of national questions, deletion of some questions and modification of some questions. A prerequisite for national modification of *KeyQuest* is consortium approval of proposed national adaptations.

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

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

DATA MANAGEMENT AT THE NATIONAL CENTRE

National modifications to the database

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

Student sampling with KeyQuest

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

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

Data entry quality control

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

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



Validation rules

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

Key violations

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

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

Monitoring of the data entry operators

The data entry efficiency report was designed specifically for PISA 2006 to keep the count of records entered by each data entry operator and the time required to enter them. The consortium recommended to all countries to use some part of these procedures (as appropriate) to assure quality of the data entry.

Double entry facilities

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

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

10.1

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

and the results are shown in Table 10.1.

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

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



Table 10.1
Double entry discrepancies per country: field trial data

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

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

Double coding of occupational data

Another new optional procedure for PISA 2006 was the double coding of occupational data. The double coding allowed national centres a check of the validity of the data and it allowed identification of the areas where supplementary coding tools could be improved. The main coding tool was the *ISCO Manual* (ILO, 1990) with the small number of additional codes described in the PISA 2006 *Data Management Manua*[1]. The supplementary coding tools would typically include coding instructions, a coding index, and training materials developed at the national centre.

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

The national centres that participated in this option commented on the usefulness of the procedures for training of the coding staff. The possibilities for analysis by the consortium of the data from this option were limited due to the language constraints. One of the results was that those countries that required their coders to enter a word description as well as four-digit code had fewer discrepancies than those that required only a four-digit code. When analysing the double entry discrepancy reports from the English speaking countries the consortium found that when one of two coders entered both the description and code while another entered the code only, the discrepancy was mostly due to the second coder being incorrect. This led to a reinforcement of the ILO recommendation that procedures should involve entering occupation descriptions first and then coding them, rather than coding directly from the questionnaires.

Validity reports

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



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

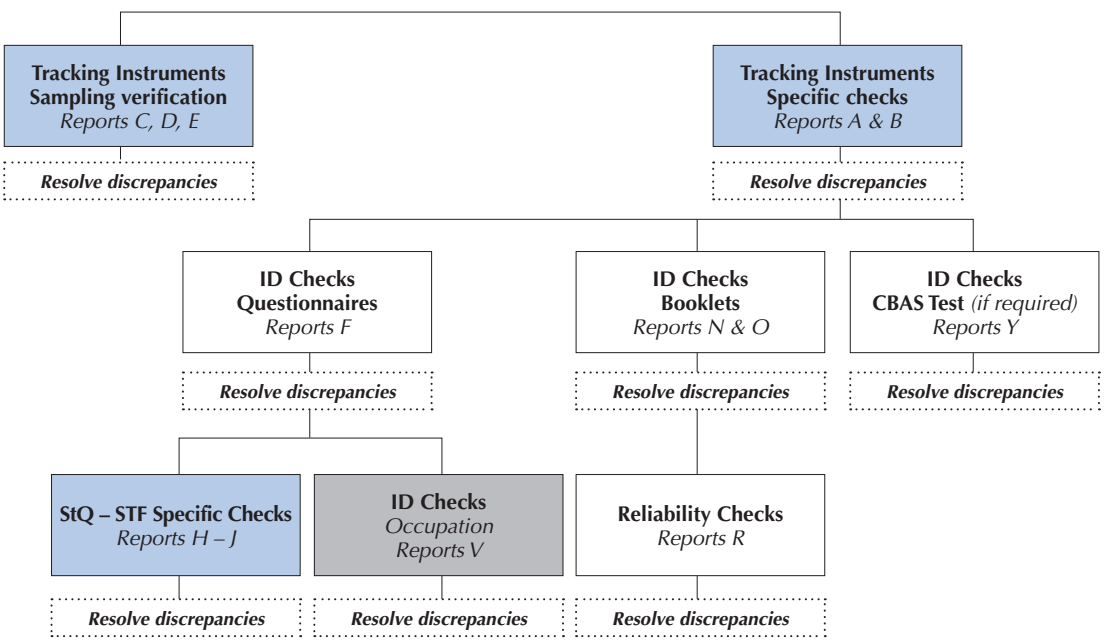
- Comparison between tracking instruments and sampling verification (tracking instruments, sampling verification);
- Data verification within tracking instruments (tracking instruments specific checks);
- Comparison of the questionnaire and tracking data (STQ-STF specific checks, ID checks questionnaires, ID checks occupation);
- Comparison of the identification variables in the test data (ID checks booklets, ID checks CBAS);
- Verification of the reliability data (reliability checks).

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

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

Due to the complexity and significant number of the validity reports, a validity report checklist was designed.

Figure 10.3
Validity reports – general hierarchy



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



DATA CLEANING AT ACER

Recoding of national adaptations

When data submitted by national centres arrived at ACER, the first step was to check the consistency of the database structure with the international database structure. An automated procedure was developed for this purpose. For each instrument the procedure identified deleted variables, added variables and variables for which the validation rules had been changed.

This report was then compared with the information provided by the NPM in the various adaptation spreadsheets such as the questionnaire adaptation sheet (*see* Chapter 3). For example, if a variable had been added to a questionnaire, the questionnaire adaptation sheet was checked to find out whether this national variable require recoding into the corresponding international one, or had to be set aside as being for purely national use and returned to the country.

Once all deviations were checked, the submitted data were recoded where necessary to fit the international structure. All additional or modified variables were set aside and returned to the national centres in a separate file so that countries could use these data for their own purposes, but they were not included in the international database.

Data cleaning organisation

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

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

Cleaning reports

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

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

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

General recodings

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



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

FINAL REVIEW OF THE DATA

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

Review of the test and questionnaire data

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

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

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

Review of the sampling data

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

NEXT STEPS IN PREPARING THE INTERNATIONAL DATABASE

When all data management procedures described in this chapter were complete, the database was ready for the next steps in preparing the public international database. Students weights and replicated weights

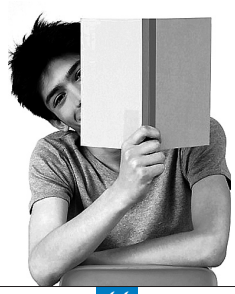


were created as described in Chapter 8. Questionnaire indices were computed or scaled as described in Chapter 16. Cognitive item responses were scaled to obtain international item parameters that were used to draw plausible values as student ability estimates (*see* Chapters 9 and 12).

Notes

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

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



11

Sampling Outcomes

Design effects and effective sample sizes 187
- Variability of the design effect 191
- Design effects in PISA for performance variables 191

Summary analyses of the design effect 203
- Countries with outlying standard errors 205

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



This chapter reports on PISA sampling outcomes. Details of the sample design are given in Chapter 4.

Table 11.1 shows the various 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.

Indices 1, 2 and 3 are intended to measure PISA population coverage. Indices 4 and 5 are intended to be diagnostic in cases where indices 1, 2 or 3 have unexpected values. Many references are made in this chapter to the various sampling forms on which NPMs documented statistics and other information needed in undertaking the sampling.

Index 1: Coverage of the national population, calculated by $P/(P+E) \times 3[c]/3[a]$:

- The national population (NP), defined by sampling form 3 response box [a] and denoted here as 3[a] (and in Table 11.1 as target populationis the population that includes all enrolled 15-year-olds in grades 7 and above in each country (with the possibility of small levels of exclusions), based on national statistics. However, the final NP reflected on each country's school sampling frame might have had some school-level exclusions. The value that represents the population of enrolled 15-year-olds minus those in excluded schools is represented initially by response box [c] on sampling form 3. It is denoted here as 3[c]. As in PISA 2003, the procedure for PISA 2006 was that very small schools having only one or two eligible students could not be excluded from the school frame but could be excluded in the field if they still had exactly only one or two eligible students at the time of data collection. Therefore, what is noted in index 1 as 3[c] (and in Table 11.1 as target minus school level exclusions is a number that excludes schools excluded from the sampling frame in addition to those schools excluded in the field. Thus, the term 3[c]/3[a] 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 eligible 15-year-olds in each country, where P is the weighted estimate of eligible non-excluded 15-year-olds and E is the weighted estimate of eligible 15-year-olds that were excluded within schools. Therefore, the term $P/(P+E)$ provides an estimate, based on the student sample, of the proportion of the eligible 15-year-old population represented by the non-excluded eligible 15-year-olds;
- Thus the result of multiplying these two proportions together (3[c]/3[a] and $P/(P+E)$) indicates the overall proportion of the NP covered by the non-excluded portion of the student sample.

Index 2: Coverage of the national enrolled population, calculated by $P/(P+E) \times 3[c]/2[b]$:

- The national enrolled population (NEP), defined by sampling form 2 response box [b] and denoted here as 2[b] (and as enrolled 15-year-oldsin Table 11.1), is the population that includes all enrolled 15-year-olds in grades 7 and above in each country, based on national statistics. The final NP, denoted here as 3[c] 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-olds 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 eligible non-excluded 15-year-olds and E is the weighted estimate of eligible 15-year-olds that were excluded within schools. Therefore, the term $P/(P+E)$ provides an estimate based on the student sample of the proportion of the eligible 15-year-old population that is represented by the non-excluded eligible 15-year-olds;
- Multiplying these two proportions together (3[c]/2[b] and $P/(P+E)$) gives the overall proportion of the NEP that is covered by the non-excluded portion of the student sample.

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



Index 1 shows the extent to which the weighted participants cover the final target population after all school exclusions.

Index 2 shows the extent to which the weighted participants cover the target population of all enrolled students in grades 7 and above.

Index 1 and Index 2 will differ when countries have excluded geographical areas or language groups apart from other school level exclusions.

Index 3: Coverage of the national 15-year-old population, calculated by *P*/2[a]:

- The national population of 15-year-olds, defined by sampling form 2 response box [a] and denoted here as 2[a] (and called all 15-year-oldsin 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 eligible non-excluded 15-year-olds from the student sample. Thus *P*/2[a] indicates the proportion of the national population of 15-year-olds covered by the non-excluded portion of the student sample;

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 eligible 15-year-olds in each country, where *P* is the weighted estimate of eligible non-excluded 15-year-olds and *E* is the weighted estimate of eligible 15-year-olds who were excluded within schools;
- The value *S* is an estimate of the 15-year-old school population in each country (called estimate of enrolled students on frame in Table 11.1). This is based on the actual or (more often) approximate number of 15-year-olds 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-olds (*ENR*) as recorded on the school sampling frame. In the infrequent case where the *ENR* value was not available, the number of 15-year-olds from the student tracking form was used;
- 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 eligible 15-year-olds. 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 Index 5.

Index 5: Coverage of the school sampling frame population, calculated by *S*/3[c]:

- The value *S*/3[c] 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 form 3 and adjusted by removing any additional excluded schools in the field. In some cases, this provides a check as to whether the data on the sampling frame give a reliable estimate of the number of 15-year-olds in each school. In other cases, however, it is evident that 3[c] has been derived using data from the sampling frame by the National Project Manager, so that this ratio may be close to 1.0 even if enrolment data on the school sampling frame are poor. Under such circumstances, Index 4 will differ noticeably from 1.0, and the figure for 3[c] will also be inaccurate.

Tables 11.2, 11.3, 11.4 present school and student-level response rates.

Table 11.2 indicates the rates calculated by using only original schools and no replacement schools. Table 11.3 indicates the improved response rates when first and second replacement schools were accounted for in the rates. Table 11.4 indicates the student response rates among the full set of participating schools.



Table 11.1 [Part 1/3]
Sampling and coverage rates

		All 15-year-olds	Enrolled 15-year-olds	Target population	School level exclusions	Target minus school level exclusions	% school level exclusions	Estimate of enrolled students on frame	Participants		Excluded	
									Actual	Weighted	Actual	Weighted
OECD	Australia	270 115	256 754	255 554	1 371	254 183	0.54	251 221.74	14 170	234 939.52	234	2 934.61
	Austria	97 337	92 149	92 149	401	91 748	0.43	92 606.34	4 927	89 925.11	94	1 585.63
	Belgium	124 943	124 557	124 216	2 957	121 259	2.38	123 596.62	8 857	123 161.45	28	401.21
	Belgium-Flanders	69 650	68 662	68 321	1 201	67 120	1.76	67 048.31	5 124	69 409.16	16	214.53
	Canada	426 967	428 876	424 238	5 141	419 097	1.21	418 565.11	22 646	370 879.36	1 681	20 339.28
	Czech Republic	127 748	124 764	124 764	1 124	123 640	0.90	125 258.79	5 932	128 827.19	8	202.51
	Denmark	66 989	65 984	65 984	1 871	64 113	2.84	57 156.10	4 532	57 012.63	170	1 960.32
	Finland	66 232	66 232	66 232	1 257	64 975	1.90	65 085.51	4 714	61 386.99	135	1 649.63
	France	809 375	809 375	777 194	19 397	757 797	2.50	757 511.93	4 716	739 428.06	28	3 876.20
	Germany	951 535	1 062 920	1 062 920	6 009	1 056 911	0.57	950 350.10	4 891	903 512.45	37	6 016.55
	Greece	107 505	110 663	110 663	640	110 023	0.58	104 827.25	4 873	96 411.69	65	1 396.91
	Hungary	124 444	120 061	120 061	3 230	116 831	2.69	114 424.54	4 490	106 010.05	31	1 103.26
	Ireland	58 667	57 648	57 510	50	57 460	0.09	57 245.39	4 585	55 114.26	93	937.20
	Italy	578 131	639 971	639 971	16	639 955	0.00	623 569.70	21 773	520 055.20	363	8 984.12
	Italy-Basilicata	7 071	8 404	8 404	0	8 404	0.00	7 736.12	1 507	6 422.46	9	41.91
	Italy-Bolzano	5 314	5 116	5 116	0	5 116	0.00	4 917.44	2 084	4 654.76	28	56.81
	Italy-Campania	76 596	80 108	80 108	0	80 108	0.00	79 658.99	1 406	67 443.20	9	323.03
	Italy-Emilia Romagna	31 879	35 926	35 926	0	35 926	0.00	35 160.37	1 531	29 500.54	34	569.50
	Italy-Friuli Venezia Giulia	9 312	10 277	10 277	0	10 277	0.00	10 123.28	1 578	8 534.10	15	84.38
	Italy-Liguria	11 739	13 839	13 839	16	13 823	0.12	13 061.63	1 753	11 747.49	45	222.09
	Italy-Lombardia	81 088	89 897	89 897	0	89 897	0.00	88 462.73	1 524	69 524.95	40	1 913.41
	Italy-Piemonte	35 309	39 070	39 070	0	39 070	0.00	38 250.67	1 478	34 069.59	31	717.74
	Italy-Puglia	48 518	50 168	50 168	0	50 168	0.00	48 922.23	1 540	45 333.52	10	351.27
	Italy-Sardegna	17 297	19 564	19 564	0	19 564	0.00	19 280.96	1 390	16 136.50	16	218.57
	Italy-Sicilia	63 369	68 146	68 146	0	68 146	0.00	66 178.54	1 354	54 116.13	28	1 135.19
	Italy-Trento	4 821	5 653	5 653	0	5 653	0.00	5 391.76	1 757	4 316.52	42	71.45
	Italy-Veneto	41 926	49 511	49 511	0	49 511	0.00	48 677.17	1 530	40 070.67	34	852.25
	Japan	1 246 207	1 222 171	1 222 171	16 604	1 205 567	1.36	1 182 687.63	5 952	1 113 700.93	0	0.00
	Korea	660 812	627 868	627 868	3 461	624 407	0.55	576 636.64	5 176	576 669.37	4	624.93
	Luxembourg	4 595	4 595	4 595	0	4 595	0.00	4 955.00	4 567	4 733.00	193	193.00
	Mexico	2 200 916	1 383 364	1 383 364	0	1 383 364	0.00	1 342 897.79	30 971	1 190 420.04	49	3 217.25
	Netherlands	197 046	193 769	193 769	57	193 712	0.03	199 533.05	4 871	189 575.82	7	226.95
	New Zealand	63 800	59 341	59 341	451	58 890	0.76	59 089.52	4 823	53 397.58	222	2 134.96
	Norway	61 708	61 449	61 373	412	60 961	0.67	60 368.65	4 692	59 884.49	156	1 764.49
	Poland	549 000	546 000	546 000	10 400	535 600	1.90	532 060.81	5 547	515 992.95	18	1 684.94
	Portugal	115 426	100 816	100 816	0	100 816	0.00	99 961.25	5 109	90 078.87	112	1 889.87
	Slovak Republic	79 989	78 427	78 427	1 355	77 072	1.73	76 671.38	4 731	76 200.83	11	193.02
	Spain	439 415	436 885	436 885	3 930	432 955	0.90	423 903.57	19 604	381 685.95	557	10 386.16
	Spain-La Rioja	2 737	2 619	2 619	11	2 608	0.42	2 641.00	1 333	2 494.35	56	107.08
	Spain-Basque Country	16 820	17 967	17 967	42	17 925	0.23	15 753.72	3 929	14 706.61	81	294.97
	Spain-Navarra	5 298	4 903	4 903	20	4 883	0.41	4 952.20	1 590	4 677.66	37	98.12
	Spain-Galicia	24 269	26 420	26 420	90	26 330	0.34	23 724.51	1 573	22 577.66	32	445.25
	Spain-Catalonia	63 240	61 491	61 491	683	60 808	1.11	61 213.50	1 527	56 987.17	62	2 147.44
	Spain-Castilla y Leon	22 011	24 089	24 089	111	23 978	0.46	21 852.57	1 512	19 697.15	64	784.65
	Spain-Cantabria	4 912	5 215	5 215	25	5 190	0.48	4 751.33	1 496	4 534.16	56	154.06
	Spain-Asturias	8 101	9 484	9 484	32	9 452	0.34	7 983.50	1 579	7 593.57	39	200.23
	Spain-Aragon	11 112	11 150	11 150	67	11 083	0.60	10 594.50	1 526	9 467.26	37	193.67
	Spain-Andalucia	93 709	93 188	93 188	335	92 853	0.36	90 552.40	1 463	81 437.14	29	1 444.61
	Sweden	129 734	127 036	127 036	2 330	124 706	1.83	127 133.27	4 443	126 392.73	122	3 470.95
	Switzerland	87 766	86 108	86 108	2 130	83 978	2.47	81 660.28	12 193	89 650.91	186	842.40
	Turkey	1 423 514	800 968	782 875	970	781 905	0.12	796 371.42	4 942	665 477.29	1	130.38
	United Kingdom	779 076	767 248	767 248	12 879	754 369	1.68	748 795.67	13 152	732 003.69	229	12 032.64
	United Kingdom-Scotland	63 245	63 087	63 087	867	62 220	1.37	63 655.81	2 444	57 332.35	95	1 691.42
	United States	4 192 939	4 192 939	4 192 939	19 710	4 173 229	0.47	3 901 130.57	5 611	3 578 039.60	254	142 517.21



Table 11.1 [Part 2/3]
Sampling and coverage rates

		Ineligible		Eligible		Within school exclusions (%)[1]	Overall exclusions (%)	Ineligible (%)	Coverage Indices				
		Actual	Weighted	Actual	Weighted				1	2	3	4	5
OECD	Australia	877	9 737.48	17 062	237 874.13	1.23	1.76	4.09	0.98	0.98	0.87	0.95	0.99
	Austria	197	3 103.30	5 642	91 510.74	1.73	2.16	3.39	0.98	0.98	0.92	0.99	1.01
	Belgium	134	2 966.90	9 520	123 562.66	0.32	2.70	2.40	0.97	0.97	0.99	1.00	1.02
	Belgium-Flanders	64	813.51	5 429	69 623.69	0.31	2.06	1.17	0.98	0.97	1.00	1.04	1.00
	Canada	1 715	23 784.08	29 143	391 218.64	5.20	6.35	6.08	0.94	0.93	0.87	0.93	1.00
	Czech Republic	42	895.68	6 583	129 029.70	0.16	1.06	0.69	0.99	0.99	1.01	1.03	1.01
	Denmark	126	1 433.58	5 255	58 972.95	3.32	6.07	2.43	0.94	0.94	0.85	1.03	0.89
	Finland	48	588.79	5 217	63 036.62	2.62	4.47	0.93	0.96	0.96	0.93	0.97	1.00
	France	87	12 158.23	5 326	743 304.26	0.52	3.00	1.64	0.97	0.93	0.91	0.98	1.00
	Germany	65	10 781.53	5 353	909 529.01	0.66	1.22	1.19	0.99	0.99	0.95	0.96	0.90
	Greece	69	1 477.14	5 186	97 808.60	1.43	2.00	1.51	0.98	0.98	0.90	0.93	0.95
	Hungary	93	2 233.76	4 854	107 113.31	1.03	3.69	2.09	0.96	0.96	0.85	0.94	0.98
	Ireland	118	1 206.67	5 562	56 051.46	1.67	1.76	2.15	0.98	0.98	0.94	0.98	1.00
	Italy	814	20 363.44	23 874	529 039.32	1.70	1.70	3.85	0.98	0.98	0.90	0.85	0.97
	Italy-Basilicata	49	186.44	1 615	6 464.37	0.65	0.65	2.88	0.99	0.99	0.91	0.84	0.92
	Italy-Bolzano	48	109.53	2 244	4 711.57	1.21	1.21	2.32	0.99	0.99	0.88	0.96	0.96
	Italy-Campania	106	4 406.00	1 561	67 766.23	0.48	0.48	6.50	1.00	1.00	0.88	0.85	0.99
	Italy-Emilia Romagna	32	598.93	1 673	30 070.05	1.89	1.89	1.99	0.98	0.98	0.93	0.86	0.98
	Italy-Friuli Venezia Giulia	29	157.17	1 689	8 618.48	0.98	0.98	1.82	0.99	0.99	0.92	0.85	0.99
	Italy-Liguria	69	392.05	1 960	11 969.57	1.86	1.97	3.28	0.98	0.98	1.00	0.92	0.94
	Italy-Lombardia	49	1 768.57	1 681	71 438.37	2.68	2.68	2.48	0.97	0.97	0.86	0.81	0.98
	Italy-Piemonte	30	574.96	1 611	34 787.33	2.06	2.06	1.65	0.98	0.98	0.96	0.91	0.98
	Italy-Puglia	64	1 563.12	1 660	45 684.79	0.77	0.77	3.42	0.99	0.99	0.93	0.93	0.98
	Italy-Sardegna	69	710.73	1 585	16 355.07	1.34	1.34	4.35	0.99	0.99	0.93	0.85	0.99
	Italy-Sicilia	135	4 774.93	1 544	55 251.32	2.05	2.05	8.64	0.98	0.98	0.85	0.83	0.97
	Italy-Trento	52	104.17	1 913	4 387.97	1.63	1.63	2.37	0.98	0.98	0.90	0.81	0.95
	Italy-Veneto	47	1 448.70	1 638	40 922.92	2.08	2.08	3.54	0.98	0.98	0.96	0.84	0.98
	Japan	408	75 104.30	5 971	1 113 700.93	0.00	1.36	6.74	0.99	0.99	0.89	0.94	0.98
	Korea	44	4 915.09	5 233	577 294.30	0.11	0.66	0.85	0.99	0.99	0.87	1.00	0.92
	Luxembourg	29	29.00	4 926	4 926.00	3.92	3.92	0.59	0.96	0.96	1.03	0.99	1.08
	Mexico	4 623	166 614.35	32 409	1 193 637.29	0.27	0.27	13.96	1.00	1.00	0.54	0.89	0.97
	Netherlands	89	3 738.26	5 437	189 802.77	0.12	0.15	1.97	1.00	1.00	0.96	0.95	1.03
	New Zealand	299	2 847.56	5 757	55 532.54	3.84	4.58	5.13	0.95	0.95	0.84	0.94	1.00
	Norway	30	333.93	5 501	61 648.98	2.86	3.51	0.54	0.96	0.96	0.97	1.02	0.99
	Poland	20	1 568.40	6 092	517 677.89	0.33	2.22	0.30	0.98	0.98	0.94	0.97	0.99
	Portugal	362	5 696.82	6 013	91 968.74	2.05	2.05	6.19	0.98	0.98	0.78	0.92	0.99
	Slovak Republic	40	622.22	5 112	76 393.85	0.25	1.98	0.81	0.98	0.98	0.95	1.00	0.99
	Spain	273	4 821.40	21 885	392 072.12	2.65	3.52	1.23	0.96	0.96	0.87	0.92	0.98
	Spain-La Rioja	13	22.58	1 530	2 601.42	4.12	4.52	0.87	0.95	0.95	0.91	0.99	1.01
	Spain-Basque Country	77	286.27	4 164	15 001.58	1.97	2.20	1.91	0.98	0.98	0.87	0.95	0.88
	Spain-Navarra	14	43.20	1 734	4 775.78	2.05	2.45	0.90	0.98	0.98	0.88	0.96	1.01
	Spain-Galicia	24	328.18	1 704	23 022.91	1.93	2.27	1.43	0.98	0.98	0.93	0.97	0.90
	Spain-Catalonia	21	706.33	1 726	59 134.61	3.63	4.70	1.19	0.95	0.95	0.90	0.97	1.01
	Spain-Castilla y Leon	22	273.33	1 700	20 481.81	3.83	4.27	1.33	0.96	0.96	0.89	0.94	0.91
	Spain-Cantabria	26	72.12	1 692	4 688.22	3.29	3.75	1.54	0.96	0.96	0.92	0.99	0.92
	Spain-Asturias	18	83.33	1 747	7 793.80	2.57	2.90	1.07	0.97	0.97	0.94	0.98	0.84
	Spain-Aragon	13	71.65	1 695	9 660.93	2.00	2.59	0.74	0.97	0.97	0.85	0.91	0.96
	Spain-Andalucia	11	526.36	1 713	82 881.75	1.74	2.10	0.64	0.98	0.98	0.87	0.92	0.98
	Sweden	33	913.64	4 973	129 863.68	2.67	4.46	0.70	0.96	0.96	0.97	1.02	1.02
	Switzerland	217	1 679.68	12 966	90 493.30	0.93	3.38	1.86	0.97	0.97	1.02	1.11	0.97
	Turkey	216	33 457.71	5 058	665 607.67	0.02	0.14	5.03	1.00	0.98	0.47	0.84	1.02
	United Kingdom	712	31 732.72	15 668	744 036.34	1.62	3.27	4.26	0.97	0.97	0.94	0.99	0.99
	United Kingdom-Scotland	145	2 657.62	3 255	59 023.77	2.87	4.20	4.50	0.96	0.96	0.91	0.93	1.02
	United States	363	228 369.18	6 433	3 720 556.81	3.83	4.28	6.14	0.96	0.96	0.85	0.95	0.93

1. Code 4 within-school exclusion is defined as students with dyslexia in Greece, Ireland and Poland, as students with dyslexia/-calculi in Denmark, as students with partial skills deficiencies (dyslexia, dysgraphia, etc.) in Hungary, as Maori students in immersion or bilingual programs in New Zealand, and for Lithuania, it includes all exclusions that were not coded to a specific exclusion category.



Table 11.1 [Part 3/3]
Sampling and coverage rates

		All 15-year-olds	Enrolled 15-year-olds	Target population	School level exclusions	Target minus school level exclusions	% school level exclusions	Estimate of enrolled students on frame	Participants		Excluded	
									Actual	Weighted	Actual	Weighted
Partners	Argentina	662 686	579 222	579 222	2 393	576 829	0.41	576 124.51	4 339	523 047.82	4	635.69
	Azerbaijan	139 119	139 119	131 235	780	130 455	0.59	130 422.82	5 184	122 208.40	0	0.00
	Brazil	3 390 471	2 374 044	2 357 355	0	2 357 355	0.00	2 347 345.55	9 295	1 875 461.15	19	6 437.58
	Bulgaria	89 751	88 071	88 071	1 733	86 338	1.97	83 281.35	4 498	74 325.71	0	0.00
	Chile	299 426	255 459	255 393	2 284	253 109	0.89	249 370.28	5 235	233 526.11	28	1 259.24
	Colombia	897 477	543 630	543 630	2 814	540 816	0.52	535 165.71	4 478	537 262.21	2	185.59
	Croatia	54 500	51 318	51 318	548	50 770	1.07	48 768.42	5 213	46 522.57	38	381.58
	Estonia	19 871	19 623	19 623	569	19 054	2.90	19 267.17	4 865	18 662.26	50	208.37
	Hong Kong-China	77 398	75 542	75 542	678	74 864	0.90	76 956.04	4 645	75 144.65	1	20.89
	Indonesia	4 238 600	3 119 393	2 983 254	9 388	2 973 866	0.31	2 256 019.14	10 647	2 248 313.41	0	0.00
	Israel	122 626	109 370	109 370	1 770	107 600	1.62	105 941.21	4 584	93 346.84	72	1 338.74
	Jordan	138 026	126 708	126 708	0	126 708	0.00	99 088.50	6 509	90 266.78	73	1 041.92
	Kyrgyzstan	128 810	94 922	92 109	1 617	90 492	1.76	90 239.71	5 904	80 674.46	42	521.05
	Latvia	34 277	33 659	33 534	932	32 602	2.78	32 531.65	4 719	29 231.86	26	129.60
	Liechtenstein	422	362	362	0	362	0.00	362.00	339	353.00	3	3.00
	Lithuania	53 931	51 808	51 761	613	51 148	1.18	50 584.35	4 744	50 329.08	28	263.81
	Montenegro	9 190	8 973	8 973	155	8 818	1.72	7 780.00	4 455	7 733.55	0	0.00
	Qatar	8 053	7 865	7 865	0	7 865	0.00	7 407.00	6 265	7 271.34	3	3.13
	Romania	341 181	241 890	240 661	2 943	237 718	1.22	231 532.75	5 118	223 887.02	0	0.00
	Russian Federation	2 243 924	2 077 231	2 077 231	43 425	2 033 806	2.09	1 848 221.08	5 799	1 810 855.92	60	20 576.00
	Serbia	88 584	80 692	80 692	1 811	78 881	2.24	77 568.27	4 798	73 906.69	6	86.07
	Slovenia	23 431	23 018	23 018	228	22 790	0.99	22 565.26	6 595	20 595.17	45	98.43
	Thailand	895 924	727 860	727 860	7 234	720 626	0.99	721 962.51	6 192	644 124.69	5	352.67
	Tunisia	153 331	153 331	153 331	0	153 331	0.00	153 009.06	4 640	138 491.18	2	51.68
	Uruguay	52 119	40 815	40 815	97	40 718	0.24	39 854.48	4 839	36 011.48	5	38.90

		Ineligible		Eligible		Within school exclusions (%)[1]	Overall exclusions (%)	Ineligible (%)	Coverage Indices				
		Actual	Weighted	Actual	Weighted				1	2	3	4	5
Partners	Argentina	259	27 533.89	4 963	523 683.51	0.12	0.53	5.26	0.99	0.99	0.79	0.91	1.00
	Azerbaijan	27	766.03	5 284	122 208.40	0.00	0.59	0.63	0.99	0.94	0.88	0.94	1.00
	Brazil	1 108	216 215.00	10 554	1 881 898.74	0.34	0.34	11.49	1.00	0.99	0.55	0.80	1.00
	Bulgaria	157	2 786.41	4 768	74 325.71	0.00	1.97	3.75	0.98	0.98	0.83	0.89	0.96
	Chile	209	8 451.50	5 615	234 785.34	0.54	1.43	3.60	0.99	0.99	0.78	0.94	0.99
	Colombia	202	26 549.37	4 789	537 447.80	0.03	0.55	4.94	0.99	0.99	0.60	1.00	0.99
	Croatia	72	595.97	5 493	46 904.15	0.81	1.87	1.27	0.98	0.98	0.85	0.96	0.96
	Estonia	63	276.44	5 169	18 870.63	1.10	3.97	1.46	0.96	0.96	0.94	0.98	1.01
	Hong Kong-China	36	617.57	5 074	75 165.54	0.03	0.93	0.82	0.99	0.99	0.97	0.98	1.03
	Indonesia	324	57 333.01	10 918	2 248 313.41	0.00	0.31	2.55	1.00	0.95	0.53	1.00	0.76
	Israel	423	7 984.81	5 130	94 685.58	1.41	3.01	8.43	0.97	0.97	0.76	0.89	0.98
	Jordan	222	2 855.45	6 864	91 308.70	1.14	1.14	3.13	0.99	0.99	0.65	0.92	0.78
	Kyrgyzstan	197	2 439.28	6 116	81 195.51	0.64	2.39	3.00	0.98	0.95	0.63	0.90	1.00
	Latvia	261	1 622.62	4 911	29 361.46	0.44	3.21	5.53	0.97	0.96	0.85	0.90	1.00
	Liechtenstein	2	2.00	356	356.00	0.84	0.84	0.56	0.99	0.99	0.84	0.98	1.00
	Lithuania	63	592.92	5 089	50 592.89	0.52	1.70	1.17	0.98	0.98	0.93	1.00	0.99
	Montenegro	41	46.45	4 951	7 733.55	0.00	1.72	0.60	0.98	0.98	0.84	0.99	0.88
	Qatar	158	158.53	7 219	7 274.47	0.04	0.04	2.18	1.00	1.00	0.90	0.98	0.94
	Romania	49	3 950.23	5 129	223 887.02	0.00	1.22	1.76	0.99	0.98	0.66	0.97	0.97
	Russian Federation	57	14 435.05	6 096	1 831 431.92	1.12	3.19	0.79	0.97	0.97	0.81	0.99	0.91
	Serbia	204	2 944.59	5 118	73 992.75	0.12	2.36	3.98	0.98	0.98	0.83	0.95	0.98
	Slovenia	168	422.74	7 288	20 693.60	0.48	1.46	2.04	0.99	0.99	0.88	0.92	0.99
	Thailand	199	22 914.23	6 271	644 477.36	0.05	1.05	3.56	0.99	0.99	0.72	0.89	1.00
	Tunisia	249	6 567.81	4 907	138 542.86	0.04	0.04	4.74	1.00	1.00	0.90	0.91	1.00
	Uruguay	462	3 395.56	5 550	36 050.38	0.11	0.34	9.42	1.00	1.00	0.69	0.90	0.98

1. Code 4 within-school exclusion is defined as students with dyslexia in Greece, Ireland and Poland, as students with dyslexia/-calculi in Denmark, as students with partial skills deficiencies (dyslexia, dysgraphia, etc.) in Hungary, as Maori students in immersion or bilingual programs in New Zealand, and for Lithuania, it includes all exclusions that were not coded to a specific exclusion category.

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



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 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 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.

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 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 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, as indicated on the sampling frame.

With the use of probability proportional-to-size sampling, in 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

$$\textit{weighted school response rate before replacement} = \frac{\sum_{i \in Y} W_i E_i}{\sum_{i \in (Y \cup N)} W_i E_i}$$

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.

The weighted school response rate, after replacement, is given by the formula:

11.2

$$\textit{weighted school response rate after replacement} = \frac{\sum_{i \in (Y \cup R)} W_i E_i}{\sum_{i \in (Y \cup R \cup N)} W_i E_i}$$

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.

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



Table 11.2 [Part 1/2]
School response rates before replacement

		Weighted school Participation Rate Before Replacement (%)	Weighted Number of Responding schools (Weighted also by enrolment)	Weighted Number of schools Sampled (responding + non-responding) (Weighted also by enrolment)	Unweighted school Participation Rate Before Replacement (%)	Number of Responding schools (Unweighted)	Number of Responding and Non-responding schools (Unweighted)
OECD	Australia	98.40	247 211.55	251 221.74	98.03	349	356
	Austria	98.77	91 471.27	92 606.34	97.04	197	203
	Belgium	81.54	100 784.59	123 596.62	81.94	236	288
	Belgium-Flanders	80.01	53 646.19	67 048.31	79.66	141	177
	Canada	83.20	348 247.71	418 565.11	90.33	850	941
	Czech Republic	72.87	91 280.51	125 258.79	75.00	198	264
	Denmark	87.24	49 864.90	57 156.10	86.70	189	218
	Finland	100.00	65 085.51	65 085.51	100.00	155	155
	France	96.68	732 365.76	757 511.93	95.72	179	187
	Germany	98.15	932 815.38	950 350.10	98.24	223	227
	Greece	92.51	96 973.38	104 827.25	91.67	176	192
	Hungary	94.70	108 354.48	114 424.54	95.24	180	189
	Iceland	98.35	4 819.00	4 900.00	89.40	135	151
	Ireland	100.00	57 245.39	57 245.39	100.00	164	164
	Italy	90.53	564 533.15	623 569.70	86.16	753	874
	Italy-Basilicata	99.61	7 706.00	7 736.12	93.22	55	59
	Italy-Bolzano	97.71	4 804.93	4 917.44	88.30	83	94
	Italy-Campania	89.21	71 059.88	79 658.99	84.21	48	57
	Italy-Emilia Romagna	96.32	33 865.72	35 160.37	86.21	50	58
	Italy-Friuli Venezia Giulia	86.80	8 786.77	10 123.28	76.81	53	69
	Italy-Liguria	91.84	11 995.37	13 061.63	93.33	70	75
	Italy-Lombardia	88.85	78 600.94	88 462.73	84.21	48	57
	Italy-Piemonte	89.19	34 117.12	38 250.67	81.03	47	58
	Italy-Puglia	91.40	44 715.50	48 922.23	90.57	48	53
	Italy-Sardegna	86.72	16 721.14	19 280.96	83.33	50	60
	Italy-Sicilia	84.93	56 204.80	66 178.54	83.05	49	59
	Italy-Trento	97.25	5 243.68	5 391.76	90.91	60	66
	Italy-Veneto	93.80	45 659.09	48 677.17	87.72	50	57
	Japan	87.27	1 032 151.56	1 182 687.63	87.24	171	196
	Korea	99.24	572 255.97	576 636.64	98.71	153	155
	Luxembourg	100.00	4 955.00	4 955.00	100.00	31	31
	Mexico	95.46	1 281 866.56	1 342 897.79	94.17	1115	1184
	Netherlands	75.70	151 038.94	199 533.05	75.26	146	194
	New Zealand	91.69	54 181.69	59 089.52	90.50	162	179
	Norway	90.47	54 613.10	60 368.65	90.61	193	213
	Poland	95.41	507 650.90	532 060.81	94.14	209	222
	Portugal	94.87	94 835.05	99 961.25	94.83	165	174
	Slovak Republic	92.42	70 860.20	76 671.38	89.47	170	190
	Spain	98.26	416 538.81	423 903.57	99.42	682	686
	Spain-La Rioja	100.00	2 641.00	2 641.00	100.00	45	45
	Spain-Basque Country	100.00	15 753.72	15 753.72	100.00	151	151
	Spain-Navarra	100.00	4 952.20	4 952.20	100.00	52	52
	Spain-Galicia	100.00	23 724.51	23 724.51	100.00	53	53
	Spain-Catalonia	95.99	58 759.14	61 213.50	96.08	49	51
	Spain-Castilla y Leon	100.00	21 852.57	21 852.57	100.00	52	52
	Spain-Cantabria	100.00	4 751.33	4 751.33	100.00	53	53
	Spain-Asturias	100.00	7 983.50	7 983.50	100.00	53	53
	Spain-Aragon	100.00	10 594.50	10 594.50	100.00	51	51
	Spain-Andalucia	100.00	90 552.40	90 552.40	100.00	51	51
	Sweden	99.59	126 611.35	127 133.27	99.00	197	199
	Switzerland	95.44	77 940.45	81 660.28	96.88	496	512
	Turkey	97.16	773 776.70	796 371.42	96.88	155	160
	United Kingdom	76.05	569 438.45	748 795.67	74.79	439	587
	United Kingdom-Scotland	63.61	40 491.76	63 655.81	63.63	70	110
	United States	68.95	2 689 741.31	3 901 130.57	69.38	145	209



Table 11.2 [Part 2/2]
School response rates before replacement

		Weighted school Participation Rate Before Replacement (%)	Weighted Number of Responding schools (Weighted also by enrolment)	Weighted Number of schools Sampled (responding + non-responding) (Weighted also by enrolment)	Unweighted school Participation Rate Before Replacement (%)	Number of Responding schools (Unweighted)	Number of Responding and Non-responding schools (Unweighted)
Partners	Argentina	95.08	547 775.36	576 124.51	93.85	168	179
	Azerbaijan	94.86	123 717.99	130 422.82	94.77	163	172
	Brazil	98.01	2 300 529.53	2 347 345.55	96.34	606	629
	Bulgaria	98.76	82 248.09	83 281.35	98.89	178	180
	Chile	83.08	207 182.85	249 370.28	82.14	161	196
	Colombia	93.53	500 566.82	535 165.71	92.22	154	167
	Croatia	98.59	48 080.63	48 768.42	97.55	159	163
	Estonia	98.98	19 070.50	19 267.17	98.82	167	169
	Hong Kong-China	68.57	52 768.08	76 956.04	67.95	106	156
	Indonesia	99.72	2 249 727.84	2 256 019.14	99.15	349	352
	Israel	89.89	95 231.11	105 941.21	83.23	139	167
	Jordan	100.00	99 088.50	99 088.50	100.00	210	210
	Kyrgyzstan	99.58	89 863.21	90 239.71	99.50	200	201
	Latvia	97.57	31 740.22	32 531.65	97.71	171	175
	Liechtenstein	100.00	362.00	362.00	100.00	12	12
	Lithuania	96.85	48 988.90	50 584.35	96.45	190	197
	Montenegro	94.64	7 363.00	7 780.00	96.08	49	51
	Qatar	98.02	7 260.00	7 407.00	93.43	128	137
	Romania	100.00	231 532.75	231 532.75	100.00	174	174
	Russian Federation	100.00	1 848 221.08	1 848 221.08	100.00	209	209
	Serbia	98.67	76 533.75	77 568.27	98.16	160	163
	Slovenia	97.42	21 983.00	22 565.26	97.26	355	365
	Thailand	97.70	705 352.94	721 962.51	98.11	208	212
	Tunisia	100.00	153 009.06	153 009.06	100.00	152	152
	Uruguay	96.30	38 377.90	39 854.48	96.43	270	280

The denominator is the number of sampled students who were age-eligible, and not explicitly excluded as student exclusions. The exception is cases where 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-olds 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 is 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 11.3 [Part 1/2]
School response rates after replacement

		Weighted school Participation Rate After all Replacement (%)	Weighted Number of Responding schools (Weighted also by enrolment)	Weighted Number of schools Sampled (responding + nonresponding) (Weighted also by enrolment)	Unweighted school Participation Rate after all Replacement (%)	Number of Responding schools (Unweighted)	Number of Responding and nonresponding schools (Unweighted)
OECD	Australia	98.85	248 320.55	251 221.74	98.31	350	356
	Austria	98.77	91 471.27	92 606.34	97.04	197	203
	Belgium	93.59	115 645.52	123 562.62	93.40	269	288
	Belgium-Flanders	91.78	61 503.35	67 014.31	91.53	162	177
	Canada	86.23	360 866.86	418 514.45	91.50	861	941
	Czech Republic	93.87	117 526.33	125 202.46	92.42	244	264
	Denmark	96.47	55 067.95	57 085.31	95.87	209	218
	Finland	100.00	65 085.51	65 085.51	100.00	155	155
	France	96.68	732 365.76	757 511.93	95.72	179	187
	Germany	99.05	941 355.81	950 350.10	99.12	225	227
	Greece	99.35	104 124.05	104 809.66	98.44	189	192
	Hungary	100.00	114 266.23	114 266.23	100.00	189	189
	Iceland	98.35	4 819.00	4 900.00	89.40	135	151
	Ireland	100.00	57 245.39	57 245.39	100.00	164	164
	Italy	97.47	607 859.64	623 618.70	91.08	796	874
	Italy-Basilicata	99.61	7 706.00	7 736.12	93.22	55	59
	Italy-Bolzano	97.71	4 804.93	4 917.44	88.30	83	94
	Italy-Campania	95.84	76 343.75	79 658.99	91.23	52	57
	Italy-Emilia Romagna	98.27	34 551.11	35 160.37	87.93	51	58
	Italy-Friuli Venezia Giulia	97.53	9 873.62	10 123.28	85.51	59	69
	Italy-Liguria	97.89	12 786.41	13 061.63	97.33	73	75
	Italy-Lombardia	99.32	87 860.16	88 462.73	94.74	54	57
	Italy-Piemonte	95.35	36 471.03	38 250.67	86.21	50	58
	Italy-Puglia	99.61	48 729.82	48 922.23	98.11	52	53
	Italy-Sardegna	96.51	18 607.86	19 280.96	91.67	55	60
	Italy-Sicilia	92.94	61 506.00	66 178.54	89.83	53	59
	Italy-Trento	97.25	5 243.68	5 391.76	90.91	60	66
	Italy-Veneto	99.15	48 310.68	48 726.17	92.98	53	57
	Japan	92.38	1 092 615.65	1 182 687.63	92.35	181	196
	Korea	99.89	575 983.97	576 636.64	99.35	154	155
	Luxembourg	100.00	4 955.00	4 955.00	100.00	31	31
	Mexico	96.20	1 291 872.06	1 342 897.79	95.27	1128	1184
	Netherlands	94.25	187 952.81	199 423.37	94.33	183	194
	New Zealand	96.06	56 761.97	59 089.52	94.97	170	179
	Norway	95.40	57 582.32	60 358.60	95.31	203	213
	Poland	99.99	532 149.94	532 197.11	99.55	221	222
	Portugal	98.73	98 593.06	99 862.92	98.85	172	174
	Slovak Republic	99.93	76 864.87	76 920.17	98.95	188	190
	Spain	100.00	424 620.57	424 620.57	100.00	686	686
	Spain-La Rioja	100.00	2 641.00	2 641.00	100.00	45	45
	Spain-Basque Country	100.00	15 753.72	15 753.72	100.00	151	151
	Spain-Navarra	100.00	4 952.20	4 952.20	100.00	52	52
	Spain-Galicia	100.00	23 724.51	23 724.51	100.00	53	53
	Spain-Catalonia	100.00	61 213.50	61 213.50	100.00	51	51
	Spain-Castilla y Leon	100.00	21 852.57	21 852.57	100.00	52	52
	Spain-Cantabria	100.00	4 751.33	4 751.33	100.00	53	53
	Spain-Asturias	100.00	7 983.50	7 983.50	100.00	53	53
	Spain-Aragon	100.00	10 594.50	10 594.50	100.00	51	51
	Spain-Andalucia	100.00	90 552.40	90 552.40	100.00	51	51
	Sweden	99.59	126 611.35	127 133.27	99.00	197	199
	Switzerland	99.09	81 345.26	82 094.93	99.41	509	512
	Turkey	100.00	794 825.58	794 825.58	100.00	160	160
	United Kingdom	88.15	660 502.84	749 269.55	84.16	494	587
	United Kingdom-Scotland	86.09	54 802.25	63 655.80	85.45	94	110
	United States	79.09	3 085 547.88	3 901 520.93	79.43	166	209



Table 11.3 [Part 2/2]
School response rates after replacement

		Weighted school Participation Rate After all Replacement (%)	Weighted Number of Responding schools (Weighted also by enrolment)	Weighted Number of schools Sampled (responding + nonresponding) (Weighted also by enrolment)	Unweighted school Participation Rate after all Replacement (%)	Number of Responding schools (Unweighted)	Number of Responding and nonresponding schools (Unweighted)
Partners	Argentina	96.19	554 186.35	576 124.51	95.53	171	179
	Azerbaijan	99.37	129 951.63	130 775.00	99.42	171	172
	Brazil	99.24	2 329 154.43	2 346 987.83	98.09	617	629
	Bulgaria	99.35	82 548.02	83 091.92	99.44	179	180
	Chile	87.89	219 082.48	249 282.99	88.27	173	196
	Colombia	99.22	530 584.59	534 764.00	98.80	165	167
	Croatia	99.80	48 727.00	48 823.00	98.77	161	163
	Estonia	100.00	19 260.50	19 260.50	100.00	169	169
	Hong Kong-China	93.76	72 564.37	77 392.26	93.59	146	156
	Indonesia	100.00	2 256 019.14	2 256 019.14	100.00	352	352
	Israel	93.45	99 541.35	106 519.85	89.22	149	167
	Jordan	100.00	99 088.50	99 088.50	100.00	210	210
	Kyrgyzstan	100.00	90 239.71	90 239.71	100.00	201	201
	Latvia	100.00	32 531.65	32 531.65	100.00	175	175
	Liechtenstein	100.00	362.00	362.00	100.00	12	12
	Lithuania	100.00	50 584.35	50 584.35	100.00	197	197
	Montenegro	94.64	7 363.00	7 780.00	96.08	49	51
	Qatar	98.02	7 260.00	7 407.00	93.43	128	137
	Romania	100.00	231 532.75	231 532.75	100.00	174	174
	Russian Federation	100.00	1 848 221.08	1 848 221.08	100.00	209	209
	Serbia	99.96	77 538.75	77 568.27	99.39	162	163
	Slovenia	97.71	22 048.86	22 565.26	97.53	356	365
	Thailand	100.00	721 551.81	721 551.81	100.00	212	212
	Tunisia	100.00	153 009.06	153 009.06	100.00	152	152
	Uruguay	96.30	38 377.90	39 854.48	96.43	270	280

Table 11.4 [Part 1/2]
Student response rates after replacement

		Weighted student Participation Rate after Second Replacement (%)	Number of students Assessed (Weighted)	Number of students Sampled (assessed + absent) (Weighted)	Unweighted student Participation Rate after Second Replacement (%)	Number of students Assessed (Unweighted)	Number of students Sampled (assessed + absent) (Unweighted)
OECD	Australia	86.30	200 410	232 221	84.82	14 071	16 590
	Austria	90.81	80 765	88 942	88.87	4 925	5 542
	Belgium	92.98	107 247	115 343	93.31	8 857	9 492
	Belgium-Flanders	94.66	60 343	63 749	94.66	5 124	5 413
	Canada	81.43	258 789	317 822	84.32	22 201	26 329
	Czech Republic	90.62	110 435	121 869	90.35	5 927	6 560
	Denmark	89.51	49 249	55 018	89.57	4 510	5 035
	Finland	92.78	56 954	61 387	92.76	4 714	5 082
	France	89.78	641 681	714 695	89.77	4 684	5 218
	Germany	92.26	825 350	894 612	92.26	4 884	5 294
	Greece	95.24	91 494	96 070	95.21	4 871	5 116
	Hungary	93.12	98 716	106 010	93.10	4 490	4 823
	Iceland	83.32	3 781	4 538	83.32	3 781	4 538
	Ireland	83.75	46 160	55 114	83.84	4 585	5 469
	Italy	92.30	467 291	506 270	92.70	21 753	23 465
	Italy-Basilicata	94.06	6 017	6 397	93.95	1 506	1 603
	Italy-Bolzano	93.58	4 263	4 556	94.04	2 084	2 216
	Italy-Campania	90.87	58 786	64 692	90.59	1 406	1 552
	Italy-Emilia Romagna	93.64	27 243	29 094	93.41	1 531	1 639
	Italy-Friuli Venezia Giulia	94.25	7 862	8 341	94.27	1 578	1 674
	Italy-Liguria	91.75	10 531	11 477	91.54	1 753	1 915
	Italy-Lombardia	93.12	64 328	69 083	92.87	1 524	1 641
	Italy-Piemonte	93.88	30 577	32 572	93.54	1 478	1 580
	Italy-Puglia	93.65	42 283	45 148	93.33	1 540	1 650
	Italy-Sardegna	87.74	13 644	15 550	88.59	1 390	1 569
	Italy-Sicilia	91.46	45 177	49 395	90.63	1 335	1 473
	Italy-Trento	95.28	3 994	4 191	93.91	1 757	1 871
	Italy-Veneto	95.47	37 958	39 761	95.39	1 530	1 604

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



Table 11.4 [Part 2/2]
Student response rates after replacement

		Weighted student Participation Rate after Second Replacement (%)	Number of students Assessed (Weighted)	Number of students Sampled (assessed + absent) (Weighted)	Unweighted student Participation Rate after Second Replacement (%)	Number of students Assessed (Unweighted)	Number of students Sampled (assessed + absent) (Unweighted)
OECD	Japan	99.55	1 028 039	1 032 727	99.68	5 952	5 971
	Korea	99.04	570 786	576 314	98.99	5 176	5 229
	Luxembourg	96.49	4 567	4 733	96.49	4 567	4 733
	Mexico	96.40	1 101 670	1 142 760	96.16	30 885	32 119
	Netherlands	90.15	161 900	179 592	90.20	4 848	5 375
	New Zealand	87.03	44 638	51 291	87.14	4 823	5 535
	Norway	87.81	50 232	57 205	87.78	4 692	5 345
	Poland	91.70	473 144	515 945	91.32	5 547	6 074
	Portugal	86.74	77 053	88 828	86.86	5 092	5 862
	Slovak Republic	93.19	70 837	76 011	92.82	4 729	5 095
	Spain	88.48	337 710	381 686	91.92	19 604	21 328
	Spain-Andalucia	86.94	70 803	81 437	86.88	1 463	1 684
	Spain-Aragon	91.71	8 682	9 467	92.04	1 526	1 658
	Spain-Asturias	92.33	7 011	7 594	92.45	1 579	1 708
	Spain-Basque Country	96.26	14 157	14 707	96.23	3 929	4 083
	Spain-Cantabria	91.36	4 142	4 534	91.44	1 496	1 636
	Spain-Castilla y Leon	92.31	18 183	19 697	92.42	1 512	1 636
	Spain-Catalonia	91.77	52 299	56 987	91.77	1 527	1 664
	Spain-Galicia	94.14	21 254	22 578	94.08	1 573	1 672
	Spain-La Rioja	89.77	2 239	2 494	90.43	1 333	1 474
	Spain-Navarra	93.38	4 368	4 678	93.69	1 590	1 697
	Sweden	91.37	115 210	126 095	91.59	4 443	4 851
	Switzerland	94.94	84 366	88 861	95.41	12 191	12 778
	Turkey	97.59	649 451	665 477	97.73	4 942	5 057
	United Kingdom	87.65	565 955	645 688	85.96	13 050	15 182
	United Kingdom-Scotland	78.57	38 688	49 237	78.78	2 384	3 026
	United States	91.00	2 589 680	2 845 841	90.81	5 611	6 179
Partners	Argentina	89.31	447 966	501 589	88.52	4 297	4 854
	Azerbaijan	98.02	119 024	121 433	98.11	5 184	5 284
	Brazil	90.83	1 692 354	1 863 114	88.84	9 246	10 408
	Bulgaria	94.47	69 821	73 907	94.34	4 498	4 768
	Chile	93.72	192 205	205 089	93.70	5 233	5 585
	Colombia	93.89	500 459	533 020	93.55	4 478	4 787
	Croatia	95.63	44 400	46 431	95.56	5 213	5 455
	Estonia	94.89	17 708	18 662	95.04	4 865	5 119
	Hong Kong-China	91.51	64 124	70 071	91.56	4 645	5 073
	Indonesia	97.81	2 199 184	2 248 313	97.52	10 647	10 918
	Israel	90.57	79 246	87 498	90.63	4 584	5 058
	Jordan	96.26	86 890	90 267	95.85	6 509	6 791
	Kyrgyzstan	97.08	78 319	80 674	97.20	5 904	6 074
	Latvia	96.66	28 255	29 232	96.60	4 719	4 885
	Liechtenstein	96.03	339	353	96.03	339	353
	Lithuania	93.76	47 189	50 329	93.74	4 744	5 061
	Macao-China	97.57	6 261	6 417	97.50	4 760	4 882
	Montenegro	93.23	6 821	7 317	93.29	4 367	4 681
	Qatar	87.34	6 224	7 126	87.34	6 224	7 126
	Romania	99.83	223 503	223 887	99.79	5 118	5 129
	Russian Federation	96.02	1 738 842	1 810 856	96.07	5 799	6 036
	Serbia	93.91	69 375	73 877	93.86	4 798	5 112
	Slovenia	91.50	18 489	20 206	91.41	6 576	7 194
	Chinese Taipei	97.75	283 168	289 675	98.08	8 815	8 988
	Thailand	98.74	636 028	644 125	98.82	6 192	6 266
	Tunisia	94.53	130 922	138 491	94.60	4 640	4 905
	Uruguay	88.24	30 693	34 784	88.83	4 779	5 380



DESIGN EFFECTS AND EFFECTIVE SAMPLE SIZES

Surveys in education and especially international surveys rarely sample students by simply selecting a random sample of students (a simple random sample). 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 they can be with a simple random sample because they are usually more similar than students attending distinct educational institutions. For instance, they 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 house, 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 than for a simple random sample of the same size.

In the case of a simple random sample, the standard error on 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 for 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 a simple random sample is inversely proportional to the number of selected students. The standard error on the mean for 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.

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



Figure 11.1 shows the standard errors of a mean for any standardized variable for a simple random sample of 5000 students and for cluster samples of 25 students per school, for different intraclass correlation coefficients. In the case of a sample of 25 students per school, this would mean that 200 schools participated.

Figure 11.1
Standard error on a mean estimate depending on the intraclass correlation

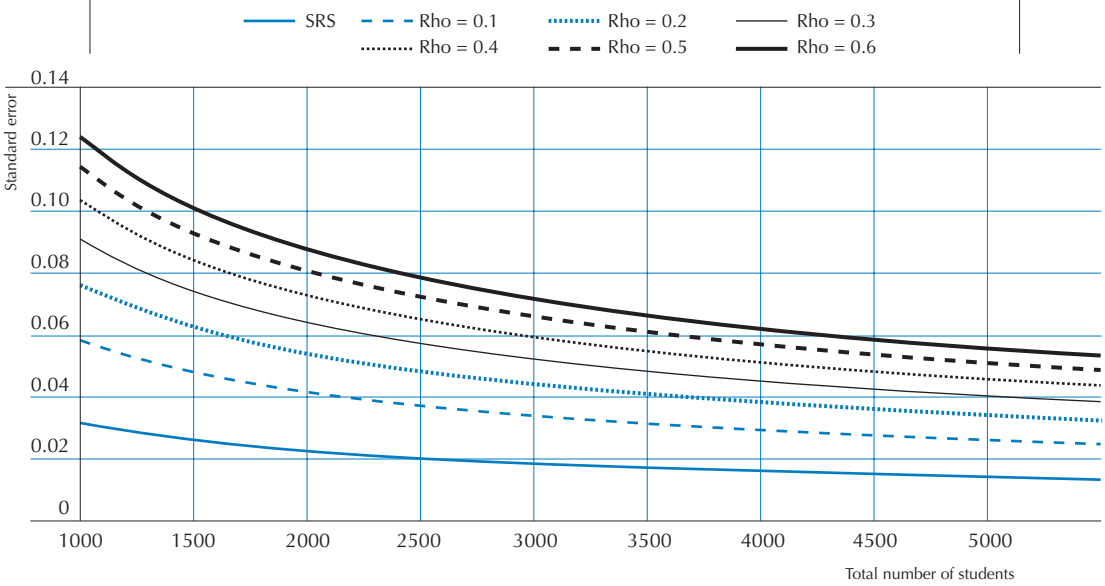


Figure 11.1 shows that the standard error on the mean is quite a lot larger for a cluster sample than it is for a simple random sample and also that the standard error is proportional of the intraclass correlation.

To limit this 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 are implemented to limit the increase in the standard error: (i) explicit and or implicit stratification of the school sample 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.

Table 11.5 provides the standard errors on the PISA 2006 combined science scale if the country sample was selected according to (i) a simple random sample; (ii) a multistage procedure without using complementary information and (iii) the BRR estimate for the actual PISA 2006 design, using the Fay's (BRR) replicates. It should be mentioned that the plausible value imputation variance was not included in these computations.

Note that the values in Table 11.5 for the standard errors for the unstratified design are overestimates for countries that had a school census (Iceland, Liechtenstein, Luxembourg, Macao - China, and Qatar) since these standard error estimates assume a simple random sample of schools.

Also note that in many of the countries where the unbiased values in Table 11.5 are greater than the values for the unstratified cluster sample, this is because of regional oversampling (Brazil, Canada, Indonesia, Mexico, Spain and Switzerland) or a three-stage design was used (Russian Federation).



Table 11.5
Standard errors for the PISA 2006 combined science scale

		Simple Random Sample	Unstratified Multi-stage Sample	BRR Estimate for PISA sample
OECD	Australia	0.84	2.47	2.26
	Austria	1.39	5.39	3.92
	Belgium	1.06	4.47	2.48
	Canada	0.63	1.64	2.03
	Czech Republic	1.28	4.95	3.48
	Denmark	1.38	2.90	3.11
	Finland	1.25	2.08	2.02
	France	1.48	5.59	3.36
	Germany	1.43	5.23	3.80
	Greece	1.32	4.95	3.23
	Hungary	1.32	5.51	2.68
	Iceland	1.57	3.26	1.64
	Ireland	1.39	3.31	3.19
	Italy	0.65	2.64	2.02
	Japan	1.30	5.23	3.37
	Korea	1.25	4.44	3.36
	Luxembourg	1.43	9.53	1.05
	Mexico	0.46	1.52	2.71
	Netherlands	1.37	5.47	2.74
	New Zealand	1.54	3.67	2.69
	Norway	1.40	2.58	3.11
	Poland	1.21	2.75	2.34
	Portugal	1.24	3.98	3.02
	Slovakia	1.35	4.56	2.59
	Spain	0.65	1.63	2.57
	Sweden	1.41	2.77	2.37
	Switzerland	0.90	2.79	3.16
	Turkey	1.18	4.88	3.84
	United Kingdom	0.93	2.50	2.29
	United States	1.42	4.19	4.22
Partners	Argentina	1.54	5.38	6.08
	Azerbaijan	0.77	3.13	2.75
	Brazil	0.93	2.69	2.79
	Bulgaria	1.59	6.03	6.11
	Chile	1.27	5.40	4.32
	Colombia	1.27	3.91	3.37
	Croatia	1.19	4.44	2.45
	Estonia	1.20	3.12	2.52
	Hong Kong-China	1.35	4.71	2.47
	Indonesia	0.68	2.37	5.73
	Israel	1.65	5.34	3.71
	Jordan	1.11	3.18	2.84
	Kyrgyzstan	1.09	3.84	2.93
	Latvia	1.23	2.99	2.97
	Liechtenstein	5.26	16.84	4.10
	Lithuania	1.31	3.64	2.76
	Macao-China	1.13	6.61	1.06
	Montenegro	1.20	6.24	1.06
	Qatar	1.06	5.86	0.86
	Romania	1.13	4.40	4.20
	Russian Federation	1.18	3.42	3.67
	Serbia	1.23	4.49	3.04
	Slovenia	1.21	4.21	1.11
	Chinese Taipei	1.01	4.27	3.57
	Thailand	0.98	3.58	2.14
	Tunisia	1.21	4.48	2.96
	Uruguay	1.36	3.77	2.75



The unbiased values in Table 11.5 are also greater than the values for the unstratified cluster sample for Argentina, Denmark and the United States. For Argentina and the United States, this may be caused by the small school strata. As described in the sampling design chapter, some countries have a substantial proportion of students attending schools with fewer than *TCS* PISA students. In such cases, to compensate the loss of assessed students, schools with fewer than *TCS* PISA students were placed in very small school strata, moderately small school strata or small school strata, depending on the percentage of students attending such schools. Schools in the very small school strata were undersampled while schools in all large school strata were slightly oversampled.

These small school strata appear, in some cases, to have an adverse impact on the standard errors. For instance, removing all small school strata in the United States reduces the standard error on the mean for the science performance estimate from 4.21 to 3.72. When a similar approach was taken for Argentina, the standard error was reduced from 6.08 to 4.61. Recall that removing schools from the sample should theoretically, all else equal, increase the standard error. This phenomenon might be due to the mixing of explicit strata in small school strata (small school strata were sorted by the explicit stratification variables).

For Denmark, there is no ready explanation as to why the unbiased estimate (3.11) is somewhat greater than that based on an unstratified design (2.90), except perhaps the fact that these estimates are based on samples and are therefore subject to random variation. However, this suggests that the stratification did not explain much between-school variance in Denmark.

It is usual to express the effect of the sampling design on the standard errors by a parameter 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 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 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 base but considering the sample as a simple random sample.

Based on Table 11.5, the standard error on the mean estimate is science in Australia is equal to 2.26. As the standard deviation of the science performance is equal to 100.205, the design effect in Australia for the mean estimate in science is therefore equal to:

11.7

$$Deff(t) = \frac{Var_{BRR}(t)}{Var_{SRS}(t)} = \frac{(2.26)^2}{[(100.205)^2/14170]} = 7.21$$

The sampling variance on the science performance mean in Australia is about seven times larger than it would have been with a simple random sample of equal size.

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



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 science performance mean is equal to:

11.9

$$Effn(t) = \frac{n}{Deff(t)} = \frac{n \times Var_{SRS}(t)}{Var_{Brr}(t)} = \frac{(100.205)^2}{(2.26)^2} = 1965.9$$

In other words, a simple random sample of 1966 students in Australia would have been as precise as the actual PISA 2006 sample for the estimation of the science performance, for the national estimate of mean science proficiency.

Variability of the design effect

Neither the design effect nor the effective sample size are a definitive characteristic of a sample. They vary both 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, and so on.

Design effects in PISA for performance variables

The notion of design effect as given earlier is here 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 error computed for the international PISA initial report, *PISA 2006: Science Competencies for Tomorrow's World* (OECD, 2007) that involves performance variables (plausible values or proficiency levels) 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 random (measurement) error.

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



For any statistic *t*, the population estimate and the sampling variance are computed for each plausible value (or each proficiency level) and then combined as described in Chapter 9.

The five design effects and their respective effective sample sizes are defined as follows:

11.10

$$Deff_1(t) = \frac{Var_{SRS}(t) + MVar(t)}{Var_{SRS}(t)}$$

where *MVar(t)* is the measurement error variance for the statistic *t*. 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 11.6 provides, per domain and per cycle, the design effect 1 values, for any country that participated in at least one cycle. Table 11.7 provides the corresponding effective sample size.

11.11

$$Deff_2(t) = \frac{Var_{BRR}(t) + MVar(t)}{Var_{SRS}(t) + MVar(t)}$$

shows the inflation of the *total* variance due only to the use of a complex sampling design. Table 11.8 provides, for each domain and PISA cycle, the design effect 2 values, for each country. Table 11.9 provides the corresponding effective sample size.

11.12

$$Deff_3(t) = \frac{Var_{BRR}(t)}{Var_{SRS}(t)}$$

shows the inflation of the sampling variance due to the use of a complex design. Table 11.9 provides, for each domain and PISA cycle, the design effect 3 values, for each country. Table 11.10 provides the corresponding effective sample size.

11.13

$$Deff_4(t) = \frac{Var_{BRR}(t) + MVar(t)}{Var_{BRR}(t)}$$

shows the inflation of the total variance due to measurement error. Table 11.11 provides, for each domain and PISA cycle, the design effect 4 values, for each country. Table 11.12 provides the corresponding effective sample size.

11.14

$$Deff_5(t) = \frac{Var_{BRR}(t) + MVar(t)}{Var_{SRS}(t)}$$

shows the inflation of the *total* variance due to the measurement error and due to the complex sampling design. Table 11.12 provides, for each domain and PISA cycle, the design effect 5 values, for each country. Table 11.13 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.



Table 11.6
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
	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
	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
	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
	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
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	1.36				1.09	1.22	1.16
	Chile	1.12	1.30	1.36				1.17	1.28	1.08
	Colombia							1.36	1.10	1.46
	Croatia							1.17	1.12	1.12
	Estonia							1.07	1.07	1.15
	Hong Kong-China	1.05	1.10	1.12	1.07	1.42	1.19	1.09	1.13	1.03
	Indonesia	1.48	1.24	1.29	1.98	1.46	1.70	1.29	1.94	1.16
	Israel	1.47	1.15	1.33				1.12	1.23	1.04
	Jordan							1.51	1.20	1.07
	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
	Montenegro							1.09	1.25	1.10
	Peru	1.10	1.20	1.89						
	Qatar							1.25	1.30	1.13
	Romania							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
	Serbia				1.11	1.29	1.36	1.14	1.33	1.05
	Slovak Republic				1.03	1.14	1.02	1.13	1.43	1.13
	Slovenia							1.16	1.23	1.07
	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
	Tunisia				1.48	1.05	1.10	1.10	1.19	1.03
	Uruguay				1.34	1.10	1.04	1.16	1.20	1.13

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



Table 11.7
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	3 983	1 923	2 374	10 328	11 335	11 055	12 176	12 841	12 654
	Austria	4 483	2 620	2 500	4 195	4 040	4 211	4 508	4 141	4 399
	Belgium	6 302	3 366	3 613	7 861	8 291	5 987	8 256	8 614	8 364
	Canada	27 294	14 682	15 047	18 723	18 559	15 320	17 465	21 011	20 048
	Czech Republic	5 019	2 964	2 841	4 681	5 221	4 006	5 377	5 195	5 604
	Denmark	3 924	1 936	2 256	3 032	3 402	3 259	3 892	3 810	3 877
	Finland	4 270	2 163	2 180	5 009	4 627	4 537	4 203	2 941	3 836
	France	4 189	2 153	2 080	3 707	3 851	3 404	4 470	3 923	4 617
	Germany	4 473	2 682	2 341	4 454	4 603	4 156	4 566	4 290	4 515
	Greece	3 930	2 108	2 553	3 054	4 192	2 366	4 497	4 459	3 485
	Hungary	4 743	2 701	2 678	4 272	3 978	3 278	3 603	3 543	4 089
	Iceland	3 045	1 505	1 804	2 940	3 164	3 179	2 341	2 421	3 387
	Ireland	3 474	1 984	2 097	3 434	3 483	3 096	3 528	3 804	3 530
	Italy	4 280	2 101	2 629	6 123	6 555	9 668	18 288	16 892	19 776
	Japan	4 753	2 655	2 489	3 595	4 308	4 296	5 086	5 774	5 680
	Korea	4 413	2 470	2 264	4 379	4 457	4 898	3 519	4 706	4 388
	Luxembourg	3 043	1 761	1 698	2 890	3 872	3 135	3 783	4 032	4 283
	Mexico	3 945	2 181	2 149	15 998	18 839	5 074	17 696	10 894	17 861
	Netherlands	2 369	1 280	1 364	3 103	3 676	3 093	3 583	4 106	4 142
	New Zealand	3 549	1 793	1 974	4 102	3 742	3 892	4 122	4 073	4 629
	Norway	3 895	1 857	2 181	3 215	3 946	3 570	4 253	4 153	4 439
	Poland	3 158	1 823	1 425	3 748	3 894	4 222	5 167	4 344	5 105
	Portugal	3 836	2 323	2 471	4 166	4 534	4 052	4 005	3 803	4 153
	Slovak Republic				7 111	6 466	7 183	4 183	3 306	4 194
	Spain	5 323	3 330	3 339	5 899	7 918	7 806	14 768	9 005	10 226
	Sweden	3 669	2 207	2 163	3 960	4 362	3 240	2 690	4 180	4 044
	Switzerland	5 798	2 841	2 626	6 883	6 596	7 033	9 335	8 456	10 732
	Turkey				3 901	3 905	3 864	3 959	3 729	4 789
	United Kingdom	8 552	4 450	4 099	6 489	7 588	7 964	10 845	11 047	9 297
	United States	3 500	1 950	1 894	3 682	4 015	4 139		4 899	5 426
Partners	Albania	4 653	2 379	2 063						
	Argentina	3 363	1 901	1 686				3 355	3 258	3 896
	Azerbaijan							3 278	4 075	4 288
	Brazil	4 112	2 175	1 660	3 244	3 639	2 381	5 804	7 668	6 672
	Bulgaria	4 128	2 538	1 879				4 114	3 688	3 873
	Chile	4 372	2 095	1 997				4 490	4 086	4 855
	Colombia							3 305	4 054	3 074
	Croatia							4 438	4 659	4 666
	Estonia							4 528	4 554	4 248
	Hong Kong-China	4 199	2 223	2 181	4 171	3 162	3 777	4 281	4 108	4 488
	Indonesia	4 980	3 304	3 153	5 436	7 375	6 340	8 244	5 500	9 191
	Israel	3 063	2 161	1 884				4 077	3 739	4 390
	Jordan							4 319	5 434	6 066
	Kyrgyzstan							5 031	5 095	5 706
	Latvia	3 240	1 826	2 059	3 851	3 920	4 026	4 136	4 481	4 368
	Liechtenstein	286	153	170	316	274	285	309	278	300
	Lithuania							4 255	3 675	4 535
	Macao-China				970	1 189	1 053	3 944	3 424	4 377
	Montenegro							4 102	3 570	4 039
	Peru	4 020	2 107	1 336						
	Qatar							5 030	4 814	5 548
	Romania							3 668	3 681	4 805
	Russian Federation	5 771	3 232	3 252	4 888	4 667	5 178	4 091	4 711	5 354
	Serbia				3 977	3 424	3 247	4 216	3 617	4 578
	Slovenia							5 693	5 373	6 146
	Chinese Taipei							5 535	7 448	8 270
	Thailand	4 726	2 406	2 698	3 073	4 177	3 934	5 193	4 898	5 721
	Tunisia				3 181	4 497	4 284	4 225	3 890	4 526
	Uruguay				4 344	5 308	5 608	4 175	4 049	4 293



Table 11.8
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
	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
	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
	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
	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
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	4.39				14.20	13.56	12.70
	Chile	6.96	3.24	2.67				10.50	11.22	10.77
	Colombia							7.34	7.48	4.87
	Croatia							4.43	3.75	3.79
	Estonia							5.37	5.31	3.86
	Hong Kong-China	5.10	2.69	2.73	7.88	6.48	7.74	3.75	3.36	3.27
	Indonesia	15.08	9.47	8.71	10.69	17.38	14.12	51.68	27.19	61.43
	Israel	18.44	10.96	9.86				6.00	6.12	4.85
	Jordan							5.21	8.47	6.05
	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
	Montenegro							0.75	0.92	0.72
	Peru	8.47	3.43	2.70						
	Qatar							0.61	0.61	0.58
	Romania							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
	Serbia				7.59	6.73	5.80	6.00	5.30	5.82
	Slovenia							0.71	0.73	0.79
	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
	Tunisia				2.74	4.30	3.68	7.21	7.21	5.83
	Uruguay				3.47	5.76	3.95	3.35	2.79	3.64



Table 11.9
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	1 085	991	889	2 549	2 184	2 675	2 406	1 703	2 201
	Austria	1 590	1 370	1 370	824	925	868	769	820	696
	Belgium	958	834	690	2 031	2 452	2 767	1 404	1 326	1 705
	Canada	4 009	4 072	3 506	3 834	3 458	4 407	2 020	2 052	2 428
	Czech Republic	1 766	1 246	1 611	1 027	887	1 400	781	964	848
	Denmark	1 875	1 556	1 405	1 367	1 374	1 520	919	1 249	1 049
	Finland	1 370	1 751	1 510	2 820	2 519	2 844	1 606	1 991	2 213
	France	1 262	1 305	1 290	1 522	1 498	1 733	690	1 093	934
	Germany	2 309	1 747	2 142	1 087	969	1 053	690	748	752
	Greece	454	466	398	985	639	1 356	698	1 058	1 138
	Hungary	581	618	633	1 549	1 301	1 791	1 031	1 261	1 192
	Iceland	4 470	1 768	1 684	4 538	4 268	4 470	4 028	3 717	3 917
	Ireland	927	1 016	847	1 228	1 352	1 498	888	1 046	1 140
	Italy	1 147	1 250	1 087	2 082	1 720	1 430	2 394	2 271	2 465
	Japan	300	276	320	947	685	764	921	765	923
	Korea	935	1 047	1 095	887	994	897	789	666	849
	Luxembourg	4 603	2 415	1 983	6 122	9 061	5 890	7 380	8 698	8 992
	Mexico	783	714	696	1 013	876	3 650	1 712	2 415	1 533
	Netherlands	739	636	601	1 137	949	1 267	1 484	1 393	1 431
	New Zealand	1 560	1 128	1 811	1 991	2 287	2 260	1 447	1 805	1 654
	Norway	1 457	1 357	1 279	1 723	1 545	1 486	1 205	1 359	1 008
	Poland	581	380	513	1 302	1 462	1 328	1 381	1 603	1 600
	Portugal	553	550	513	683	673	829	982	1 173	1 056
	Slovak Republic				908	883	776	1 338	1 605	1 465
	Sweden	2 106	1 609	1 558	1 821	1 454	2 191	1 350	1 475	1 730
	Switzerland	607	618	656	1 023	1 080	1 020	1 234	1 376	1 121
	Turkey				337	301	334	609	480	485
	United Kingdom	1 682	1 570	1 687	2 138	1 817	1 984	2 476	2 050	3 079
	United States	243	181	215	1 462	1 418	1 437		571	652
Partners	Albania	977	1 410	1 427						
	Argentina	144	194	214				388	350	309
	Azerbaijan							800	574	494
	Brazil	920	864	1 253	810	521	956	1 200	1 193	1 431
	Bulgaria	488	387	581				317	332	354
	Chile	702	844	1 020				498	467	486
	Colombia							610	598	920
	Croatia							1 177	1 389	1 374
	Estonia							907	917	1 259
	Hong Kong-China	863	907	893	568	691	578	1 237	1 384	1 422
	Indonesia	489	432	468	1 007	619	762	206	392	173
	Israel	244	227	255				764	749	944
	Jordan							1 249	769	1 076
	Kyrgyzstan							1 012	754	846
	Latvia	451	632	317	730	671	654	675	787	870
	Liechtenstein	600	216	185	664	700	666	649	593	630
	Lithuania							1 144	1 217	1 115
	Macao-China				1 239	956	1 002	5 857	5 820	5 947
	Montenegro							5 938	4 837	6 226
	Peru	523	738	937						
	Qatar							10 254	10 257	10 791
	Romania							535	553	398
	Russian Federation	568	418	501	687	618	670	659	660	647
	Serbia				580	654	759	800	906	824
	Slovenia							9 244	9 015	8 373
	Chinese Taipei							995	748	747
	Thailand	633	648	694	1 320	937	1 205	1 189	1 537	1 403
	Tunisia				1 725	1 097	1 282	643	643	795
	Uruguay				1 683	1 012	1 478	1 444	1 734	1 329



Table 11.10
Design effect 3 by country, by domain and by 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
	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
	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
	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
	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
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	6.96	5.61				15.44	16.32	14.58
	Chile	7.66	3.92	3.28				12.08	14.09	11.53
	Colombia							9.60	8.16	6.63
	Croatia							5.03	4.08	4.12
	Estonia							5.69	5.60	4.28
	Hong Kong-China	5.31	2.85	2.93	8.39	8.76	8.99	3.99	3.66	3.35
	Indonesia	21.83	11.49	10.96	20.17	24.89	23.28	66.45	51.69	71.00
	Israel	26.61	12.44	12.82				6.63	7.28	5.02
	Jordan							7.35	9.94	6.42
	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
	Montenegro							0.73	0.90	0.69
	Peru	9.24	3.91	4.22						
	Qatar							0.52	0.49	0.53
	Romania							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
	Serbia				8.30	8.38	7.52	6.69	6.70	6.06
	Slovenia							0.67	0.67	0.77
	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
	Tunisia				3.58	4.47	3.96	7.82	8.41	5.96
	Uruguay				4.31	6.24	4.07	3.73	3.14	3.98



Table 11.11
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	751	779	2 176	2 007	2 417	2 118	1 560	1 999
	Austria	1 531	1 365	1 327	764	833	808	713	708	631
	Belgium	912	761	674	1 861	2 349	2 093	1 323	1 295	1 627
	Canada	3 726	3 726	3 260	2 690	2 396	2 601	1 591	1 916	2 176
	Czech Republic	1 688	1 221	1 554	794	751	966	717	862	808
	Denmark	1 796	1 440	1 383	1 081	1 182	1 279	812	1 099	929
	Finland	1 246	1 610	1 363	2 609	2 204	2 492	1 486	1 477	1 973
	France	1 164	1 184	1 148	1 380	1 393	1 498	659	946	918
	Germany	2 152	1 711	2 031	1 050	959	963	651	669	702
	Greece	388	390	393	701	586	810	652	986	872
	Hungary	566	601	612	1 437	1 138	1 395	866	1 058	1 112
	Iceland	4 633	1 741	1 679	4 774	4 338	4 552	4 191	3 677	3 933
	Ireland	856	979	838	1 128	1 258	1 296	715	903	931
	Italy	1 018	1 066	1 054	1 197	1 035	1 213	2 046	1 804	2 263
	Japan	273	253	277	759	635	707	805	745	887
	Korea	846	974	968	737	842	821	564	613	738
	Luxembourg	4 838	2 480	1 988	7 655	9 220	6 739	8 461	9 884	9 610
	Mexico	688	633	613	549	556	687	1 002	895	903
	Netherlands	711	610	593	944	891	1 057	1 187	1 229	1 273
	New Zealand	1 531	1 060	1 805	1 886	2 077	2 094	1 293	1 619	1 609
	Norway	1 398	1 234	1 246	1 495	1 517	1 366	1 119	1 244	965
	Poland	513	356	387	1 164	1 349	1 293	1 309	1 339	1 507
	Portugal	472	511	499	626	664	745	803	928	893
	Slovak Republic				882	789	761	1 223	1 249	1 346
	Spain	1 005	848	1 057	1 502	1 413	1 550	1 625	1 589	1 323
	Sweden	1 903	1 546	1 488	1 653	1 396	1 788	929	1 415	1 631
	Switzerland	580	533	531	855	870	869	968	989	997
	Turkey				275	245	269	500	371	471
	United Kingdom	1 564	1 406	1 433	1 567	1 504	1 716	2 112	1 766	2 337
	United States	222	167	193	1 081	1 120	1 164		505	633
Partners	Albania	925	1 301	1 224						
	Argentina	122	167	167				306	268	279
	Azerbaijan							537	462	416
	Brazil	797	739	935	621	435	569	788	1 007	1 074
	Bulgaria	438	376	455				291	276	308
	Chile	638	697	831				433	372	454
	Colombia							467	549	675
	Croatia							1 037	1 278	1 265
	Estonia							855	869	1 137
	Hong Kong-China	830	855	831	534	511	498	1 164	1 268	1 389
	Indonesia	337	356	372	533	432	462	160	206	150
	Israel	169	200	196				692	630	912
	Jordan							886	655	1 014
	Kyrgyzstan							885	662	821
	Latvia	383	562	305	624	581	580	602	754	817
	Liechtenstein	658	224	185	699	911	798	713	710	709
	Lithuania							1 052	1 001	1 077
	Macao-China				1 236	945	967	6 151	6 374	6 079
	Montenegro							6 114	4 943	6 492
	Peru	480	647	600						
	Qatar							12 151	12 697	11 900
	Romania							395	410	375
	Russian Federation	495	369	446	574	494	589	481	547	602
	Serbia				530	526	586	718	716	792
	Slovenia							9 872	9 837	8 541
	Chinese Taipei							653	640	704
	Thailand	568	549	645	865	775	961	1 029	1 282	1 319
	Tunisia				1 320	1 057	1 193	593	552	779
	Uruguay				1 353	935	1 435	1 299	1 541	1 217



Table 11.12
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
	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
	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
	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
	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
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
	Chile	1.02	1.08	1.11				1.01	1.02	1.01
	Colombia							1.04	1.01	1.07
	Croatia							1.03	1.03	1.03
	Estonia							1.01	1.01	1.03
	Hong Kong-China	1.01	1.03	1.04	1.01	1.05	1.02	1.02	1.04	1.01
	Indonesia	1.02	1.02	1.03	1.05	1.02	1.03	1.00	1.02	1.00
	Israel	1.02	1.01	1.03				1.02	1.03	1.01
	Jordan							1.07	1.02	1.01
	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
	Montenegro							1.12	1.28	1.15
	Peru	1.01	1.05	1.21						
	Qatar							1.48	1.62	1.25
	Romania							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
	Serbia				1.01	1.03	1.05	1.02	1.05	1.01
	Slovenia							1.24	1.34	1.10
	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
	Tunisia				1.14	1.01	1.03	1.01	1.02	1.00
	Uruguay				1.08	1.02	1.01	1.04	1.06	1.03



Table 11.13
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	4 926	2 534	2 709	12 098	12 339	12 231	13 831	14 010	13 934
	Austria	4 657	2 630	2 582	4 525	4 485	4 524	4 862	4 796	4 852
	Belgium	6 617	3 692	3 702	8 579	8 655	7 911	8 762	8 821	8 762
	Canada	29 364	16 041	16 181	26 687	26 790	25 958	22 183	22 498	22 367
	Czech Republic	5 251	3 025	2 946	6 053	6 166	5 806	5 859	5 812	5 885
	Denmark	4 097	2 090	2 292	3 833	3 952	3 872	4 402	4 333	4 380
	Finland	4 697	2 352	2 414	5 412	5 287	5 177	4 540	3 964	4 301
	France	4 542	2 373	2 337	4 090	4 143	3 938	4 680	4 532	4 696
	Germany	4 800	2 738	2 466	4 612	4 648	4 546	4 845	4 799	4 833
	Greece	4 600	2 516	2 587	4 292	4 567	3 962	4 819	4 783	4 549
	Hungary	4 870	2 777	2 772	4 604	4 550	4 205	4 286	4 224	4 383
	Iceland	2 936	1 527	1 809	2 793	3 113	3 121	2 246	2 444	3 372
	Ireland	3 762	2 059	2 119	3 739	3 741	3 577	4 380	4 406	4 323
	Italy	4 822	2 464	2 712	10 650	10 887	11 397	21 390	21 264	21 547
	Japan	5 227	2 899	2 867	4 483	4 649	4 640	5 818	5 929	5 910
	Korea	4 875	2 656	2 561	5 270	5 264	5 354	4 923	5 116	5 047
	Luxembourg	2 893	1 713	1 691	2 301	3 804	2 730	3 291	3 542	3 999
	Mexico	4 489	2 460	2 439	29 508	29 656	26 950	30 236	29 401	30 322
	Netherlands	2 463	1 334	1 382	3 738	3 917	3 706	4 478	4 652	4 657
	New Zealand	3 617	1 908	1 980	4 330	4 120	4 200	4 613	4 542	4 756
	Norway	4 058	2 042	2 237	3 703	4 019	3 883	4 579	4 535	4 638
	Poland	3 575	1 947	1 888	4 194	4 220	4 334	5 452	5 199	5 419
	Portugal	4 495	2 497	2 536	4 542	4 597	4 508	4 897	4 809	4 911
	Slovak Republic				7 317	7 240	7 329	4 576	4 247	4 565
	Spain	6 050	3 403	3 420	9 673	10 301	10 228	19 085	17 896	18 461
	Sweden	4 059	2 295	2 265	4 362	4 541	3 966	3 906	4 355	4 287
	Switzerland	6 070	3 295	3 248	8 230	8 186	8 251	11 903	11 770	12 058
	Turkey				4 789	4 796	4 787	4 821	4 824	4 927
	United Kingdom	9 198	4 968	4 826	8 852	9 164	9 208	12 717	12 823	12 248
	United States	3 824	2 119	2 105	4 980	5 081	5 109		5 539	5 590
Partners	Albania	4 916	2 577	2 403						
	Argentina	3 961	2 201	2 160				4 251	4 252	4 307
	Azerbaijan							4 890	5 061	5 099
	Brazil	4 746	2 544	2 220	4 232	4 357	4 005	8 844	9 086	8 891
	Bulgaria	4 601	2 608	2 399				4 471	4 438	4 449
	Chile	4 815	2 536	2 451				5 162	5 131	5 198
	Colombia							4 318	4 421	4 189
	Croatia							5 038	5 065	5 069
	Estonia							4 802	4 806	4 705
	Hong Kong-China	4 365	2 358	2 343	4 439	4 275	4 387	4 548	4 485	4 597
	Indonesia	7 210	4 006	3 970	10 262	10 566	10 447	10 600	10 457	10 623
	Israel	4 420	2 454	2 450				4 499	4 446	4 544
	Jordan							6 088	6 382	6 436
	Kyrgyzstan							5 754	5 801	5 876
	Latvia	3 817	2 054	2 142	4 504	4 524	4 542	4 635	4 679	4 654
	Liechtenstein	261	147	169	300	210	238	281	231	266
	Lithuania							4 626	4 469	4 695
	Macao-China				969	1 203	1 089	3 741	3 104	4 276
	Montenegro							3 983	3 478	3 872
	Peru	4 381	2 406	2 088						
	Qatar							4 236	3 875	5 025
	Romania							4 966	4 962	5 093
	Russian Federation	6 622	3 664	3 656	5 849	5 839	5 885	5 604	5 675	5 749
	Serbia				4 349	4 259	4 205	4 701	4 575	4 760
	Slovenia							5 322	4 915	6 022
	Chinese Taipei							8 444	8 699	8 769
	Thailand	5 267	2 838	2 903	4 690	5 047	4 936	6 000	5 870	6 085
	Tunisia				4 154	4 669	4 602	4 582	4 536	4 620
	Uruguay				5 403	5 743	5 777	4 640	4 556	4 689

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



Table 11.14
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
	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
	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
	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
	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
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	6.99	5.97				15.53	16.54	14.74
	Chile	7.78	4.23	3.64				12.24	14.37	11.61
	Colombia							9.95	8.27	7.09
	Croatia							5.20	4.20	4.24
	Estonia							5.77	5.67	4.43
	Hong Kong-China	5.35	2.95	3.05	8.46	9.18	9.18	4.07	3.80	3.38
	Indonesia	22.31	11.72	11.25	21.15	25.35	23.97	66.74	52.62	71.16
	Israel	27.07	12.59	13.15				6.75	7.51	5.07
	Jordan							7.86	10.14	6.49
	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
	Montenegro							0.81	1.15	0.79
	Peru	9.34	4.11	5.12						
	Qatar							0.76	0.79	0.66
	Romania							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
	Serbia				8.41	8.66	7.87	6.83	7.02	6.10
	Slovenia							0.83	0.90	0.85
	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
	Tunisia				4.06	4.52	4.06	7.92	8.60	5.98
	Uruguay				4.66	6.34	4.11	3.88	3.33	4.10



Table 11.15
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	835	666	738	2 098	1 973	2 356	2 067	1 543	1 966
	Austria	1 502	1 360	1 284	752	813	795	704	689	622
	Belgium	905	742	670	1 815	2 311	1 883	1 308	1 290	1 610
	Canada	3 686	3 626	3 199	2 568	2 296	2 416	1 558	1 904	2 150
	Czech Republic	1 652	1 204	1 495	761	732	888	708	844	801
	Denmark	1 737	1 264	1 351	982	1 108	1 174	789	1 050	898
	Finland	1 203	1 402	1 214	2 437	2 011	2 226	1 431	1 242	1 801
	France	1 131	1 082	1 036	1 312	1 342	1 372	654	909	914
	Germany	2 036	1 656	1 757	1 039	957	939	644	656	694
	Greece	382	377	392	650	579	694	645	968	814
	Hungary	564	597	606	1 388	1 086	1 232	827	995	1 086
	Iceland	4 037	1 414	1 634	3 983	4 031	4 241	2 488	2 375	3 501
	Ireland	836	948	833	1 087	1 213	1 195	684	868	878
	Italy	985	950	1 033	1 095	969	1 188	2 010	1 762	2 239
	Japan	271	250	273	723	627	697	787	742	880
	Korea	828	934	899	713	814	807	536	606	719
	Luxembourg	3 970	2 170	1 727	4 509	8 942	4 706	6 113	7 681	8 432
	Mexico	671	606	587	541	550	618	978	850	884
	Netherlands	699	589	587	884	874	982	1 092	1 174	1 217
	New Zealand	1 510	988	1 762	1 811	1 897	1 950	1 237	1 524	1 587
	Norway	1 369	1 093	1 208	1 363	1 500	1 305	1 092	1 202	954
	Poland	502	350	357	1 114	1 299	1 279	1 286	1 255	1 472
	Portugal	462	502	496	618	663	729	770	873	858
	Slovak Republic				879	778	759	1 183	1 121	1 298
	Spain	979	841	1 046	1 346	1 349	1 469	1 582	1 451	1 246
	Sweden	1 749	1 441	1 379	1 559	1 371	1 535	817	1 387	1 574
	Switzerland	577	517	507	836	846	852	945	954	986
	Turkey				271	242	266	488	362	470
	United Kingdom	1 540	1 345	1 336	1 455	1 446	1 657	2 042	1 722	2 176
	United States	221	166	191	987	1 043	1 090		498	630
Partners	Albania	913	1 206	1 063						
	Argentina	121	165	163				300	262	277
	Azerbaijan							506	451	409
	Brazil	773	692	768	591	426	512	749	984	1 027
	Bulgaria	433	375	427				290	272	305
	Chile	628	647	748				428	364	451
	Colombia							450	542	632
	Croatia							1 002	1 242	1 230
	Estonia							844	858	1 099
	Hong Kong-China	823	827	799	529	488	488	1 140	1 224	1 374
	Indonesia	330	349	362	509	424	449	160	202	150
	Israel	166	197	191				679	611	905
	Jordan							829	642	1 003
	Kyrgyzstan							862	651	817
	Latvia	376	537	303	607	568	569	592	748	806
	Liechtenstein	547	189	178	632	579	573	591	486	557
	Lithuania							1 026	943	1 066
	Macao-China				962	910	845	4 853	4 186	5 469
	Montenegro							5 467	3 877	5 645
	Peru	474	615	495						
	Qatar							8 232	7 881	9 556
	Romania							383	398	373
	Russian Federation	490	363	438	562	483	580	465	536	597
	Serbia				524	509	559	703	683	786
	Slovenia							7 979	7 344	7 803
	Chinese Taipei							625	632	701
	Thailand	560	527	632	775	747	906	997	1 216	1 297
	Tunisia				1 163	1 045	1 163	586	539	776

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



SUMMARY ANALYSES OF THE DESIGN EFFECT

To better understand the evolution of the design effect for a particular country across the three PISA cycles, some information related to the design effects and their respective effective sample sizes, are presented in appendix 3. In particular, as the design effect and the effective sample size depends on:

- ***The sample size***, the number of participating schools, the number of participating students and the average school sample size, which are provided in Table A3.2;
- ***The school variance***, school variance estimates and the intraclass correlation, which are provided respectively in Table A3.3 and Table A3.4;
- ***The stratification variables***, the intraclass correlation coefficient within explicit strata (provided in Table A3.5), and the percentage of school variance explained by explicit stratification variables (provided in Table A3.6).

Finally, the standard errors on the mean performance estimates are provided in Table A3.1.

Table 11.16 to Table 11.21 present the median of the indices presented in Table 11.6 and in Table A3.1 to Table A3.6 by cycle and per domain.

Table 11.16
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

In PISA 2000, student performance estimates for a particular domain were only provided for students who responded to testing material from that domain, while in 2003 and 2006, student proficiency estimates were provided for all domains. For PISA 2000 about five-ninths of the students were assessed in the minor domains (Adams & Wu, 2002). This difference explains why the design effects in mathematics and science for 2000 are so low in comparison with all other design effects.

The design effect associated with scientific literacy is always the smallest for any data collection. As shown by Table 11.16, this outcome seems to result from the smaller school variance estimates in scientific literacy in comparison with reading literacy and mathematical literacy. Indeed, for the three cycles, the school variance in science literacy is always the smallest. However, as will be explained below, the school variance estimates in PISA 2000 and PISA 2003 are suspected to be biased downwards.

Table 11.17 presents summary information about the standard errors of national mean achievement across PISA cycles.

Table 11.17
Median of the standard errors of the student performance mean estimate for each domain and PISA cycle for the 35 countries that participated in every cycle

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



With the exception of reading literacy in 2006, the standard errors, on average, have decreased between the 2000 data collection and the 2006 data collection. This decrease is associated with the continuously increasing school sample size. Note that, generally speaking, the sample size increase in a given country, in 2006 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 is 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 country mean achievement estimates.

This reduction of the standard errors might also be explained by a better efficiency of the explicit stratification variables. Indeed, the median percentage of school variance explained by explicit stratification variables have slightly increased, mainly between 2003 and 2006 data collection, as shown by Table 11.18.

Table 11.18 shows that school sample sizes have generally been increasing across PISA cycles.

Table 11.18
Median of the number of participating schools for each domain and PISA cycle for the 35 countries that participated in every cycle

	Number of schools
PISA 2000	176
PISA 2003	193
PISA 2006	199

Table 11.19 shows information about the size of the between-school variance across PISA cycles.

Table 11.19
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	2744	2520

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 school average of student performance weighted maximum likelihood estimates in the major domain. In 2006, the conditioning variables consist of *n*-1 dummy variables, with *n* being the number of participating schools (see Chapter 9). The method used in the first two PISA studies seems 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 2000 and in 2003 was associated with the major domain, respectively reading literacy and mathematic literacy.

Table 11.20
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 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009



Table 11.21
Median of the within explicit strata intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle

	Reading	Mathematics	Science
PISA 2000	0.21	0.19	0.18
PISA 2003	0.18	0.20	0.14
PISA 2006	0.27	0.22	0.20

Table 11.22
Median of the percentages of school variances explained by explicit stratification variables, for each domain and PISA cycle for the 35 countries that participated in every cycle

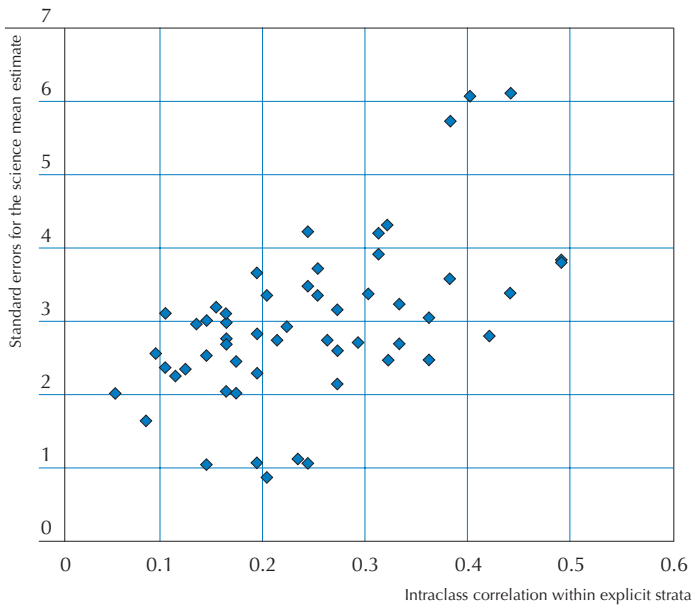
	Reading	Mathematics	Science
PISA 2000	23	22	24
PISA 2003	23	22	21
PISA 2006	34	28	34

Countries with outlying standard errors

Figure 11.2 presents the relationship between the intraclass correlation within explicit strata and the standard errors for the science performance mean estimates. The correlation between these 2 variables is equal to 0.54.

The three outlying dots in the scatter plot represent Indonesia, Argentina and Bulgaria. The large standard error for Indonesia is due to an error in the school frame for the measure of size of a single school. Removing that school from the PISA database reduces the standard errors from 5.73 to 3.33. In Bulgaria and in Argentina, the school variance within explicit stratification variable is quite large (the intraclass correlation is above 0.40) and the percentage of school variance explained by explicit stratification variable is quite low (about 0.30). This suggests that, in future cycles, efforts might be needed to improve the effectiveness of the explicit stratification in these two countries.

Figure 11.2
Relationship between the standard error for the science performance mean and the intraclass correlation within explicit strata (PISA 2006)



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



Five countries have an intraclass correlation within explicit strata higher than, or equal to, 0.30 but present a percentage of school variance explained by explicit stratification variables above 0.50 – Austria, Belgium, Chile, Hungary and Romania.

Greece has an intraclass correlation within explicit strata equal to 0.33 and a percentage of explained school variance equal to 0.43. This suggests that stratification variables used are quite efficient for explaining the school variance but can still be improved.

The following countries have an intraclass correlation within explicit strata above or equal to 0.30 and a percentage of explained variance close or below 0.30: Columbia, Hong Kong-China, Serbia, Chinese Taipei, Indonesia, Argentina, Azerbaijan, Brazil, Japan, Bulgaria, Germany and Turkey. In these countries, the sampling design should be revised and more efficient stratification variables should be identified.



12

Scaling Outcomes

International characteristics of the item pool 208
- Test targeting 208
- Test reliability 208
- Domain inter-correlations 208
- Science scales 215

Scaling outcomes 216
- National item deletions 216
- International scaling 219
- Generating student scale scores 219

Test length analysis 219
Booklet effects 221
- Overview of the PISA cognitive reporting scales 232
- PISA overall literacy scales 234
- PISA literacy scales 234
- Special purpose scales 234

Observations concerning the construction of the PISA overall literacy scales 235
- Framework development 235
- Testing time and item characteristics 236
- Characteristics of each of the links 237

Transforming the plausible values to PISA scales 246
- Reading 246
- Mathematics 246
- Science 246
- Attitudinal scales 247

Link error 247

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



INTERNATIONAL CHARACTERISTICS OF THE ITEM POOL

When main study 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 participating country (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) consisted of 398 750 students from 57 participating countries. Table 12.1 shows the total number of sampled students, broken down by participating country and test booklet.

Test targeting

Each of the domains was separately scaled to examine the targeting of the tests. Figure 12.1 shows the match between the international (OECD countries only) item difficulty distribution and the distribution of OECD's student achievement for each of reading, mathematics, science, interest and support, respectively. The figures consist of two panels. The left panel, students, shows the distribution of students' Rasch-scaled achievement estimates. Students at the top end of this distribution have higher proficiency estimates than students at the lower end of the distribution. The right panel, item difficulties, shows the distribution of Rasch-estimated item difficulties.

In each of Figure 12.2 to Figure 12.5 the student proficiency distribution, shown by *X*s, is well matched to the item difficulty distribution. For the interest scale (Figure 12.4) the items are well-matched in terms of average difficulty, but the items are not widely dispersed on the scale. For the support items, shown in Figure 12.5 shows it is clear that the items were very easy to agree with for students in OECD countries. The figures are constructed so that when a student and an item are located at the same height on the scale then the student has a 50% chance of responding correctly to the item.

Test reliability

A second test characteristic that is of importance is the test reliability. Table 12.2 shows the reliability for each of the five overall scales (mathematical literacy, reading literacy, combined scientific literacy and the attitude scales interest and support) before conditioning and based upon five separate scalings, using plausible values and using WLEs. The reliabilities for the minor domains are higher when using WLEs, because students that were not assessed in mathematics or reading were excluded form the calculation of the WLE reliabilities. These students do get plausible values, but there is no information available about these students (no scores on other domains, because of using uni-dimensional models and no background information because these are the reliabilities before conditioning). The international reliability for each domain after conditioning is reported later in Table 12.6.

Domain inter-correlations

Correlations between the ability estimates for individual students in each of the five domains, the latent correlations, as estimated by ConQuest® (Wu, Adams and Wilson, 1997) are given in Table 12.3. 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 12.3 are reported for both OECD countries and for all participating countries.[1]



Table 12.1
Number of sampled student by country and booklet

		Booklet														
		1	2	3	4	5	6	7	8	9	10	11	12	13	UH	Total
OECD	Australia	1088	1058	1081	1118	1124	1092	1074	1071	1066	1086	1106	1092	1114		14170
	Austria	370	386	383	394	395	378	372	366	370	371	387	369	367	19	4927
	Belgium	672	668	680	647	665	660	677	662	674	676	671	672	661	172	8857
	Canada	1725	1738	1692	1736	1744	1768	1762	1769	1738	1768	1745	1717	1744		22646
	Czech Republic	449	463	438	457	442	437	439	452	451	446	439	447	450	122	5932
	Denmark	355	362	348	349	367	366	361	339	335	334	340	331	345		4532
	Finland	361	372	366	366	358	349	361	359	362	367	367	366	360		4714
	France	361	360	358	356	358	361	368	363	362	367	371	364	367		4716
	Germany	366	361	360	363	354	361	364	366	370	368	371	360	367	160	4891
	Greece	370	381	390	381	384	378	367	359	370	367	383	366	377		4873
	Hungary	355	356	359	345	342	345	332	342	327	341	351	347	348		4490
	Iceland	297	294	290	303	299	298	289	284	265	288	294	289	299		3789
	Ireland	341	360	362	347	348	347	350	363	351	351	344	357	364		4585
	Italy	1674	1666	1671	1686	1656	1661	1684	1712	1716	1678	1658	1665	1646		21773
	Japan	469	471	467	451	450	453	457	454	458	446	451	461	464		5952
	Korea	396	407	400	389	389	385	394	405	406	404	408	394	399		5176
	Luxembourg	362	367	358	353	349	351	352	355	346	340	345	347	342		4567
	Mexico	2328	2391	2406	2415	2369	2356	2423	2356	2373	2402	2373	2379	2400		30971
	Netherlands	368	358	366	363	358	373	368	375	366	372	372	360	370	102	4871
	New Zealand	394	368	371	372	379	368	361	363	360	371	380	365	371		4823
	Norway	368	356	366	367	369	369	366	351	352	353	362	355	358		4692
	Poland	425	421	422	433	429	434	441	445	440	413	409	414	421		5547
	Portugal	378	388	387	381	391	397	382	395	405	386	413	397	409		5109
	Slovak Republic	373	368	362	365	369	365	364	353	350	355	358	359	352	38	4731
	Spain	1538	1534	1514	1489	1479	1507	1522	1513	1485	1501	1499	1505	1518		19604
	Sweden	337	336	346	346	344	336	336	336	336	350	360	342	338		4443
	Switzerland	932	953	933	931	940	932	951	939	948	925	927	950	931		12192
	Turkey	380	381	379	377	377	379	376	377	376	382	391	381	386		4942
	United Kingdom	993	996	997	997	1008	1034	1034	1020	1003	1027	1011	1021	1011		13152
	United States	437	438	425	427	419	419	434	423	430	420	443	451	445		5611
Partners	Argentina	347	340	332	336	340	333	325	328	330	322	331	333	342		4339
	Azerbaijan	404	407	399	407	405	407	387	387	391	398	399	397	396		5184
	Brazil	702	723	701	737	713	718	719	704	714	693	739	711	721		9295
	Bulgaria	342	345	349	342	351	347	344	349	350	347	346	345	341		4498
	Chile	413	391	395	395	414	412	410	392	394	401	401	408	407		5233
	Colombia	336	350	338	350	350	347	346	341	349	355	337	336	343		4478
	Croatia	414	421	412	410	407	404	398	399	389	392	381	378	408		5213
	Estonia	370	365	367	360	373	369	375	378	383	383	379	380	383		4865
	Hong Kong-China	348	349	353	349	357	362	371	360	363	368	368	352	345		4645
	Indonesia	822	812	806	812	816	804	815	820	825	837	825	833	820		10647
	Israel	362	357	346	363	364	350	345	351	341	335	346	352	372		4584
	Jordan	507	505	489	492	481	485	498	499	508	513	505	510	517		6509
	Kyrgyzstan	455	459	460	460	445	456	447	462	461	453	457	444	445		5904
	Latvia	368	357	359	353	359	364	362	356	369	354	368	380	370		4719
	Liechtenstein	27	26	25	26	26	27	26	25	26	25	26	26	28		339
	Lithuania	349	355	354	370	373	368	366	373	372	375	367	368	354		4744
	Macao-China	369	369	365	359	372	368	370	364	362	365	368	371	358		4760
	Montenegro	342	346	351	335	354	344	351	347	343	349	336	327	330		4455
	Qatar	476	471	488	483	478	493	477	475	482	480	496	493	473		6265
	Romania	392	394	401	406	399	396	385	388	393	393	394	388	389		5118
	Russian Federation	456	455	444	448	450	440	453	449	439	439	446	441	439		5799
	Serbia	369	370	372	363	371	375	371	364	379	377	358	365	364		4798
	Slovenia	487	508	488	505	497	485	491	492	480	488	483	484	490	217	6595
	Chinese Taipei	680	687	679	673	678	680	674	673	676	686	685	676	668		8815
	Thailand	474	475	468	466	469	483	488	478	473	474	482	486	476		6192
	Tunisia	360	343	361	368	364	361	363	364	342	349	349	356	360		4640
	Uruguay	387	390	389	367	374	368	368	348	369	368	366	364	381		4839



Figure 12.1
Item plot for mathematics items

```
   Students        Item difficulties
-----------------------------------------
      4             |
                    |
                    |
                    |
                    |
                    |
                    |25
      3             |
                    |
                  X |
                 XX |
                 XX |
                  X |
      2          XX |14 20 27
                XXX |
               XXXX |
                XXX |5 44
               XXXX |28
             XXXXXX |13 41 47 48
      1      XXXXXX |11
             XXXXXX |21 38 39
           XXXXXXXX |45
          XXXXXXXXX |23
           XXXXXXXX |2 7
          XXXXXXXXX |15 17 33 34 37
      0    XXXXXXXX |30 35
             XXXXXX |16 18
           XXXXXXXX |6 8 46
            XXXXXXX |4
           XXXXXXXX |12 36
         XXXXXXXXXX |3 19 31 32 42
     -1    XXXXXXXX |24
            XXXXXXX |26 43
               XXXX |29
              XXXX  |
             XXXXX  |1
               XXX  |22
     -2         XX  |10
                XX  |
                XX  |
                 X  |
                 X  |
                 X  |40
     -3          X  |
                    |
                    |
                    |
                    |
                    |
                    |9
     -4             |
=========================================
Note: Each "X" represents 92.4 cases.
```


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



Figure 12.2
Item plot for reading items

```
     Students          Item difficulties
-----------------------------------------
          5              |
                         |
                         |
                         |
                         |
                         |
          4              |
                         |
                         |
                         |
                         |
                       X |
          3           XX |
                      XX |
                       X |
                      XX |
                     XXX |
                    XXXX |13
          2         XXXX |
                    XXXX |
                  XXXXXX |
                  XXXXXX |8
                  XXXXXX |12
                XXXXXXXX |15
          1   XXXXXXXXX |9 20
               XXXXXXXX |16
             XXXXXXXXXX |
               XXXXXXXX |2
                XXXXXXX |6
                XXXXXXX |18 22 25 27
          0     XXXXXXX |3 21 26
               XXXXXXXX |7
              XXXXXXXXX |17 24
                XXXXXXX |14
                 XXXXXX |28
                    XXX |4
         -1         XXX |
                   XXXX |19
                     XX |23
                      X |10 11
                      X |1
                      X |
         -2           X |5
                      X |
                      X |
                      X |
                        |
                        |
         -3             |
                        |
                        |
                        |
=========================================
Note: Each "X" represents 91.8 cases.
```


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



Figure 12.3
Item plot for science items

```
          Students        Item difficulties
-----------------------------------------------------
             4             |
                           |
                           |
                           |
                           |
                           |
                           |
             3             |
                           |
                           |
                         X |
                        XX |
                        XX |
                         X |61 106
             2          XX |3
                      XXXX |
                      XXXX |22
                      XXXX |
                      XXXX |101
                    XXXXXXX|24 39
                     XXXXXX|5
             1      XXXXXXX|10 14 58 105
                   XXXXXXXX|2 29 34 65 80 86 99
                  XXXXXXXXX|17 26 38 54 89 95
                  XXXXXXXXX|13 27 33 37 55 75 83 84
                 XXXXXXXXXX|6 15 47 57 88 97
                   XXXXXXXX|4 25 49
                     XXXXXX|11 18 63 81 87 98 100
             0      XXXXXXX|1 32 76 90 94 108
                    XXXXXXX|19 21 50 62 67 85 102
                   XXXXXXXX|12 56 64 71 74 78 107
                 XXXXXXXXXX|16 20 23 36 44 45 91 104
                   XXXXXXXX|
                    XXXXXXX|28 40 53 77 79
                       XXXX|9 31 42 69 70
            -1        XXXXX|48 66 72 73
                      XXXXX|46 51 92
                        XXX|30 43 68
                        XXX|7
                        XXX|82
                         XX|
                          X|52 96
            -2            X|
                          X|8 103
                          X|
                           |
                           |
                           |
                           |
            -3             |
                           |
                           |
=====================================================
Note: Each "X" represents 87.0 cases.
```


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



Figure 12.4
Item plot for interest items

```
            Students        Item difficulties
-------------------------------------------------------
                            |
                            |
                            |
                            |
                            |
               4            |
                            |
                            |
                            |
                            |
                            |
                            |
               3            |
                            |
                            |
                          X |
                          X |
                          X |
               2          X |
                         XX |
                        XXX |
                        XXX |
                        XXX |
                       XXXXX|
                       XXXXX|
               1      XXXXXX|16
                     XXXXXXX|6 15
                   XXXXXXXXX|45 46
                     XXXXXXX|2 3 17
                  XXXXXXXXXX|1 4 5 33 36 37 38 47
                    XXXXXXXX|9 32 40 49
                     XXXXXXX|10 11 26 48
               0     XXXXXXXX|7 13 22 34 39
                     XXXXXXX|27 30 31 50
                    XXXXXXXX|12 18 21 24 25 28 29 35 52
                   XXXXXXXXX|14 19
                      XXXXXX|8 20 23 41 44
                       XXXXX|51
              -1         XXX|43
                         XXX|42
                         XXX|
                          XX|
                          XX|
                           X|
                           X|
              -2           X|
                           X|
                           X|
                           X|
                            |
                            |
                            |
              -3            |
                            |
=======================================================
 Note: Each "X" represents 101.5 cases.
```


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



Figure 12.5
Item plot for support items

```
  Students          Item difficulties
 ------------------------------------
                      |
       4              |
                      |
                      |
                      |
                      |
                    X |
                      |
                    X |
                   XX |
       3           XX |
                    X |
                   XX |
                  XXX |
                  XXX |
                  XXX |
                  XXX |
                 XXXXX|
       2         XXXXX|
                 XXXXX|
               XXXXXXX|
              XXXXXXXX|
               XXXXXXX|
             XXXXXXXXX|
              XXXXXXXX|
                XXXXXX|
               XXXXXXX|
       1      XXXXXXXX|
              XXXXXXXX|3
            XXXXXXXXXX|2
               XXXXXXX|1 11
                XXXXXX|16
                  XXXX|13 17 19 27
                  XXXX|
                  XXXX|4 8 12 20 29 30
       0            XX|5 14 21 31
                    XX|10 15 28 32
                    XX|6 37
                     X|7 22 23 36
                     X|18 26 33 34
                     X|9 35
                      |24
                     X|
                      |25
      -1              |
                      |
 ====================================
 Note: Each "X" represents 99.3 cases.
```


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



Table 12.2
Reliabilities of each of the four overall scales when scaled separately

Domain	Reliability (PV)	Reliability (WLE)
Mathematics	0.613	0.784
Reading	0.429	0.780
Science	0.856	0.832
Interest	0.886	0.867
Support	0.725	0.705

Table 12.3
Latent correlation between the five domains

	Reading		Science		Interest		Support	
	r	SE	r	SE	r	SE	r	SE
Mathematics								
OECD	0.80	0.0009	0.89	0.0006	-0.09	0.0022	0.19	0.0025
All	0.79	0.0008	0.88	0.0004	-0.21	0.0027	0.14	0.0020
Reading								
OECD			0.84	0.0008	-0.09	0.0016	0.22	0.0014
All			0.83	0.0007	-0.18	0.0022	0.18	0.0019
Science								
OECD					-0.06	0.0022	0.25	0.0022
All					-0.19	0.0018	0.19	0.0019
Interest								
OECD							0.60	0.0014
All							0.60	0.0009

Science scales

A five-dimensional scaling was performed on the achievement data, consisting of mathematics, reading, and the three competency scales of science: *explaining phenomena scientifically, identifying scientific issues* and *using scientific evidence*. Responses within these domains were included in the scaling model to improve the estimation of posterior distributions of the science competency scales. The correlations between the scales are given in Table 12.4.

Table 12.4
Latent correlation between the science scales

	Identifying scientific issues		Using scientific evidence	
	r	SE	r	SE
Explaining phenomena scientifically				
OECD	0.90	0.0005	0.93	0.0002
All	0.89	0.0003	0.93	0.0002
Identifying scientific issues				
OECD			0.91	0.0002
All			0.90	0.0003
Using scientific evidence				
OECD	0.90	0.0005	0.93	0.0002
All	0.89	0.0003	0.93	0.0002

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



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 NPMs. The consultations resulted in the deletion of a number of items at the national level.

At the international level, five science items were deleted from scaling (*S421Q02*, *S456Q01T*, *S456Q02*, *S426Q01* and *S508Q04*). Of these five items, *S421Q02*, *S456Q01T*, and *S456Q02* were deleted because they were misconceived by students, not because of an error in the source version. For this reason, they were added to the public student questionnaire database, but excluded from the data files with responses to cognitive items. The nationally deleted items are listed in Table 12.5. All deleted items were recoded as not applicable and were excluded from both international scaling and generating plausible values

Table 12.5
Items deleted at the national level

Item	Country
M273Q01	Lithuania (booklet 13)
M302Q01T	Turkey
M302Q02	Turkey
M420Q01T	Korea
M442Q02	Iceland (booklet 7)
M464Q01T	Croatia
M800Q01	Uruguay
R055Q01	Lithuania (booklets 6 and 11)
R102Q04A	Israel (booklets 2 and 12 of Arabic-language version), Korea
R102Q05	Chile, Israel (booklets 2 and 12 of Arabic-language version), Tunisia
R102Q07	Israel (booklets 2 and 12 of Arabic-language version), Luxembourg, Austria
R111Q02B	Lithuania, Slovak Republic
R219Q01E	Turkey
R219Q01T	Turkey
R220Q02B	Brazil (booklet 12), Denmark
R220Q06	Estonia (Russian-language version)
R227Q02	Azerbaijan (booklet 13 of Azerbaijani-language version)
S131Q04T	Mexico
S268Q02	Norway
S437Q03	Russian Federation
S447Q02	Switzerland (Italian-language version)
S447Q03	Hungary, Slovak Republic
S465Q04	Switzerland (Italian-language version)
S466Q01	Japan
S495Q03	Azerbaijan (booklet 2 of Azerbaijani-language version)
S495Q04T	Switzerland (Italian-language version), Poland
S519Q01	Azerbaijan (Russian-language version)
S519Q03	Sweden
S524Q07	Norway

Table 12.6
Final reliability of the PISA scales

Domain	Reliability
Mathematics	0.892
Reading	0.891
Science	0.920
Explaining phenomena scientifically	0.912
Identifying scientific issues	0.904
Using scientific evidence	0.923
Interest	0.892
Support	0.818

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



Table 12.7
National reliabilities for the main domains

		Mathematics	Reading	Science	Interest	Support
OECD	Australia	0.88	0.88	0.92	0.90	0.83
	Austria	0.90	0.93	0.94	0.89	0.82
	Belgium	0.92	0.91	0.94	0.89	0.80
	Canada	0.88	0.88	0.91	0.91	0.83
	Czech Republic	0.90	0.91	0.92	0.89	0.77
	Denmark	0.88	0.90	0.92	0.90	0.80
	Finland	0.87	0.87	0.90	0.90	0.82
	France	0.90	0.89	0.93	0.89	0.80
	Germany	0.92	0.92	0.93	0.89	0.83
	Greece	0.86	0.87	0.91	0.88	0.80
	Hungary	0.91	0.90	0.91	0.88	0.80
	Iceland	0.88	0.89	0.92	0.91	0.85
	Ireland	0.89	0.88	0.92	0.89	0.80
	Italy	0.91	0.92	0.93	0.88	0.85
	Japan	0.89	0.86	0.91	0.91	0.86
	Korea	0.89	0.89	0.91	0.90	0.82
	Luxembourg	0.89	0.89	0.93	0.89	0.82
	Mexico	0.87	0.86	0.89	0.85	0.80
	Netherlands	0.93	0.92	0.93	0.90	0.78
	New Zealand	0.89	0.90	0.93	0.91	0.84
	Norway	0.88	0.88	0.91	0.91	0.85
	Poland	0.89	0.87	0.91	0.88	0.79
	Portugal	0.89	0.90	0.92	0.87	0.83
	Slovak Republic	0.90	0.89	0.92	0.87	0.80
	Spain	0.90	0.91	0.92	0.89	0.82
	Sweden	0.88	0.89	0.92	0.90	0.83
	Switzerland	0.90	0.90	0.93	0.89	0.81
	Turkey	0.88	0.85	0.91	0.90	0.85
	United Kingdom	0.89	0.88	0.93	0.90	0.82
	United States	0.90	m	0.93	0.91	0.83
Partners	Argentina	0.86	0.85	0.90	0.86	0.78
	Azerbaijan	0.85	0.80	0.84	0.83	0.80
	Brazil	0.87	0.85	0.90	0.86	0.79
	Bulgaria	0.88	0.89	0.92	0.90	0.83
	Chile	0.88	0.83	0.90	0.88	0.79
	Colombia	0.85	0.81	0.87	0.83	0.77
	Croatia	0.88	0.91	0.91	0.88	0.77
	Estonia	0.88	0.91	0.91	0.87	0.78
	Hong Kong-China	0.89	0.88	0.92	0.91	0.83
	Indonesia	0.86	0.85	0.87	0.83	0.81
	Israel	0.88	0.87	0.91	0.91	0.82
	Jordan	0.86	0.86	0.90	0.85	0.83
	Kyrgyzstan	0.83	0.83	0.85	0.81	0.77
	Latvia	0.87	0.87	0.90	0.86	0.77
	Liechtenstein	0.88	0.91	0.93	0.88	0.81
	Lithuania	0.89	0.88	0.92	0.88	0.80
	Macao-China	0.85	0.81	0.89	0.89	0.79
	Montenegro	0.85	0.88	0.89	0.88	0.77
	Qatar	0.85	0.86	0.88	0.89	0.87
	Romania	0.87	0.86	0.90	0.86	0.83
	Russian Federation	0.84	0.82	0.89	0.85	0.79
	Serbia	0.87	0.87	0.90	0.88	0.78
	Slovenia	0.90	0.93	0.93	0.90	0.80
	Chinese Taipei	0.90	0.87	0.92	0.91	0.81
	Thailand	0.84	0.85	0.88	0.85	0.86
	Tunisia	0.86	0.83	0.87	0.85	0.83
	Uruguay	0.86	0.84	0.90	0.86	0.76



Table 12.8
National reliabilities for the science scales

		Explaining phenomena scientifically	Identifying scientific issues	Using scientific evidence
OECD	Australia	0.90	0.90	0.92
	Austria	0.92	0.93	0.93
	Belgium	0.93	0.92	0.94
	Canada	0.90	0.90	0.91
	Czech Republic	0.91	0.92	0.93
	Denmark	0.91	0.90	0.92
	Finland	0.89	0.88	0.90
	France	0.92	0.92	0.93
	Germany	0.93	0.92	0.93
	Greece	0.90	0.90	0.92
	Hungary	0.91	0.91	0.92
	Iceland	0.92	0.90	0.93
	Ireland	0.92	0.91	0.93
	Italy	0.92	0.92	0.93
	Japan	0.92	0.90	0.92
	Korea	0.90	0.91	0.91
	Luxembourg	0.91	0.90	0.93
	Mexico	0.89	0.88	0.90
	Netherlands	0.92	0.89	0.93
	New Zealand	0.92	0.91	0.92
	Norway	0.90	0.90	0.92
	Poland	0.91	0.89	0.92
	Portugal	0.92	0.88	0.92
	Slovak Republic	0.91	0.92	0.93
	Spain	0.91	0.92	0.93
	Sweden	0.92	0.89	0.93
	Switzerland	0.93	0.91	0.93
	Turkey	0.90	0.90	0.92
	United Kingdom	0.92	0.91	0.93
	United States	0.92	0.90	0.93
Partners	Argentina	0.89	0.89	0.91
	Azerbaijan	0.85	0.87	0.87
	Brazil	0.89	0.88	0.92
	Bulgaria	0.91	0.91	0.92
	Chile	0.89	0.88	0.91
	Colombia	0.88	0.86	0.90
	Croatia	0.91	0.91	0.92
	Estonia	0.89	0.90	0.91
	Hong Kong-China	0.91	0.92	0.92
	Indonesia	0.84	0.82	0.85
	Israel	0.91	0.89	0.92
	Jordan	0.88	0.86	0.90
	Kyrgyzstan	0.83	0.88	0.89
	Latvia	0.89	0.89	0.92
	Liechtenstein	0.92	0.91	0.93
	Lithuania	0.91	0.90	0.92
	Macao-China	0.88	0.85	0.89
	Montenegro	0.87	0.87	0.88
	Qatar	0.87	0.86	0.87
	Romania	0.88	0.88	0.90
	Russian Federation	0.88	0.88	0.90
	Serbia	0.88	0.88	0.91
	Slovenia	0.91	0.91	0.93
	Chinese Taipei	0.91	0.88	0.91
	Thailand	0.88	0.88	0.89
	Tunisia	0.86	0.88	0.90
	Uruguay	0.90	0.88	0.92



International scaling

The international scaling was performed on the calibration data set of 15 000 students (500 randomly selected students from each of the 30 countries). The item parameter estimates from this scaling are reported in Appendix 1. The item parameters were estimated using four separate one-dimensional models. As in previous cycles, a booklet facet was used in the item response model.

Generating student scale scores

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 12.6 gives the reliabilities at the international level for the generated scale scores. The increase in reliability of the results reported in Table 12.6 over those presented in Table 12.2 is due to the use of multi-dimensional scaling and conditioning.

TEST LENGTH ANALYSIS

Table 12.9 shows the number of missing responses and the number of missing responses recoded as not reached, by booklet. A response is coded as missing 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 non-reached code, except for the first value of the missing series, which is coded as missing (see Chapter 18).

Table 12.9
Average number of not-reached items and missing items by booklet

	Missing		Not Reached	
Booklet	Weighted	Unweighted	Weighted	Unweighted
1	4.44	4.51	3.32	2.38
2	5.40	5.46	1.84	1.50
3	4.80	4.82	1.68	1.38
4	5.33	5.41	1.71	1.38
5	5.22	5.39	5.39	4.10
6	5.48	5.94	3.81	2.95
7	5.87	5.86	1.97	1.63
8	5.65	5.70	2.62	1.97
9	5.74	6.01	2.70	2.18
10	5.46	5.59	3.48	2.88
11	5.83	6.21	3.30	2.64
12	6.07	6.20	3.37	2.66
13	5.05	5.25	2.03	1.58
UH	5.27	4.04	0.66	0.42
Total	**5.41**	**5.56**	**2.86**	**2.25**

Table 12.10 shows this information by country over all booklets. 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 12.10 provides the percentage distribution of not-reached items per booklet. The percentage of students who reached the last item ranges from 76 to 87% when using weighted data and 79 to 90% when using unweighted data (*i.e.*, the percentages of students with zero not-reached items).

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



Table 12.10
Average number of not-reached items and missing items by country

		Missing		Not Reached	
		Weighted	Unweighted	Weighted	Unweighted
OECD	Australia	3.27	3.52	0.70	0.85
	Austria	4.70	4.62	0.41	0.38
	Belgium	4.14	3.89	0.99	0.90
	Canada	2.57	2.95	1.04	0.92
	Czech Republic	5.68	4.83	0.54	0.50
	Denmark	5.21	5.21	1.29	1.29
	Finland	2.50	2.50	0.59	0.58
	France	5.60	5.56	1.88	1.84
	Germany	4.83	4.81	0.67	0.66
	Greece	6.93	6.91	1.93	1.89
	Hungary	4.62	4.40	0.79	0.69
	Iceland	4.23	4.21	1.24	1.21
	Ireland	3.44	3.37	0.72	0.73
	Italy	7.47	6.88	1.74	1.54
	Japan	5.31	5.22	0.91	0.89
	Korea	2.82	2.85	0.32	0.32
	Luxembourg	5.45	5.40	0.88	0.87
	Mexico	4.14	3.89	5.42	4.78
	Netherlands	1.26	1.15	0.18	0.17
	New Zealand	3.30	3.21	0.94	0.88
	Norway	5.78	5.78	1.36	1.30
	Poland	4.89	4.70	0.84	0.82
	Portugal	5.74	5.46	1.46	1.42
	Slovak Republic	6.00	5.82	0.75	0.71
	Spain	5.96	4.96	1.70	1.35
	Sweden	4.93	4.92	1.29	1.33
	Switzerland	4.16	4.24	0.77	0.76
	Turkey	6.81	6.45	1.62	1.60
	United Kingdom	4.16	4.10	1.11	0.87
	United States	2.78	2.87	0.46	0.43
Partners	Argentina	11.23	10.87	9.97	9.34
	Azerbaijan	13.37	13.17	0.26	0.26
	Brazil	9.29	9.69	6.04	6.42
	Bulgaria	11.45	11.11	3.09	2.87
	Chile	7.67	7.44	5.09	4.98
	Colombia	5.97	5.92	12.72	12.47
	Croatia	3.91	3.93	0.75	0.75
	Estonia	3.32	3.33	0.56	0.58
	Hong Kong-China	2.39	2.23	0.74	0.71
	Indonesia	5.29	5.96	4.09	4.36
	Israel	9.73	9.62	3.16	3.16
	Jordan	6.75	6.17	3.37	3.19
	Kyrgyzstan	16.67	16.43	12.37	11.84
	Latvia	4.03	3.91	1.41	1.42
	Liechtenstein	3.71	3.73	0.42	0.42
	Lithuania	5.36	5.41	1.04	1.10
	Macao-China	3.17	3.31	1.58	1.54
	Montenegro	12.04	12.19	1.26	1.33
	Qatar	11.40	11.27	3.22	3.15
	Romania	6.95	7.29	1.09	1.24
	Russian Federation	6.18	6.09	4.33	4.26
	Serbia	10.69	10.58	2.01	1.91
	Slovenia	5.00	6.17	0.31	0.43
	Chinese Taipei	3.16	2.76	0.56	0.49
	Thailand	4.36	4.32	2.07	2.12
	Tunisia	8.56	8.51	6.10	6.02
	Uruguay	10.65	10.15	7.79	7.31

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



Table 12.11
Distribution of not-reached items by booklet

Number of not-reached items	Booklet													
	1	2	3	4	5	6	7	8	9	10	11	12	13	UH
	Weighted percentage													
0	81.6	82.4	76.2	80.5	77.9	82.3	80.2	86.8	76.9	82.3	82.6	80.5	84.3	84.2
1	1.8	0.6	7.2	2.3	1.0	0.8	0.5	0.4	1.7	1.1	0.3	0.3	0.6	3.5
2	0.7	0.5	0.8	3.4	0.7	0.3	1.7	0.9	3.4	1.8	0.4	1.7	0.6	5.6
3	1.1	0.5	2.4	3.0	1.3	0.3	2.5	0.4	1.5	0.4	0.2	1.8	0.6	1.2
4	0.3	3.9	4.0	1.1	1.1	0.5	0.6	0.8	0.8	0.5	0.1	0.5	1.4	0.9
5	1.2	0.1	0.7	1.7	0.5	0.1	0.8	0.4	3.7	0.6	1.1	1.1	0.4	1.4
6	0.1	0.3	0.6	0.2	0.1	0.6	2.7	0.2	0.8	0.5	2.0	1.5	0.6	0.4
7	1.0	1.8	0.3	0.5	0.1	0.5	1.3	0.5	0.7	1.2	0.6	0.3	1.6	0.1
8	1.0	1.2	0.5	0.3	0.7	0.6	0.6	0.1	1.3	0.1	0.9	0.3	0.8	0.1
>8	11.5	8.6	7.2	7.0	16.5	14.1	9.0	9.6	9.5	11.3	11.8	12.0	9.1	2.4
	Unweighted percentage													
0	85.0	84.6	79.9	82.3	81.7	85.4	82.6	89.8	79.2	84.4	85.7	82.6	86.8	88.7
1	1.8	0.6	5.6	2.6	1.0	1.0	0.7	0.3	1.7	1.1	0.3	0.4	0.6	2.8
2	0.5	0.4	0.8	3.0	0.7	0.4	1.6	0.8	3.3	1.9	0.4	1.9	0.6	3.5
3	1.0	0.6	2.5	3.2	1.3	0.2	2.7	0.4	1.4	0.3	0.1	2.2	0.6	1.0
4	0.4	4.0	3.5	0.8	1.2	0.4	0.6	0.8	0.7	0.5	0.1	0.6	1.4	0.4
5	1.1	0.2	0.8	1.6	0.4	0.1	1.0	0.2	3.9	0.7	1.0	0.7	0.5	1.9
6	0.1	0.3	0.7	0.2	0.1	0.5	2.1	0.1	0.7	0.5	1.7	1.7	0.6	0.5
7	0.8	1.6	0.2	0.5	0.1	0.5	1.3	0.2	0.5	0.9	0.5	0.3	1.5	0.1
8	0.7	1.0	0.4	0.3	0.6	0.4	0.5	0.1	1.1	0.2	0.7	0.3	0.6	0.1
>8	8.6	6.6	5.4	5.4	12.9	11.0	7.0	7.1	7.6	9.5	9.5	9.3	7.0	1.1

Table 12.12
Estimated booklet effects on the PISA scale

Booklet	Domains				
	Mathematics	Reading	Science	Interest	Support
1			–3.1	0.6	2.6
2	2.4	10.1	–20.0	4.9	–3.3
3	11.8		–20.5	–10.5	1.1
4	4.2		–6.3	–10.7	–8.4
5			1.6	–4.5	–1.5
6		–1.2	6.7	–5.2	–1.1
7	13.6	–17.6	–19.9	13.9	3.2
8	–10.8		17.6	3.3	–3.3
9	–10.1	40.4	0.2	–0.6	4.1
10	–11.7		21.4	–1.2	–2.0
11	–13.9	19.2	12.1	–2.2	1.7
12	–5.1	–33.6	20.4	–2.5	5.9
13	19.5	–17.2	–10.4	14.6	0.9

BOOKLET EFFECTS

The booklet parameters that are described in Chapter 9 are reported in Table 12.13. 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 12.13 shows the booklet effects after transformation to the PISA scales.

Table 12.13
Estimated booklet effects in logits

Booklet	Domains				
	Mathematics	Reading	Science	Interest	Support
1			–0.033	0.007	0.023
2	0.031	0.126	–0.214	0.055	–0.029
3	0.152		–0.220	–0.118	0.010
4	0.054		–0.068	–0.120	–0.073
5			0.017	–0.050	–0.013
6		–0.015	0.072	–0.058	–0.010
7	0.175	–0.220	–0.213	0.156	0.028
8	–0.139		0.189	0.037	–0.029
9	–0.130	0.506	0.002	–0.007	0.036
10	–0.150		0.229	–0.013	–0.017
11	–0.178	0.240	0.130	–0.025	0.015
12	–0.065	–0.421	0.219	–0.028	0.051
13	0.250	–0.216	–0.112	0.163	0.008



Table 12.14 [Part 1/2]
Variance in mathematics booklet means

		Expected mean	Booklet 1		Boolet 2		Booklet 3		Boolet 4		Boolet 5		Boolet 6		Boolet 7	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	Australia	520	521	0.14	526	2.99	515	1.82	517	0.90	521	0.25	521	0.18	524	1.49
	Austria	507	505	0.26	513	1.35	515	1.65	507	0.00	503	0.65	506	0.04	509	0.08
	Belgium	526	526	0.00	520	2.51	531	1.17	521	2.03	527	0.03	526	0.01	529	0.60
	Canada	527	528	0.02	529	0.11	528	0.02	532	1.63	536	**5.17**	529	0.28	529	0.11
	Czech Republic	516	513	0.15	521	1.11	520	0.62	513	0.20	515	0.00	515	0.00	511	0.58
	Denmark	513	508	0.92	519	1.01	521	2.01	510	0.43	511	0.17	518	0.70	514	0.02
	Finland	548	550	0.28	570	**19.88**	550	0.25	546	0.10	546	0.07	553	1.58	550	0.41
	France	496	492	0.52	491	0.63	489	1.73	499	0.51	496	0.00	502	1.93	493	0.44
	Germany	509	516	1.71	504	1.14	515	0.94	512	0.44	505	0.48	505	0.60	511	0.11
	Greece	460	456	0.33	457	0.18	448	**5.35**	464	0.74	456	0.30	469	2.50	452	1.72
	Hungary	491	494	0.38	501	**3.97**	489	0.11	499	2.97	485	1.01	490	0.01	485	1.40
	Iceland	506	510	0.33	495	**3.96**	501	0.74	493	**5.75**	505	0.03	500	1.34	510	0.48
	Ireland	502	500	0.19	515	**4.93**	490	**4.31**	515	**9.59**	501	0.01	499	0.43	492	3.64
	Italy	461	456	2.04	472	**6.78**	459	0.43	463	0.27	460	0.08	460	0.12	456	2.84
	Japan	524	524	0.00	496	**26.38**	540	**12.49**	517	2.16	525	0.06	527	0.53	527	0.41
	Korea	548	553	0.91	540	2.11	545	0.23	551	0.28	545	0.18	545	0.16	552	0.47
	Luxembourg	490	486	0.41	473	**13.72**	497	1.87	485	0.96	488	0.07	494	0.74	500	3.82
	Mexico	407	403	0.71	404	0.32	398	**4.81**	397	**4.27**	407	0.00	412	1.14	393	**9.21**
	Netherlands	535	536	0.01	533	0.14	538	0.47	531	0.83	539	0.54	533	0.23	541	1.83
	New Zealand	522	518	0.64	526	0.41	523	0.04	529	2.21	515	1.73	523	0.10	527	1.06
	Norway	490	486	0.68	488	0.10	497	1.61	489	0.02	484	1.13	484	1.35	498	2.20
	Poland	495	494	0.05	496	0.03	499	0.50	496	0.01	502	1.93	496	0.00	489	1.94
	Portugal	466	466	0.00	468	0.10	464	0.19	475	3.80	467	0.00	458	2.37	471	0.71
	Slovak Republic	494	494	0.00	489	0.77	495	0.03	500	1.42	492	0.21	495	0.12	501	1.58
	Spain	480	482	0.18	482	0.32	483	0.48	477	0.56	482	0.24	477	0.22	471	**4.63**
	Sweden	502	506	0.57	502	0.00	508	0.77	504	0.12	497	1.13	499	0.24	502	0.00
	Switzerland	529	530	0.05	516	**9.18**	536	2.18	527	0.22	537	2.93	531	0.09	534	0.84
	Turkey	424	425	0.03	439	**4.92**	433	1.88	424	0.00	431	1.08	424	0.00	414	2.21
	United Kingdom	496	492	0.65	499	0.49	499	1.02	502	2.11	498	0.19	498	0.26	490	1.65
	United States	475	481	1.16	471	0.53	466	2.54	480	0.78	476	0.04	470	0.62	470	0.71
	OECD average	**499**	**498**	**0.00**	**499**	**0.00**	**500**	**0.05**	**499**	**0.01**	**498**	**0.00**	**499**	**0.00**	**498**	**0.01**
Partners	Argentina	382	380	0.06	389	0.64	368	2.59	385	0.17	384	0.06	374	1.17	355	**6.72**
	Azerbaijan	474	475	0.09	454	**56.07**	**510**	**164.86**	442	**152.10**	477	0.58	484	**4.44**	490	**26.52**
	Brazil	372	366	1.39	369	0.26	362	2.57	359	**4.25**	376	0.40	372	0.00	365	2.22
	Bulgaria	414	416	0.08	401	2.88	403	1.79	414	0.00	410	0.25	415	0.00	414	0.01
	Chile	413	411	0.03	431	**11.89**	398	**5.47**	415	0.33	408	0.50	414	0.04	398	**5.28**
	Chinese Taipei	549	548	0.03	546	0.50	561	**4.36**	543	1.61	550	0.01	549	0.01	546	0.40
	Colombia	371	379	1.18	362	1.99	351	**6.63**	368	0.35	363	1.63	367	0.44	348	**13.74**
	Croatia	468	468	0.01	466	0.10	478	**5.73**	462	1.69	472	1.60	466	0.06	468	0.00
	Estonia	514	516	0.08	503	**5.21**	520	1.68	521	1.57	515	0.00	519	0.75	505	3.77
	Hong Kong-China	548	551	0.35	554	2.08	548	0.01	544	0.51	548	0.00	547	0.04	550	0.28
	Indonesia	391	396	0.62	393	0.12	386	0.45	390	0.00	390	0.02	388	0.12	390	0.02
	Israel	443	447	0.36	438	0.36	436	0.68	435	2.05	434	1.19	443	0.00	429	**3.87**
	Jordan	384	384	0.00	392	2.79	385	0.02	385	0.01	382	0.16	379	0.91	364	**16.38**
	Kyrgyzstan	312	310	0.15	316	0.64	282	**28.46**	315	0.51	313	0.05	308	0.48	300	**5.08**
	Latvia	486	487	0.01	484	0.14	484	0.24	483	0.24	490	0.55	493	1.18	473	**5.23**
	Liechtenstein	525	518	0.12	523	0.04	523	0.01	539	0.40	551	1.86	497	1.68	514	0.30
	Lithuania	486	492	1.20	496	**3.95**	494	2.55	480	1.95	484	0.13	486	0.00	473	**7.66**
	Macao-China	525	522	0.27	527	0.06	531	1.48	533	2.48	524	0.07	527	0.05	516	2.91
	Montenegro	399	405	1.16	392	2.19	404	0.69	394	0.81	406	1.26	401	0.07	402	0.49
	Qatar	318	314	0.71	302	**11.84**	301	**14.69**	311	2.42	318	0.00	315	0.31	320	0.23
	Romania	414	414	0.00	408	0.98	410	0.45	398	**10.04**	414	0.00	414	0.01	412	0.07
	Russian Federation	476	473	0.21	479	0.38	469	0.97	470	1.04	473	0.18	477	0.07	465	**5.51**
	Serbia	436	442	0.84	432	0.68	434	0.10	426	3.76	432	0.35	437	0.04	433	0.28
	Slovenia	506	505	0.09	495	2.99	513	1.41	503	0.22	509	0.43	502	0.58	507	0.03
	Thailand	417	416	0.03	436	**16.77**	412	1.58	422	1.18	418	0.06	420	0.36	408	3.74
	Tunisia	368	369	0.02	392	**14.45**	345	**13.40**	362	1.10	364	0.38	363	0.56	350	**7.26**
	Uruguay	430	436	1.14	425	0.67	401	**21.24**	407	**14.40**	422	2.60	435	0.52	398	**20.78**

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



Table 12.14 [Part 2/2]
Variance in mathematics booklet means

		Boolet 8		Boolet 9		Boolet 10		Boolet 11		Boolet 12		Boolet 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	0.07	515	1.83	518	0.18	521	0.10	517	0.45	521	0.20	10.6
	Austria	497	3.33	505	0.13	504	0.25	513	1.16	500	1.44	517	2.48	12.8
	Belgium	522	1.15	516	**6.09**	523	0.66	527	0.01	540	**9.07**	536	**6.16**	**29.5**
	Canada	527	0.03	519	**5.20**	519	2.98	520	3.44	523	0.86	533	2.22	**22.1**
	Czech Republic	511	0.69	504	3.16	512	0.31	524	1.87	513	0.20	528	**4.41**	13.3
	Denmark	515	0.25	508	0.97	513	0.00	500	**4.95**	519	1.29	514	0.01	12.7
	Finland	540	3.01	531	**10.41**	546	0.05	535	**10.75**	544	0.93	566	**12.07**	**59.8**
	France	497	0.02	504	3.26	491	1.07	496	0.01	493	0.19	498	0.16	10.5
	Germany	508	0.07	508	0.09	515	1.02	514	0.71	499	2.65	507	0.34	10.3
	Greece	461	0.10	466	1.99	473	**8.36**	474	**8.21**	452	2.60	442	**12.22**	**44.6**
	Hungary	491	0.00	505	**6.68**	495	0.42	481	**3.95**	484	2.29	483	2.20	**25.4**
	Iceland	512	0.94	522	**11.96**	503	0.27	508	0.10	511	0.55	505	0.07	**26.5**
	Ireland	503	0.12	505	0.52	490	**5.31**	484	**12.32**	511	3.39	513	**4.57**	**49.3**
	Italy	458	0.78	463	0.11	470	**7.74**	479	**16.99**	454	**4.72**	452	**5.93**	**48.8**
	Japan	526	0.23	526	0.14	514	**4.38**	525	0.04	535	**6.08**	520	0.52	**53.4**
	Korea	546	0.07	550	0.15	534	**5.38**	549	0.05	557	3.08	551	0.44	13.5
	Luxembourg	495	1.28	496	1.23	489	0.02	493	0.47	488	0.14	487	0.30	**25.0**
	Mexico	409	0.27	401	2.12	419	**9.34**	428	**35.67**	404	0.44	401	3.23	**71.5**
	Netherlands	539	0.39	536	0.02	521	**8.72**	523	**5.82**	546	**4.29**	540	1.00	**24.3**
	New Zealand	523	0.05	521	0.00	518	0.76	513	3.34	526	0.81	526	0.50	11.6
	Norway	494	0.58	493	0.24	493	0.21	490	0.01	483	1.28	489	0.04	9.5
	Poland	500	0.73	486	**4.48**	487	2.25	493	0.17	508	**6.81**	495	0.01	18.9
	Portugal	461	1.14	466	0.01	472	0.93	473	1.13	457	2.82	464	0.11	13.3
	Slovak Republic	491	0.26	509	**8.03**	498	0.50	496	0.23	480	**8.32**	482	**5.51**	**27.0**
	Spain	477	0.58	478	0.31	481	0.04	484	1.38	485	1.82	480	0.00	10.8
	Sweden	500	0.31	504	0.17	497	0.94	507	0.77	496	1.01	506	0.44	6.5
	Switzerland	530	0.04	524	1.68	529	0.02	528	0.07	530	0.05	532	0.25	17.6
	Turkey	413	2.66	413	3.04	435	3.78	419	0.68	414	2.04	427	0.23	**22.6**
	United Kingdom	492	0.49	491	1.28	497	0.08	482	**9.52**	499	0.78	503	2.72	**21.2**
	United States	484	2.12	488	**6.40**	465	2.31	447	**23.67**	487	**5.08**	482	1.48	**47.4**
	OECD average	**498**	**0.01**	**498**	**0.01**	**497**	**0.05**	**498**	**0.04**	**499**	**0.00**	**500**	**0.07**	**0.3**
Partners	Argentina	394	1.89	386	0.18	401	**5.81**	395	3.16	397	2.84	348	**17.89**	**43.2**
	Azerbaijan	465	**8.45**	440	**132.24**	470	2.52	480	2.90	503	**93.00**	498	**42.54**	**686.3**
	Brazil	365	2.78	373	0.01	390	**12.84**	402	**57.32**	366	1.60	341	**30.03**	**115.7**
	Bulgaria	413	0.05	422	1.11	429	3.65	434	**8.63**	398	**4.17**	405	1.23	**23.8**
	Chile	421	2.05	412	0.00	421	2.73	428	**6.09**	396	**7.13**	395	**7.70**	**49.2**
	Chinese Taipei	549	0.02	539	**4.44**	545	0.60	537	**4.68**	561	**6.17**	570	**11.84**	**34.7**
	Colombia	380	1.82	370	0.03	399	**28.38**	416	**59.88**	376	0.53	330	**50.73**	**167.3**
	Croatia	458	**4.25**	454	**10.60**	472	1.32	481	**8.57**	458	**4.60**	471	0.86	**39.4**
	Estonia	520	1.01	519	0.72	512	0.30	508	1.48	520	1.34	511	0.41	18.3
	Hong Kong-China	548	0.00	532	**12.01**	539	2.82	540	2.25	555	2.29	565	**11.77**	**34.4**
	Indonesia	403	2.81	401	2.29	388	0.21	379	**4.08**	393	0.13	386	0.59	11.5
	Israel	451	1.59	454	2.25	453	3.74	448	1.08	450	1.04	428	**4.20**	**22.4**
	Jordan	387	0.19	376	2.76	385	0.01	394	**4.01**	407	**20.87**	373	**4.52**	**52.6**
	Kyrgyzstan	319	2.55	323	**5.29**	329	**10.00**	340	**38.30**	305	1.26	276	**47.53**	**140.3**
	Latvia	489	0.36	490	0.39	485	0.02	479	1.87	495	3.32	487	0.01	13.6
	Liechtenstein	534	0.23	510	0.72	528	0.02	535	0.29	529	0.03	525	0.00	5.7
	Lithuania	477	2.44	478	2.33	483	0.40	485	0.01	495	2.71	504	**9.41**	**34.7**
	Macao-China	529	0.35	525	0.00	520	1.35	509	**8.17**	532	1.42	531	0.94	19.6
	Montenegro	394	0.78	384	**10.86**	399	0.00	420	**20.42**	392	2.18	397	0.17	**41.1**
	Qatar	325	2.13	350	**40.60**	342	**31.24**	338	**18.66**	306	**8.80**	290	**23.76**	**155.4**
	Romania	415	0.05	419	0.84	418	0.58	444	**19.60**	406	1.06	420	0.77	**34.5**
	Russian Federation	485	3.75	478	0.26	484	1.60	493	**9.40**	478	0.18	460	**6.60**	**30.2**
	Serbia	433	0.30	433	0.22	439	0.20	456	**17.95**	432	0.50	432	0.49	**25.7**
	Slovenia	506	0.00	509	0.24	502	0.57	511	0.79	504	0.20	510	0.72	8.3
	Thailand	405	**7.68**	407	**5.05**	436	**20.96**	413	1.05	410	2.16	418	0.04	**60.7**
	Tunisia	370	0.25	353	**4.29**	395	**24.59**	386	**15.84**	377	2.65	328	**38.62**	**123.4**
	Uruguay	445	**8.02**	438	2.34	450	**14.42**	451	**20.21**	450	**11.00**	396	**25.08**	**142.4**



Table 12.15 [Part 1/2]
Variance in reading booklet means

		Expected mean	Booklet 1		Boolet 2		Booklet 3		Boolet 4		Boolet 5		Boolet 6		Boolet 7	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	Australia	513	514	0.16	518	2.37	509	1.10	508	1.64	514	0.14	510	0.52	516	0.71
	Austria	491	493	0.08	500	2.19	490	0.04	484	1.45	495	0.35	493	0.09	483	2.54
	Belgium	507	509	0.11	507	0.02	502	0.82	507	0.00	508	0.09	506	0.01	501	1.37
	Canada	527	529	0.11	527	0.00	527	0.00	527	0.02	536	**3.92**	530	0.38	522	1.73
	Czech Republic	490	485	0.55	491	0.06	495	0.99	483	1.06	489	0.01	486	0.30	489	0.00
	Denmark	495	490	0.71	488	1.13	497	0.25	491	0.38	495	0.00	502	1.80	489	1.02
	Finland	546	551	1.09	563	**8.27**	549	0.22	542	0.81	544	0.18	550	0.43	556	**4.97**
	France	489	486	0.13	475	**4.09**	487	0.04	483	0.76	485	0.28	498	**4.41**	495	1.14
	Germany	505	509	0.36	507	0.10	502	0.23	508	0.25	497	1.43	503	0.18	496	2.09
	Greece	460	453	1.31	460	0.01	455	0.74	461	0.00	460	0.00	473	**4.53**	471	3.31
	Hungary	483	483	0.03	495	**4.44**	481	0.06	484	0.06	477	0.98	483	0.00	479	0.32
	Iceland	485	489	0.32	475	2.91	480	0.61	473	3.69	485	0.00	485	0.00	486	0.04
	Ireland	517	515	0.12	528	2.93	514	0.29	518	0.03	514	0.41	508	2.33	522	0.82
	Italy	469	470	0.01	457	**9.42**	476	1.97	473	0.85	465	0.67	470	0.02	462	2.73
	Japan	498	500	0.12	484	**6.25**	498	0.00	501	0.23	499	0.03	494	0.76	491	1.45
	Korea	556	562	1.09	555	0.01	558	0.11	551	0.91	554	0.17	554	0.21	543	**5.21**
	Luxembourg	479	474	1.17	487	2.06	483	0.32	479	0.00	478	0.06	480	0.01	490	3.04
	Mexico	412	411	0.13	378	**45.75**	409	0.50	408	0.62	408	0.46	429	**14.21**	417	1.04
	Netherlands	513	512	0.03	525	**5.10**	512	0.00	508	0.54	518	1.16	508	0.97	516	0.46
	New Zealand	521	514	1.06	529	1.78	528	1.46	524	0.27	511	3.48	518	0.33	526	0.77
	Norway	485	474	2.11	478	1.14	494	2.38	490	0.47	476	1.47	473	3.63	485	0.00
	Poland	508	508	0.01	509	0.07	508	0.00	515	1.28	516	1.95	503	0.83	519	**4.89**
	Portugal	473	469	0.33	476	0.33	477	0.50	481	2.16	471	0.06	471	0.11	496	**15.41**
	Slovak Republic	469	467	0.14	463	0.89	468	0.03	474	1.28	462	1.15	481	**7.15**	462	1.40
	Spain	461	464	0.22	450	**5.00**	462	0.13	462	0.14	462	0.11	460	0.02	462	0.07
	Sweden	508	509	0.01	501	0.59	500	0.83	506	0.18	512	0.44	500	1.65	513	0.87
	Switzerland	499	503	0.66	491	**4.48**	499	0.02	502	0.29	507	3.49	496	0.59	490	**5.71**
	Turkey	447	442	0.55	435	3.38	454	0.86	455	1.44	452	0.50	450	0.16	448	0.00
	United Kingdom	496	491	0.80	492	0.66	503	2.75	498	0.16	499	0.28	496	0.01	492	0.57
	United States															
	OECD average	**476**	**476**	**0.01**	**475**	**0.08**	**477**	**0.02**	**476**	**0.00**	**476**	**0.00**	**477**	**0.01**	**477**	**0.02**
Partners	Argentina	378	377	0.00	334	**12.97**	382	0.20	379	0.04	378	0.00	386	0.81	394	1.71
	Azerbaijan	353	352	0.06	350	0.54	352	0.07	351	0.34	351	0.34	362	3.46	336	**11.80**
	Brazil	393	390	0.15	373	**6.33**	395	0.06	383	1.61	401	1.67	398	0.95	407	**6.58**
	Bulgaria	402	406	0.23	393	0.95	400	0.05	411	0.83	395	0.42	412	1.30	410	0.89
	Chile	442	440	0.08	418	**11.07**	443	0.00	436	0.67	441	0.05	454	3.12	464	**8.39**
	Chinese Taipei	497	496	0.03	501	1.14	495	0.17	497	0.02	498	0.16	490	1.91	489	3.16
	Colombia	389	395	0.61	325	**58.41**	384	0.20	389	0.00	386	0.16	398	2.76	424	**22.12**
	Croatia	477	480	0.45	482	0.65	478	0.02	476	0.09	478	0.04	479	0.14	482	1.21
	Estonia	501	504	0.25	504	0.40	493	3.01	501	0.02	497	0.48	514	**7.35**	484	**8.54**
	Hong Kong-China	535	537	0.23	544	**4.72**	538	0.34	533	0.19	537	0.08	520	**18.12**	525	**4.76**
	Indonesia	392	396	0.37	385	1.43	397	0.67	393	0.01	393	0.00	398	0.47	399	0.83
	Israel	439	446	0.74	414	**10.67**	433	0.47	440	0.02	430	1.15	437	0.05	433	0.56
	Jordan	401	399	0.12	393	1.93	400	0.00	403	0.21	401	0.02	409	1.95	399	0.05
	Kyrgyzstan	284	285	0.01	267	**7.25**	282	0.13	286	0.12	289	0.55	291	1.22	309	**18.86**
	Latvia	479	483	0.33	476	0.16	479	0.00	477	0.10	484	0.58	490	3.11	488	1.74
	Liechtenstein	511	503	0.23	504	0.16	508	0.02	512	0.00	535	1.69	476	2.87	497	0.54
	Lithuania	470	477	1.62	490	**10.74**	463	1.39	473	0.18	464	1.04	472	0.07	472	0.07
	Macao-China	492	491	0.10	495	0.28	495	0.27	490	0.15	492	0.02	495	0.35	488	1.17
	Montenegro	392	397	0.79	390	0.21	391	0.01	390	0.08	397	0.86	394	0.19	380	**4.79**
	Qatar	313	305	1.44	308	0.65	306	1.84	313	0.00	317	0.53	307	0.88	309	0.41
	Romania	397	400	0.18	393	0.37	400	0.28	392	0.41	392	0.30	397	0.00	407	3.45
	Russian Federation	440	446	1.04	417	**12.27**	442	0.03	438	0.14	433	2.11	452	2.50	454	**7.60**
	Serbia	401	403	0.07	405	0.49	402	0.03	398	0.26	401	0.00	411	3.68	400	0.03
	Slovenia	498	495	0.22	487	3.40	499	0.12	492	0.90	501	0.47	504	2.21	477	**15.31**
	Thailand	417	417	0.00	411	1.24	411	1.58	416	0.07	421	0.67	423	2.36	426	**4.41**
	Tunisia	382	382	0.01	365	**6.04**	386	0.24	375	1.13	379	0.25	395	**5.38**	386	0.54
	Uruguay	414	416	0.07	360	**50.73**	410	0.24	409	0.67	411	0.20	429	**4.35**	427	2.43



Table 12.15 [Part 2/2]
Variance in reading booklet means

		Boolet 8		Boolet 9		Boolet 10		Boolet 11		Boolet 12		Boolet 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	518	1.40	518	1.99	511	0.36	512	0.01	510	0.81	509	0.77	12.0
	Austria	492	0.01	502	2.36	493	0.03	498	1.17	479	**4.72**	490	0.02	15.1
	Belgium	504	0.38	506	0.01	503	0.71	513	1.47	511	1.01	510	0.46	6.5
	Canada	529	0.13	526	0.16	523	0.66	528	0.01	520	2.84	530	0.64	10.6
	Czech Republic	486	0.22	487	0.19	490	0.00	494	0.29	488	0.06	498	2.36	6.1
	Denmark	495	0.00	509	**5.05**	494	0.00	502	1.69	479	**6.96**	498	0.47	19.5
	Finland	543	0.51	545	0.04	549	0.43	547	0.04	536	**4.94**	533	**9.96**	**31.9**
	France	491	0.17	482	1.21	488	0.01	489	0.01	491	0.10	490	0.03	12.4
	Germany	509	0.46	517	**4.54**	506	0.01	498	1.23	502	0.24	507	0.11	11.2
	Greece	457	0.21	435	**15.62**	467	1.01	445	**5.82**	476	**7.94**	464	0.30	**40.8**
	Hungary	481	0.04	484	0.05	481	0.06	481	0.09	489	2.21	471	**3.91**	12.3
	Iceland	485	0.00	513	**24.00**	483	0.10	486	0.01	475	2.62	485	0.01	**34.3**
	Ireland	524	1.15	528	3.62	511	1.01	512	0.90	515	0.18	515	0.16	14.0
	Italy	469	0.02	461	3.42	468	0.07	463	1.62	475	2.76	483	**14.56**	**38.1**
	Japan	500	0.10	508	3.82	502	0.48	498	0.00	497	0.04	503	0.65	14.0
	Korea	554	0.09	571	**6.96**	551	0.67	568	**5.35**	552	0.79	556	0.01	**21.6**
	Luxembourg	480	0.00	471	1.69	482	0.20	474	0.96	477	0.13	477	0.12	9.8
	Mexico	409	0.38	378	**39.74**	409	0.27	423	3.56	426	**7.45**	433	**25.39**	**139.5**
	Netherlands	510	0.12	525	**5.64**	511	0.05	516	0.28	509	0.76	499	**8.83**	**23.9**
	New Zealand	517	0.44	532	3.00	518	0.27	519	0.12	524	0.34	514	1.35	14.7
	Norway	489	0.50	489	0.46	486	0.04	488	0.18	481	0.35	493	2.21	14.9
	Poland	505	0.18	498	3.24	503	0.43	500	1.93	512	0.68	503	0.74	16.2
	Portugal	473	0.01	449	**14.20**	472	0.01	456	**7.89**	491	**10.24**	461	**5.38**	**56.6**
	Slovak Republic	467	0.10	478	2.36	467	0.08	459	2.36	461	1.87	474	0.79	19.6
	Spain	459	0.26	461	0.01	460	0.11	463	0.41	460	0.05	465	0.83	7.4
	Sweden	508	0.00	517	2.36	505	0.35	509	0.03	511	0.25	504	0.66	8.2
	Switzerland	501	0.12	513	**10.95**	503	0.41	508	**4.82**	479	**20.64**	500	0.02	**52.2**
	Turkey	447	0.01	436	3.12	447	0.00	454	1.33	449	0.12	444	0.30	11.8
	United Kingdom	490	0.87	485	3.69	499	0.45	488	2.42	503	2.95	499	0.38	16.0
	United States													
	OECD average	**476**	**0.00**	**477**	**0.03**	**476**	**0.00**	**476**	**0.00**	**476**	**0.01**	**477**	**0.01**	**0.2**
Partners	Argentina	367	0.60	319	**25.30**	370	0.51	355	**6.72**	413	**20.38**	403	**8.39**	**77.6**
	Azerbaijan	353	0.01	367	**11.04**	352	0.04	358	1.19	337	**11.51**	368	**7.35**	**47.7**
	Brazil	396	0.24	362	**30.46**	395	0.22	389	0.77	418	**12.77**	400	1.05	**62.8**
	Bulgaria	404	0.04	393	1.27	397	0.24	397	0.50	400	0.08	409	0.58	7.4
	Chile	442	0.00	414	**13.85**	437	0.77	434	1.57	461	**7.33**	461	**7.60**	**54.5**
	Chinese Taipei	499	0.16	508	**7.79**	491	1.09	502	1.65	492	1.11	493	0.66	19.0
	Colombia	391	0.06	309	**100.96**	377	1.43	388	0.00	422	**25.20**	428	**24.28**	**236.2**
	Croatia	473	0.57	471	1.73	475	0.19	471	2.29	486	**3.87**	473	0.92	12.2
	Estonia	507	1.30	512	3.77	500	0.00	504	0.22	489	**4.91**	501	0.01	**30.3**
	Hong Kong-China	537	0.14	554	**15.88**	531	1.20	552	**13.65**	519	**12.94**	541	1.84	**74.1**
	Indonesia	395	0.13	385	1.06	395	0.18	395	0.24	381	**4.86**	397	0.39	10.7
	Israel	440	0.02	437	0.04	442	0.10	439	0.00	448	1.85	461	**8.39**	**24.1**
	Jordan	402	0.08	388	**5.68**	398	0.24	407	1.51	406	1.23	402	0.11	13.1
	Kyrgyzstan	283	0.12	245	**54.62**	283	0.02	276	1.88	321	**37.45**	285	0.01	**122.2**
	Latvia	478	0.02	474	0.97	480	0.00	475	0.85	475	0.58	475	0.58	9.0
	Liechtenstein	512	0.00	521	0.31	516	0.06	543	2.46	489	1.17	521	0.26	9.8
	Lithuania	465	0.66	466	0.67	468	0.21	461	2.18	467	0.44	474	0.52	19.8
	Macao-China	497	0.73	492	0.01	493	0.04	494	0.11	481	**5.71**	498	1.97	10.9
	Montenegro	393	0.05	396	0.65	383	1.96	402	3.15	378	**7.58**	405	**5.34**	**25.7**
	Qatar	307	0.77	295	**8.77**	321	2.24	312	0.00	326	**9.79**	331	**8.52**	**35.9**
	Romania	387	1.51	380	**7.00**	396	0.02	379	**4.43**	415	**7.08**	411	2.89	**27.9**
	Russian Federation	439	0.09	418	**13.51**	441	0.03	443	0.20	450	2.31	446	1.05	**42.9**
	Serbia	400	0.06	396	0.90	396	0.74	405	0.59	394	1.88	401	0.00	8.7
	Slovenia	498	0.03	517	**15.35**	496	0.10	509	**5.57**	478	**15.51**	504	2.48	**61.7**
	Thailand	419	0.24	415	0.25	418	0.11	411	1.12	407	**5.59**	421	1.00	18.6
	Tunisia	382	0.01	347	**23.32**	376	0.80	376	0.83	394	**6.08**	398	**7.50**	**52.1**
	Uruguay	416	0.04	367	**38.05**	417	0.09	402	2.63	445	**23.34**	455	**35.22**	**158.1**

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



Table 12.16 [Part 1/2]
Variance in science booklet means

		Expected mean	Booklet 1		Boolet 2		Booklet 3		Boolet 4		Boolet 5		Boolet 6		Boolet 7	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	Australia	526	531	1.32	525	0.07	517	**6.85**	525	0.14	530	1.25	529	0.30	516	**7.59**
	Austria	512	512	0.00	510	0.12	510	0.12	501	3.19	507	0.79	520	1.76	501	3.66
	Belgium	514	513	0.08	514	0.00	511	1.02	513	0.21	520	2.27	514	0.01	523	**4.92**
	Canada	534	535	0.04	536	0.12	525	**6.72**	532	0.27	545	**7.60**	534	0.01	532	0.19
	Czech Republic	518	520	0.18	516	0.18	519	0.08	511	2.25	517	0.04	516	0.17	532	**5.96**
	Denmark	496	488	1.60	496	0.00	502	1.22	499	0.36	496	0.00	516	**12.61**	488	1.64
	Finland	564	565	0.10	553	**3.86**	563	0.04	563	0.01	570	1.28	581	**11.15**	571	2.55
	France	495	484	3.83	513	**8.90**	500	1.17	493	0.14	498	0.33	491	0.82	510	**8.64**
	Germany	522	530	2.77	522	0.02	519	0.32	522	0.03	516	0.94	523	0.05	505	**7.32**
	Greece	473	473	0.03	479	0.91	493	**12.87**	479	1.34	471	0.19	469	0.63	496	**14.27**
	Hungary	504	506	0.20	507	0.39	500	0.62	501	0.33	501	0.38	510	1.43	501	0.41
	Iceland	491	490	0.05	479	**4.41**	483	1.99	487	0.51	487	0.50	492	0.04	495	0.50
	Ireland	508	506	0.19	505	0.17	505	0.34	510	0.16	496	**7.69**	508	0.00	474	**32.09**
	Italy	476	472	1.19	490	**15.61**	486	**7.28**	477	0.18	471	2.15	479	0.92	476	0.00
	Japan	531	531	0.00	528	0.21	522	3.50	517	**9.82**	531	0.00	523	2.13	533	0.22
	Korea	522	526	0.74	498	**20.85**	500	**23.17**	519	0.47	521	0.01	523	0.07	523	0.06
	Luxembourg	486	479	2.11	489	0.21	490	0.58	485	0.06	484	0.20	494	1.78	488	0.13
	Mexico	410	399	**7.15**	419	**4.60**	416	1.87	409	0.12	410	0.01	402	3.43	417	2.29
	Netherlands	529	534	1.34	527	0.14	525	1.04	525	0.54	539	**3.96**	532	0.52	529	0.00
	New Zealand	531	518	**4.36**	530	0.00	527	0.36	534	0.43	520	**3.93**	536	0.99	527	0.28
	Norway	487	476	3.82	482	0.66	499	**5.07**	498	3.47	481	0.92	487	0.01	498	3.51
	Poland	498	495	0.24	491	1.77	498	0.00	506	2.42	509	**5.12**	491	1.82	503	1.27
	Portugal	475	481	1.15	481	1.42	472	0.22	476	0.14	475	0.00	459	**7.97**	473	0.07
	Slovak Republic	490	492	0.35	480	3.18	492	0.19	487	0.21	489	0.01	486	0.86	498	1.76
	Spain	488	487	0.04	490	0.07	490	0.06	482	3.03	492	0.72	478	**3.87**	499	**6.11**
	Sweden	503	510	1.22	498	0.69	505	0.03	503	0.01	501	0.23	505	0.08	518	**7.60**
	Switzerland	511	512	0.10	506	1.92	505	1.54	516	1.10	519	2.87	514	0.26	506	1.53
	Turkey	425	427	0.16	424	0.03	430	1.07	434	**4.24**	431	1.89	413	3.45	416	2.36
	United Kingdom	516	514	0.20	515	0.04	530	**15.08**	516	0.00	513	0.40	517	0.04	504	**5.37**
	United States	490	492	0.17	489	0.04	497	1.10	496	1.16	486	0.36	477	3.40	478	3.60
	OECD average	**501**	**500**	**0.02**	**500**	**0.04**	**501**	**0.00**	**501**	**0.00**	**501**	**0.00**	**501**	**0.00**	**501**	**0.00**
Partners	Argentina	392	392	0.00	418	**10.15**	410	**5.71**	406	**4.52**	384	0.75	363	**16.00**	406	1.96
	Azerbaijan	381	395	**10.46**	380	0.11	393	**9.50**	383	0.25	367	**16.15**	405	**28.49**	383	0.21
	Brazil	391	379	**4.57**	410	**15.18**	421	**45.49**	402	**5.42**	396	1.27	377	**8.55**	388	0.18
	Bulgaria	434	433	0.05	435	0.00	441	0.88	440	0.56	428	0.56	429	0.37	450	**3.93**
	Chile	439	442	0.37	455	**9.13**	448	2.49	446	2.49	429	2.56	421	**8.87**	438	0.00
	Chinese Taipei	532	533	0.03	531	0.07	526	1.83	517	**10.04**	528	0.72	519	**8.04**	543	**4.84**
	Colombia	389	388	0.04	409	**10.42**	416	**13.02**	411	**18.29**	394	0.71	363	**26.20**	405	**5.07**
	Croatia	493	496	0.42	498	0.78	495	0.18	485	3.19	484	**5.36**	490	0.43	483	**6.70**
	Estonia	531	533	0.09	531	0.01	537	1.33	527	0.66	525	1.50	550	**10.29**	517	**6.96**
	Hong Kong-China	541	546	1.35	540	0.06	533	**3.96**	535	1.88	539	0.24	538	0.66	544	0.24
	Indonesia	394	395	0.02	398	0.44	414	**11.82**	406	2.37	392	0.09	376	**5.83**	400	0.51
	Israel	453	450	0.13	461	1.94	470	**4.96**	451	0.15	450	0.18	458	0.68	475	**9.06**
	Jordan	423	422	0.03	434	**4.96**	427	1.14	422	0.01	420	0.21	403	**13.35**	415	2.48
	Kyrgyzstan	324	324	0.00	347	**21.36**	343	**16.92**	344	**24.56**	318	1.28	295	**35.63**	334	**6.35**
	Latvia	490	486	0.27	482	1.26	490	0.03	492	0.15	487	0.21	493	0.23	493	0.27
	Liechtenstein	524	511	0.36	530	0.15	511	0.30	538	0.48	544	1.03	488	2.74	512	0.28
	Lithuania	488	494	1.49	487	0.06	489	0.04	492	0.65	481	1.75	492	0.64	495	1.81
	Macao-China	512	508	0.65	500	**4.39**	513	0.08	512	0.05	507	1.11	505	1.68	519	**3.88**
	Montenegro	412	421	2.34	417	1.13	415	0.34	408	0.98	408	0.61	401	**5.41**	413	0.00
	Qatar	350	336	**9.77**	370	**20.51**	354	1.37	357	3.65	352	0.44	333	**12.56**	350	0.00
	Romania	418	422	0.41	434	**8.18**	423	0.99	409	**4.05**	411	0.91	409	2.86	413	0.97
	Russian Federation	479	475	0.67	498	**11.49**	494	**4.99**	486	1.50	475	1.22	466	3.39	487	2.60
	Serbia	435	447	3.77	435	0.01	436	0.00	431	0.80	426	2.80	431	0.92	438	0.37
	Slovenia	522	523	0.04	504	**8.39**	523	0.00	503	**7.83**	523	0.01	524	0.14	529	1.22
	Thailand	421	412	**8.28**	421	0.04	438	**19.19**	432	**7.12**	428	2.78	413	3.34	430	**4.33**
	Tunisia	385	396	**4.28**	409	**18.40**	418	**44.11**	399	**8.60**	377	3.23	366	**13.71**	388	0.18
	Uruguay	429	429	0.00	459	**36.38**	448	**14.95**	425	0.40	422	1.58	413	**6.14**	445	**6.28**

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



Table 12.16 [Part 2/2]
Variance in science booklet means

		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	536	3.68	529	0.35	529	0.30	534	2.59	529	0.45	519	2.53	**27.4**
	Austria	517	0.58	517	0.76	517	0.73	513	0.02	511	0.02	522	2.72	14.4
	Belgium	508	2.32	505	**6.33**	507	3.20	518	0.84	517	0.33	526	**8.07**	**29.6**
	Canada	534	0.00	530	1.26	532	0.25	539	1.04	535	0.03	538	0.94	18.5
	Czech Republic	520	0.08	511	1.41	517	0.05	516	0.08	511	1.72	531	**5.25**	17.4
	Denmark	509	**4.58**	490	0.81	498	0.16	484	**4.55**	483	**4.93**	496	0.00	**32.5**
	Finland	558	1.08	540	**23.41**	561	0.22	573	**4.15**	562	0.11	564	0.00	**48.0**
	France	492	0.35	494	0.02	475	**13.77**	500	0.55	485	2.54	501	1.17	**42.2**
	Germany	529	1.63	521	0.02	533	3.68	517	0.77	520	0.05	520	0.08	17.6
	Greece	445	**16.91**	482	2.52	469	0.79	460	**6.65**	460	**8.62**	478	0.60	**66.3**
	Hungary	504	0.00	505	0.04	513	3.26	499	1.07	509	1.38	495	3.08	12.6
	Iceland	491	0.00	505	**7.29**	490	0.02	493	0.15	491	0.00	498	1.19	16.7
	Ireland	518	2.78	530	**15.83**	522	**5.60**	513	1.01	528	**10.90**	494	**4.80**	**81.5**
	Italy	471	1.83	480	1.70	468	**3.85**	464	**9.18**	471	1.45	476	0.02	**45.3**
	Japan	544	**5.46**	544	**7.65**	549	**11.66**	532	0.03	530	0.01	524	1.60	**42.3**
	Korea	532	3.09	511	3.70	529	1.59	538	**9.85**	544	**19.62**	525	0.46	**83.7**
	Luxembourg	495	3.08	478	2.04	489	0.13	485	0.03	477	3.00	489	0.13	13.5
	Mexico	394	**16.51**	422	**7.82**	392	**18.74**	411	0.08	405	1.38	429	**29.20**	**93.2**
	Netherlands	536	1.37	525	0.75	533	0.58	533	0.65	521	2.77	522	1.92	15.6
	New Zealand	531	0.01	538	1.30	538	1.49	528	0.15	546	**7.67**	523	1.48	**22.5**
	Norway	491	0.39	488	0.02	480	1.06	482	0.63	478	2.44	484	0.29	**22.3**
	Poland	494	0.53	495	0.24	499	0.08	493	1.09	497	0.00	499	0.14	14.7
	Portugal	468	2.00	480	1.07	470	0.69	473	0.15	487	**6.94**	472	0.29	**22.1**
	Slovak Republic	485	0.65	490	0.00	487	0.22	487	0.37	490	0.04	505	**8.69**	16.5
	Spain	490	0.14	482	2.29	485	0.89	490	0.16	488	0.04	499	**4.11**	**21.5**
	Sweden	504	0.01	509	1.39	503	0.01	490	**7.17**	497	1.39	502	0.10	19.9
	Switzerland	524	**7.10**	500	**6.23**	515	0.69	513	0.24	506	1.51	515	0.42	**25.5**
	Turkey	430	0.99	416	2.43	421	0.48	434	3.83	420	1.06	414	3.01	**25.0**
	United Kingdom	510	1.10	519	0.34	524	2.54	508	2.78	516	0.00	506	3.00	**30.9**
	United States	475	**4.37**	505	**7.76**	489	0.03	496	0.78	495	0.75	481	1.90	**25.4**
	OECD average	**501**	**0.00**	**501**	**0.02**	**501**	**0.00**	**501**	**0.00**	**500**	**0.01**	**502**	**0.03**	**0.1**
Partners	Argentina	369	**4.48**	404	1.87	362	**10.26**	381	2.29	378	3.12	413	**7.15**	**68.3**
	Azerbaijan	389	3.62	374	**4.03**	360	**28.00**	389	3.77	373	**7.34**	379	0.27	**112.2**
	Brazil	372	**16.86**	403	**6.11**	369	**25.26**	398	2.96	387	0.66	373	**11.14**	**143.7**
	Bulgaria	426	1.28	446	2.70	421	2.61	434	0.00	421	3.41	441	0.64	17.0
	Chile	423	**6.33**	450	3.65	419	**11.69**	441	0.16	427	**3.99**	458	**9.17**	**60.9**
	Chinese Taipei	541	2.76	544	**7.31**	535	0.23	530	0.16	529	0.76	547	**8.59**	**45.4**
	Colombia	355	**28.94**	396	1.44	343	**38.82**	385	0.69	364	**17.45**	417	**26.10**	**187.2**
	Croatia	499	1.40	501	**4.17**	498	1.62	493	0.00	494	0.03	496	0.44	**24.7**
	Estonia	539	1.88	534	0.22	535	0.40	516	**7.10**	537	1.05	528	0.48	**32.0**
	Hong Kong-China	555	**4.55**	538	0.52	538	0.38	547	1.76	536	1.19	560	**13.20**	**30.0**
	Indonesia	375	**6.03**	399	0.65	377	**8.47**	402	1.71	392	0.18	390	0.58	**38.7**
	Israel	456	0.28	460	0.95	441	2.51	436	**6.56**	431	**11.39**	460	0.70	**39.5**
	Jordan	408	**6.99**	443	**17.89**	415	2.06	420	0.38	432	3.74	424	0.04	**53.3**
	Kyrgyzstan	284	**78.75**	336	**11.20**	288	**46.81**	337	**7.18**	316	2.25	323	0.00	**252.3**
	Latvia	500	3.20	488	0.13	485	0.59	484	1.14	492	0.25	492	0.12	7.9
	Liechtenstein	536	0.50	510	0.64	528	0.06	540	0.66	510	0.45	532	0.26	7.9
	Lithuania	478	3.10	486	0.14	485	0.26	471	**10.80**	486	0.11	508	**12.49**	**33.3**
	Macao-China	517	1.43	508	0.79	513	0.07	507	0.95	505	1.47	528	**11.89**	**28.4**
	Montenegro	409	0.36	417	1.29	402	3.42	420	2.70	412	0.03	412	0.00	18.6
	Qatar	330	**22.28**	368	**16.29**	341	**5.12**	366	**15.79**	349	0.03	334	**9.71**	**117.5**
	Romania	418	0.00	433	**6.81**	408	2.37	424	0.47	413	0.53	421	0.18	**28.7**
	Russian Federation	469	**3.94**	482	0.30	475	0.34	473	0.90	470	3.59	481	0.04	**35.0**
	Serbia	435	0.00	442	2.42	437	0.05	430	1.33	431	0.51	443	2.14	15.1
	Slovenia	510	**4.12**	522	0.00	534	**3.99**	510	**4.37**	525	0.23	541	**15.84**	**46.2**
	Thailand	398	**30.91**	431	**6.94**	411	**7.02**	421	0.00	422	0.03	417	1.37	**91.3**
	Tunisia	372	**8.83**	393	2.84	359	**29.92**	385	0.00	374	**4.80**	376	2.89	**141.8**
	Uruguay	421	1.73	430	0.12	398	**26.81**	422	1.34	400	**22.59**	449	**8.77**	**127.1**

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



Table 12.17 [Part 1/2]
Variance in interest booklet means

		Expected mean	Booklet 1		Boolet 2		Booklet 3		Boolet 4		Boolet 5		Boolet 6		Boolet 7	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	Australia	466	473	**5.09**	471	1.79	463	0.79	463	0.62	462	1.25	461	1.83	454	**8.86**
	Austria	506	511	1.50	498	2.60	502	0.48	506	0.00	506	0.01	516	2.74	509	0.37
	Belgium	501	502	0.07	498	0.51	499	0.21	507	1.89	499	0.24	505	0.77	511	**6.98**
	Canada	470	477	**3.84**	477	3.39	465	1.37	464	1.58	466	0.71	467	0.29	459	**5.38**
	Czech Republic	487	476	**5.40**	495	2.34	500	**6.48**	491	0.43	485	0.19	476	3.02	488	0.03
	Denmark	463	463	0.00	454	3.02	457	1.22	464	0.06	460	0.28	482	**12.25**	452	**4.77**
	Finland	449	459	3.77	452	0.46	450	0.07	448	0.08	449	0.00	433	**6.30**	452	0.55
	France	520	520	0.02	519	0.02	522	0.17	526	1.22	514	1.27	537	**11.76**	514	1.34
	Germany	512	513	0.04	522	**4.00**	512	0.00	515	0.42	512	0.00	515	0.25	503	3.16
	Greece	548	543	1.26	545	0.34	535	**6.99**	553	1.00	556	2.64	554	1.17	559	2.44
	Hungary	522	525	0.47	529	2.26	515	2.24	517	1.46	527	1.18	523	0.10	525	0.23
	Iceland	467	465	0.11	472	0.53	477	2.56	458	1.65	465	0.08	446	**8.87**	481	**3.84**
	Ireland	481	482	0.05	480	0.04	480	0.04	476	1.04	492	3.46	481	0.00	482	0.00
	Italy	529	523	**3.85**	520	**9.65**	522	3.63	534	2.59	528	0.09	534	2.88	532	0.96
	Japan	512	517	1.03	490	**17.46**	513	0.06	528	**8.65**	513	0.09	509	0.23	497	**5.10**
	Korea	486	482	0.49	481	0.65	494	3.15	476	2.36	484	0.08	482	0.58	478	2.70
	Luxembourg	515	520	1.22	519	0.66	503	**6.28**	511	0.61	512	0.28	518	0.33	522	2.20
	Mexico	611	596	**11.28**	599	**7.08**	612	0.25	611	0.02	618	2.62	620	**5.62**	606	1.25
	Netherlands	449	465	**11.97**	439	3.83	431	**13.48**	459	3.61	435	**5.30**	456	1.21	451	0.05
	New Zealand	461	461	0.01	474	**6.18**	459	0.38	466	0.77	471	2.45	456	0.95	452	2.53
	Norway	473	467	1.10	485	3.24	496	**25.27**	477	0.46	470	0.33	470	0.34	470	0.42
	Poland	500	495	2.15	494	2.11	497	0.49	488	**7.33**	503	0.29	494	1.62	508	3.42
	Portugal	571	565	1.62	563	2.59	564	1.97	576	1.10	562	3.09	576	1.44	576	1.45
	Slovak Republic	522	522	0.00	522	0.01	517	1.07	523	0.08	521	0.05	526	0.60	528	1.62
	Spain	534	526	**4.34**	540	2.56	540	1.65	527	2.65	532	0.22	534	0.00	531	0.69
	Sweden	455	466	**4.72**	455	0.00	446	1.00	447	2.36	449	1.15	453	0.20	451	0.33
	Switzerland	504	506	0.33	508	1.39	492	**8.87**	506	0.24	506	0.14	508	0.71	508	1.48
	Turkey	542	535	0.99	526	**4.53**	565	**20.84**	536	0.63	536	0.67	553	**4.41**	559	**8.05**
	United Kingdom	464	471	3.49	468	0.67	461	0.56	458	1.74	465	0.04	462	0.23	443	**18.75**
	United States	480	472	2.03	479	0.01	474	1.59	485	0.97	494	**5.75**	464	**5.40**	489	2.14
	OECD average	**500**	**500**	**0.00**	**499**	**0.01**	**499**	**0.03**	**500**	**0.00**	**500**	**0.00**	**500**	**0.01**	**500**	**0.00**
Partners	Argentina	566	563	0.32	565	0.05	558	2.29	572	1.14	571	0.55	579	**4.17**	576	2.43
	Azerbaijan	611	609	0.27	607	0.73	600	**7.34**	620	2.87	603	3.30	624	**7.67**	613	0.17
	Brazil	592	577	**6.07**	572	**14.13**	583	2.86	604	3.83	590	0.13	601	3.48	603	**4.06**
	Bulgaria	522	520	0.09	526	0.30	517	0.66	519	0.24	521	0.04	528	0.76	542	**6.19**
	Chile	591	579	**4.53**	585	1.00	598	1.54	605	**5.67**	596	0.72	595	0.38	600	1.74
	Chinese Taipei	534	531	0.40	522	**5.37**	533	0.01	541	1.42	527	2.16	530	0.38	524	**3.86**
	Colombia	642	626	3.36	637	0.46	638	0.39	630	1.86	653	2.55	664	**4.95**	633	1.20
	Croatia	536	527	**4.39**	539	0.30	541	1.08	538	0.11	535	0.03	539	0.41	548	**7.50**
	Estonia	503	503	0.02	502	0.00	500	0.38	497	1.43	498	1.17	508	1.31	503	0.00
	Hong Kong-China	535	530	1.16	530	0.38	522	**5.79**	543	2.17	523	**4.43**	530	0.86	535	0.01
	Indonesia	608	607	0.01	591	**7.57**	582	**33.15**	622	**4.96**	606	0.10	619	**7.44**	618	**4.08**
	Israel	510	515	0.79	516	0.87	508	0.14	505	0.40	513	0.15	491	**5.76**	512	0.08
	Jordan	608	603	1.35	606	0.13	606	0.19	602	1.44	611	0.36	604	0.89	626	**9.51**
	Kyrgyzstan	581	584	0.69	572	3.66	562	**11.64**	578	0.45	589	3.09	597	**13.82**	581	0.00
	Latvia	503	494	3.42	499	0.73	498	1.68	499	0.87	500	0.50	501	0.41	510	1.75
	Liechtenstein	505	534	3.03	512	0.12	534	3.38	499	0.08	484	1.10	490	0.45	515	0.24
	Lithuania	544	544	0.00	544	0.01	541	0.30	540	1.02	542	0.22	549	0.91	552	1.78
	Macao-China	524	525	0.03	507	**8.88**	511	**5.11**	540	**5.49**	517	1.08	531	1.16	521	0.23
	Montenegro	561	561	0.00	579	**9.15**	559	0.07	549	**3.89**	571	3.63	573	**5.05**	561	0.00
	Qatar	566	556	**3.92**	561	0.91	565	0.02	560	0.85	576	**5.87**	562	0.62	582	**8.65**
	Romania	591	573	**9.96**	581	3.27	588	0.25	590	0.01	599	1.47	599	0.77	604	**5.36**
	Russian Federation	541	541	0.00	541	0.00	534	**3.89**	545	0.68	542	0.07	531	**4.56**	547	1.53
	Serbia	524	515	1.98	520	0.43	536	**8.14**	529	1.12	522	0.08	529	1.17	528	1.03
	Slovenia	505	508	0.14	501	0.72	518	3.82	497	1.73	500	1.11	507	0.15	514	2.90
	Thailand	641	633	3.59	624	**18.42**	628	**6.16**	642	0.07	648	1.60	657	**8.44**	657	**7.79**
	Tunisia	589	567	**20.24**	596	2.08	590	0.00	579	**4.33**	598	2.23	610	**14.98**	599	3.37
	Uruguay	567	557	3.23	558	2.85	561	1.07	571	0.44	569	0.10	568	0.03	569	0.20

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



Table 12.17 [Part 2/2]
Variance in interest booklet means

		Boolet 8		Boolet 9		Boolet 10		Boolet 11		Boolet 12		Boolet 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	476	**8.72**	469	0.45	460	2.56	472	2.54	464	0.28	464	0.47	**35.3**
	Austria	505	0.08	504	0.20	513	1.71	501	0.71	507	0.07	499	1.55	12.0
	Belgium	503	0.40	500	0.12	493	2.76	497	1.30	492	**5.23**	506	1.19	**21.7**
	Canada	472	0.36	470	0.01	468	0.11	478	3.27	477	3.48	458	**7.91**	**31.7**
	Czech Republic	488	0.07	484	0.27	491	0.42	482	0.67	492	0.85	485	0.10	20.3
	Denmark	468	0.73	465	0.14	478	**8.16**	470	1.46	457	1.24	451	**5.13**	**38.5**
	Finland	437	3.07	454	0.75	450	0.01	439	3.36	454	0.99	448	0.05	19.5
	France	519	0.01	507	**6.37**	527	1.48	526	1.11	519	0.04	509	2.72	**27.5**
	Germany	508	0.61	510	0.22	510	0.16	512	0.00	520	3.02	502	3.63	15.5
	Greece	531	**8.44**	544	0.54	546	0.20	550	0.06	548	0.03	566	**9.81**	**34.9**
	Hungary	525	0.40	527	0.99	537	**8.58**	509	**6.11**	511	**4.15**	516	1.71	**29.9**
	Iceland	464	0.19	476	2.68	444	**8.43**	467	0.00	461	0.88	482	**5.17**	**35.0**
	Ireland	479	0.13	492	**4.58**	478	0.29	479	0.20	472	2.72	482	0.04	12.6
	Italy	535	2.31	526	0.79	534	2.67	527	0.40	534	2.26	530	0.10	**32.2**
	Japan	509	0.20	512	0.01	515	0.25	513	0.03	523	3.71	514	0.10	**36.9**
	Korea	491	0.68	491	1.57	494	1.60	486	0.00	491	1.03	481	0.95	15.8
	Luxembourg	520	1.10	519	0.86	508	1.25	516	0.05	510	0.54	512	0.34	15.7
	Mexico	615	1.08	607	0.64	614	0.50	623	**5.68**	609	0.13	612	0.05	**36.2**
	Netherlands	449	0.01	443	1.70	442	2.44	458	1.84	459	3.08	453	0.40	**48.9**
	New Zealand	453	1.88	454	1.61	461	0.00	461	0.00	470	2.19	457	0.44	19.4
	Norway	469	0.50	474	0.04	465	1.72	473	0.00	464	2.43	457	**7.23**	**43.1**
	Poland	512	**5.88**	497	0.41	514	**7.52**	497	0.43	505	1.11	503	0.37	**33.1**
	Portugal	574	0.53	567	0.57	578	2.48	569	0.08	574	0.39	574	0.47	17.8
	Slovak Republic	528	1.17	518	0.67	522	0.02	520	0.16	517	0.52	519	0.24	6.2
	Spain	532	0.23	545	**6.75**	538	0.58	539	1.35	534	0.02	529	2.03	**23.1**
	Sweden	458	0.32	452	0.55	467	**5.09**	441	**4.90**	452	0.41	467	**4.72**	**25.7**
	Switzerland	508	1.61	499	0.95	500	0.87	500	0.56	496	**4.19**	508	0.87	**22.2**
	Turkey	520	**11.06**	536	0.74	534	1.43	533	2.49	541	0.01	545	0.26	**56.1**
	United Kingdom	475	**5.87**	462	0.20	475	**6.17**	475	**4.18**	464	0.00	447	**11.40**	**53.3**
	United States	470	2.51	471	2.46	473	0.96	498	**7.63**	477	0.24	489	3.50	**35.2**
	OECD average	**500**	**0.00**	**499**	**0.01**	**501**	**0.05**	**500**	**0.01**	**500**	**0.00**	**499**	**0.03**	**0.2**
Partners	Argentina	556	1.59	554	**7.42**	570	0.31	565	0.06	576	2.19	569	0.14	**22.7**
	Azerbaijan	607	0.91	602	**4.14**	621	**4.87**	604	1.86	639	**32.82**	603	2.88	**69.8**
	Brazil	590	0.14	577	**11.86**	601	3.09	605	**7.47**	601	3.22	594	0.25	**60.6**
	Bulgaria	519	0.17	514	1.27	510	3.40	522	0.00	511	2.36	544	**9.72**	**25.2**
	Chile	590	0.06	575	**5.09**	589	0.17	594	0.19	584	1.01	591	0.01	**22.1**
	Chinese Taipei	546	**7.81**	534	0.00	552	**15.05**	529	1.00	535	0.05	528	1.02	**38.5**
	Colombia	661	3.25	622	**9.87**	664	**7.30**	660	**6.90**	635	0.88	640	0.07	**43.0**
	Croatia	526	2.78	538	0.17	540	0.68	526	**4.03**	526	2.63	536	0.00	**24.1**
	Estonia	507	1.25	502	0.04	511	**4.61**	488	**7.97**	500	0.27	510	2.55	21.0
	Hong Kong-China	544	2.08	532	0.60	548	**5.41**	547	2.32	544	1.70	540	0.62	**27.5**
	Indonesia	626	**15.25**	581	**40.08**	603	0.73	622	**11.70**	604	0.58	621	**8.49**	**134.1**
	Israel	520	2.49	508	0.06	493	**4.03**	518	1.77	505	0.77	511	0.03	17.3
	Jordan	618	3.66	597	**5.08**	604	0.59	610	0.26	599	**5.38**	633	**22.98**	**51.8**
	Kyrgyzstan	590	**4.91**	567	**10.99**	595	**7.99**	584	0.31	572	2.55	572	3.01	**63.1**
	Latvia	509	2.31	500	0.43	510	1.71	505	0.10	500	0.63	522	**13.65**	**28.2**
	Liechtenstein	500	0.08	494	0.62	492	0.74	497	0.19	482	1.93	520	0.99	12.9
	Lithuania	548	0.94	533	**4.52**	548	0.40	546	0.10	545	0.00	545	0.01	10.2
	Macao-China	531	2.76	522	0.08	539	**6.99**	515	1.38	525	0.04	524	0.00	**33.2**
	Montenegro	567	0.69	566	0.66	556	0.74	539	**14.51**	552	2.39	557	0.47	**41.3**
	Qatar	568	0.23	549	**11.26**	554	3.81	573	2.19	558	1.98	585	**14.48**	**54.8**
	Romania	596	0.63	583	1.44	605	**5.55**	591	0.00	579	3.64	601	2.29	**34.6**
	Russian Federation	548	2.87	540	0.02	544	0.33	537	0.77	540	0.07	546	1.46	16.3
	Serbia	508	**7.15**	518	1.15	526	0.13	515	2.72	518	0.77	529	1.09	**26.9**
	Slovenia	503	0.14	511	1.26	500	0.63	485	**7.22**	504	0.09	512	1.60	**21.5**
	Thailand	648	2.84	630	**5.99**	656	**12.86**	645	0.70	628	**6.39**	650	2.82	**77.7**
	Tunisia	607	**10.27**	577	**7.67**	586	0.29	597	1.87	578	**4.94**	590	0.01	**72.3**
	Uruguay	566	0.03	565	0.07	568	0.02	578	3.60	572	0.69	571	0.57	12.9



Table 12.18 [Part 1/2]
Variance in support booklet means

		Expected mean	Booklet 1		Boolet 2		Booklet 3		Boolet 4		Boolet 5		Boolet 6		Boolet 7	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	Australia	488	492	1.11	492	1.54	497	**8.11**	500	**13.87**	483	1.37	488	0.07	477	**9.25**
	Austria	516	508	1.55	537	**12.47**	504	3.40	503	3.66	531	**6.39**	505	2.65	498	**7.44**
	Belgium	493	495	0.41	497	1.08	497	1.76	509	**15.20**	491	0.22	484	**5.04**	502	**4.96**
	Canada	500	496	1.12	513	**5.07**	501	0.07	501	0.00	513	**9.64**	486	**7.13**	507	1.76
	Czech Republic	486	490	0.53	491	0.65	493	1.68	480	1.41	484	0.22	461	**16.81**	490	0.50
	Denmark	482	474	2.70	489	2.33	478	0.84	465	**7.86**	485	0.30	472	**4.32**	484	0.14
	Finland	478	469	3.71	480	0.11	484	1.39	476	0.31	497	**11.02**	502	**19.61**	467	**5.67**
	France	507	517	3.10	513	1.74	495	**4.50**	523	**6.71**	507	0.00	492	**6.25**	511	0.50
	Germany	522	523	0.04	534	**4.93**	521	0.01	497	**13.30**	524	0.08	510	3.38	506	**4.96**
	Greece	533	548	**7.34**	519	**6.89**	530	0.34	533	0.01	524	2.64	542	3.37	531	0.13
	Hungary	512	509	0.46	518	1.08	504	2.08	521	2.66	507	0.61	524	**5.89**	511	0.00
	Iceland	492	485	1.09	495	0.18	510	**7.81**	469	**8.51**	484	1.15	483	1.87	492	0.00
	Ireland	484	478	1.95	495	3.22	482	0.27	488	0.43	492	1.92	476	2.38	471	**7.56**
	Italy	511	508	0.76	509	0.59	509	0.40	516	1.53	503	**7.87**	528	**27.40**	525	**16.01**
	Japan	470	469	0.05	467	0.47	484	**8.39**	447	**14.20**	469	0.06	497	**24.04**	448	**12.26**
	Korea	495	500	0.60	471	**17.03**	496	0.02	504	1.92	473	**12.80**	507	3.58	508	**4.38**
	Luxembourg	522	513	2.38	541	**8.39**	525	0.23	486	**24.18**	520	0.20	517	0.76	528	1.04
	Mexico	536	556	**14.71**	491	**72.81**	501	**108.81**	570	**48.31**	528	2.46	550	**12.33**	545	**4.91**
	Netherlands	448	438	**7.75**	449	0.09	456	**6.36**	456	**5.21**	445	0.28	470	**18.04**	432	**13.86**
	New Zealand	470	463	1.83	467	0.48	488	**14.17**	481	**4.24**	476	1.02	490	**12.45**	457	**6.38**
	Norway	485	463	**9.55**	495	1.55	495	3.78	486	0.02	479	0.72	466	**9.38**	503	**7.58**
	Poland	513	511	0.19	489	**22.75**	500	**7.04**	493	**14.79**	508	0.89	520	2.23	537	**25.49**
	Portugal	537	538	0.00	523	**6.27**	522	**9.58**	534	0.31	538	0.01	544	2.06	544	1.93
	Slovak Republic	497	504	1.87	489	2.53	476	**30.78**	492	1.13	500	0.38	501	0.46	495	0.21
	Spain	529	530	0.04	534	1.64	530	0.04	531	0.13	524	1.85	525	1.01	521	3.35
	Sweden	471	468	0.28	460	1.30	477	1.00	462	1.95	465	1.12	454	**10.15**	479	1.64
	Switzerland	511	509	0.07	522	**5.21**	505	1.56	512	0.03	521	**4.58**	499	**4.05**	510	0.02
	Turkey	563	575	1.22	531	**16.07**	564	0.00	612	**39.04**	543	**6.27**	565	0.08	599	**20.84**
	United Kingdom	470	471	0.03	477	2.07	483	**11.51**	469	0.08	475	1.42	473	0.55	459	**4.92**
	United States	491	490	0.04	494	0.31	493	0.20	481	1.74	504	**4.52**	488	0.31	501	3.06
	OECD average	**500**	**500**	**0.02**	**499**	**0.03**	**500**	**0.00**	**500**	**0.01**	**500**	**0.01**	**501**	**0.01**	**501**	**0.03**
Partners	Argentina	507	513	0.68	494	3.51	492	**8.96**	544	**35.13**	502	0.76	503	0.46	504	0.28
	Azerbaijan	539	554	**5.89**	500	**38.60**	511	**23.93**	583	**52.73**	518	**14.84**	577	**29.95**	567	**14.99**
	Brazil	519	525	1.53	507	3.68	503	**11.45**	561	**85.98**	503	**10.37**	529	3.59	522	0.35
	Bulgaria	528	538	2.19	515	2.50	506	**8.17**	529	0.02	514	3.21	547	**8.80**	546	**6.05**
	Chile	565	558	1.20	571	0.85	557	1.92	592	**17.28**	583	**9.06**	581	**5.77**	557	1.80
	Chinese Taipei	544	564	**14.57**	524	**15.69**	525	**20.32**	565	**15.29**	516	**33.77**	525	**12.54**	563	**15.59**
	Colombia	546	562	**7.21**	519	**11.01**	533	3.76	599	**46.20**	539	0.99	537	1.55	543	0.14
	Croatia	514	516	0.19	494	**18.81**	505	**4.12**	542	**31.47**	509	1.28	516	0.35	520	1.83
	Estonia	497	501	0.49	494	0.44	471	**29.60**	476	**12.27**	506	2.52	495	0.22	506	2.16
	Hong Kong-China	529	526	0.42	517	2.72	512	**7.66**	533	0.47	520	2.07	538	2.79	574	**52.72**
	Indonesia	517	527	**6.14**	483	**62.16**	502	**9.29**	546	**16.39**	502	**8.14**	564	**55.98**	538	**14.84**
	Israel	512	516	0.25	511	0.03	515	0.24	509	0.16	514	0.08	526	2.22	538	**17.46**
	Jordan	555	577	**13.45**	522	**27.46**	527	**18.32**	578	**11.83**	540	**5.74**	602	**49.51**	559	0.51
	Kyrgyzstan	500	516	**5.26**	452	**50.02**	472	**27.87**	533	**35.26**	497	0.20	532	**35.14**	540	**26.87**
	Latvia	494	486	1.72	491	0.32	483	**4.79**	495	0.09	489	0.76	489	1.07	508	**10.09**
	Liechtenstein	526	534	0.09	528	0.01	540	0.64	515	0.13	540	0.41	474	**4.57**	548	0.90
	Lithuania	540	557	**7.16**	531	2.31	502	**58.22**	504	**37.13**	528	**4.36**	549	2.46	552	**5.56**
	Macao-China	522	530	3.00	497	**15.82**	496	**27.76**	544	**15.41**	517	0.61	534	**5.15**	537	**10.81**
	Montenegro	529	542	**4.83**	530	0.04	517	3.41	544	**6.44**	524	0.57	562	**31.68**	525	0.35
	Qatar	519	520	0.02	494	**13.26**	530	2.75	557	**23.98**	521	0.04	572	**62.40**	549	**15.05**
	Romania	538	539	0.01	521	**4.94**	529	2.24	557	3.57	521	**6.72**	571	**15.43**	546	1.71
	Russian Federation	506	522	**10.69**	484	**30.87**	480	**38.95**	510	0.68	502	0.78	510	0.44	524	**17.98**
	Serbia	521	525	0.43	516	0.97	514	2.05	514	1.29	515	1.36	531	3.50	535	**7.69**
	Slovenia	503	493	1.48	526	**11.37**	512	1.72	471	**32.42**	515	**5.08**	495	1.73	515	**5.03**
	Thailand	565	581	**8.25**	514	**130.97**	526	**38.98**	614	**62.41**	546	**10.21**	618	**77.17**	578	**3.97**
	Tunisia	534	541	1.07	516	**7.24**	510	**18.82**	552	**7.20**	524	2.74	543	2.27	567	**24.44**
	Uruguay	510	519	2.81	491	**18.11**	493	**12.08**	534	**22.55**	498	**4.97**	513	0.45	520	3.45

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



Table 12.18 [Part 2/2]
Variance in support booklet means

		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	489	0.08	488	0.01	473	**13.15**	495	3.52	479	**4.23**	479	**4.74**	**61.1**
	Austria	528	2.85	519	0.23	523	0.83	531	3.77	527	2.11	495	**7.39**	**54.8**
	Belgium	484	**5.85**	491	0.13	475	**16.80**	489	0.84	492	0.02	497	0.94	**53.3**
	Canada	494	2.68	497	0.69	485	**10.91**	520	**15.65**	499	0.11	497	0.48	**55.3**
	Czech Republic	468	**16.72**	497	**6.48**	488	0.07	483	0.35	502	**7.04**	493	1.54	**54.0**
	Denmark	481	0.08	499	**7.19**	484	0.09	489	1.55	483	0.01	489	1.60	**29.0**
	Finland	490	3.04	462	**8.38**	483	0.98	493	**6.79**	463	**6.89**	466	**8.43**	**76.3**
	France	493	**5.20**	511	0.65	493	**5.79**	512	0.70	499	1.61	521	**5.30**	**42.1**
	Germany	532	2.00	527	0.60	543	**10.73**	525	0.22	534	2.66	505	**7.00**	**49.9**
	Greece	529	0.56	527	0.84	541	1.48	522	3.77	543	3.24	539	1.19	**31.8**
	Hungary	505	1.41	518	1.30	532	**14.71**	496	**5.80**	512	0.01	494	**13.56**	**49.6**
	Iceland	496	0.42	508	**5.23**	480	2.48	488	0.16	491	0.00	505	3.55	**32.5**
	Ireland	486	0.05	495	**5.09**	481	0.37	489	1.27	482	0.13	482	0.26	**24.9**
	Italy	502	**5.49**	514	0.44	505	2.71	492	**25.01**	510	0.20	520	**6.21**	**94.6**
	Japan	492	**14.55**	473	0.32	466	0.41	471	0.08	441	**14.55**	457	3.82	**93.2**
	Korea	490	0.92	481	**6.55**	507	3.69	492	0.36	504	1.77	507	**5.07**	**58.7**
	Luxembourg	531	2.00	528	0.85	530	1.00	526	0.31	513	1.31	522	0.01	**42.7**
	Mexico	522	**9.82**	535	0.01	547	**6.33**	513	**16.42**	559	**36.93**	550	**11.85**	**345.7**
	Netherlands	447	0.02	452	0.85	447	0.05	452	0.69	448	0.00	429	**15.73**	**68.9**
	New Zealand	466	0.54	465	0.87	467	0.38	454	**9.62**	467	0.24	472	0.16	**52.4**
	Norway	494	1.40	488	0.28	482	0.17	499	3.66	466	**7.83**	489	0.30	**46.2**
	Poland	512	0.02	513	0.00	535	**13.10**	482	**40.38**	535	**16.80**	531	**18.99**	**162.7**
	Portugal	547	3.00	541	0.46	549	3.21	518	**15.36**	543	0.88	551	**7.22**	**50.3**
	Slovak Republic	504	2.63	500	0.21	503	1.37	505	3.04	514	**12.10**	490	1.88	**58.6**
	Spain	529	0.00	533	0.59	530	0.03	535	1.55	527	0.18	531	0.06	10.5
	Sweden	485	**4.56**	468	0.32	473	0.08	488	**8.53**	469	0.09	474	0.20	**31.2**
	Switzerland	510	0.02	508	0.24	513	0.22	510	0.00	510	0.02	506	0.99	17.0
	Turkey	543	**7.11**	527	**21.04**	543	**5.51**	558	0.34	568	0.25	590	**12.24**	**130.0**
	United Kingdom	472	0.32	464	1.72	475	1.92	475	1.12	461	**4.95**	453	**15.69**	**46.3**
	United States	485	0.85	482	1.93	490	0.01	496	0.91	478	**4.58**	493	0.21	18.7
	OECD average	**500**	**0.00**	**500**	**0.00**	**501**	**0.03**	**500**	**0.00**	**501**	**0.00**	**501**	**0.01**	**0.2**
Partners	Argentina	478	**10.15**	513	1.23	519	3.00	498	1.81	529	**11.66**	493	**4.56**	**82.2**
	Azerbaijan	506	**33.45**	511	**29.68**	555	**6.23**	507	**16.15**	587	**61.76**	563	**15.10**	**343.3**
	Brazil	504	**6.90**	511	2.05	526	1.83	503	**12.24**	548	**25.59**	501	**13.95**	**179.5**
	Bulgaria	516	2.78	522	0.57	528	0.00	515	2.57	537	1.19	537	1.57	**39.6**
	Chile	564	0.04	559	0.59	564	0.03	560	0.74	535	**12.25**	551	**4.73**	**56.2**
	Chinese Taipei	566	**18.13**	533	**6.49**	548	0.51	551	1.48	555	**4.05**	558	**7.10**	**165.5**
	Colombia	524	**9.79**	543	0.14	542	0.21	520	**12.01**	585	**32.08**	547	0.07	**125.2**
	Croatia	507	1.59	513	0.04	522	2.49	499	**11.50**	522	1.91	522	**4.76**	**80.3**
	Estonia	513	**10.99**	489	2.39	496	0.03	492	0.72	507	3.25	513	**9.88**	**75.0**
	Hong Kong-China	533	0.23	506	**21.87**	526	0.39	519	2.33	505	**12.40**	567	**38.95**	**145.0**
	Indonesia	508	2.22	501	**14.09**	535	**9.87**	500	**14.45**	541	**17.49**	530	**7.51**	**238.6**
	Israel	518	0.69	492	**6.06**	505	0.52	496	**5.65**	483	**16.97**	525	3.60	**53.9**
	Jordan	553	0.05	553	0.09	561	1.05	512	**45.01**	573	**9.10**	560	0.61	**182.7**
	Kyrgyzstan	469	**25.66**	474	**25.40**	528	**15.89**	488	**4.38**	507	1.09	517	**6.67**	**259.7**
	Latvia	491	0.15	493	0.01	499	0.95	489	0.96	498	0.56	506	**4.47**	**25.9**
	Liechtenstein	511	0.41	524	0.01	514	0.36	545	0.92	515	0.15	527	0.00	8.6
	Lithuania	550	2.64	533	1.51	567	**24.51**	542	0.20	569	**31.25**	541	0.05	**177.4**
	Macao-China	520	0.18	499	**15.89**	540	**8.27**	501	**9.41**	513	2.33	543	**16.98**	**131.6**
	Montenegro	545	**5.00**	520	2.84	521	1.95	509	**11.12**	514	**5.16**	524	0.40	**73.8**
	Qatar	493	**12.62**	508	2.85	514	0.45	474	**50.01**	509	1.62	516	0.27	**185.3**
	Romania	529	1.40	531	0.95	564	**8.58**	532	0.36	541	0.14	540	0.06	**46.1**
	Russian Federation	508	0.26	504	0.20	522	**6.70**	503	0.39	526	**11.78**	511	1.01	**120.7**
	Serbia	528	1.34	515	1.20	523	0.17	501	**6.74**	522	0.02	524	0.26	**27.0**
	Slovenia	503	0.00	505	0.06	497	1.33	486	**4.68**	509	0.90	506	0.25	**66.1**
	Thailand	549	**9.39**	556	3.52	594	**17.64**	540	**13.96**	596	**32.74**	578	**3.92**	**413.1**
	Tunisia	513	**6.79**	537	0.18	539	0.53	486	**40.39**	554	**8.86**	558	**11.41**	**131.9**
	Uruguay	515	0.98	504	0.90	509	0.03	496	**6.72**	524	**6.44**	521	**4.43**	**83.9**

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



Booklets with the domain at the end of the booklet (mathematics in booklets 3, 7 and 13 and reading in booklets 2 and 9) have the highest parameters. The booklet effects for reading are more extreme than last cycle, possibly because the items in the major domain (science) include more words than the items of the major domain of last cycle (mathematics).

After scaling the PISA 2006 data for each country separately, the booklet parameters were added to the students' achievement scores for mathematics, reading, science, interest and support and mean performance scores could be compared across countries and across booklets. Tables 12.14 to 12.18 present results of testing the variance in booklet means by country (UH booklet excluded). 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. Significant values are in bold.

Taking the square root of the squared difference between observed and expected mean, divided by the error variance gives a *z*-score and is an indication of the magnitude of the difference between observed booklet mean and expected booklet mean. Significantly easier than expected booklets are bold and italic, significantly harder booklets than expected are bold. Shaded columns are booklets without items in the domain.

There is no significant booklet effect at the OECD level, because the booklet corrections controlled for this effect. Therefore, the booklet effects within countries are relative to the effect at OECD level. A plausible explanation for high chi-squares across domains of most countries is fatigue or speediness (Mexico, Colombia, Kyrgyzstan, Tunisia and Uruguay). In these cases the booklet means deviate most from the expected mean if the items of that domain appear at the end of the booklet. For some other countries, the reason for their relative high chi-squares across domains is less obvious (Azerbaijan, Brazil and Qatar).

The vast majority of booklets means for domains that are not included in the booklet (shaded columns for mathematics and reading) do not significantly differ from the expected booklet means, which is to be expected using the deviation contrast codes for booklets in the conditioning model.

Overview of the PISA cognitive reporting scales

PISA 2006 is the third PISA assessment and as such it is the third occasion on which reading, mathematics and science literacy scores have been reported. A central aim of PISA is to monitor trends over time in indicators based upon reading, mathematics and science literacy. In this section we review the stability of the PISA scales over time, with a view to:

- Setting out the range of scales that have been prepared over the past three PISA assessments;
- Describing their special features and appropriate use; and,
- Asking recommendations regarding future design elements of PISA.

Table 12.19 provides a listing of the 19 distinct cognitive scales that have been produced as part of PISA 2000, 2003 and 2006.[2] For the purpose of this overview, the cognitive scales are classified into three types: PISA overall literacy scales, PISA literacy scales and special purpose scales. PISA overall literacy scales are the key reporting scales that have been established for each domain, when that domain has been the major domain. The PISA literacy scales are sub-components of PISA overall literacy scales that were provided when a domain was the major domain. The special purpose scales are additional scales that can be used as interim and trend scales prior to the establishment of the related PISA overall literacy scales.



Table 12.19
Summary of PISA cognitive reporting scales

Name	Established	2000	2003	2006	Comment
PISA literacy scale					
PISA reading	2000	✓	✓	✓	Trends can be reported between any of the three cycles, by country or by subgroups within countries
PISA mathematics	2003		✓	✓	Trends can be reported between 2003 and 2006, by country or by subgroups within countries
PISA science	2006			✓	Provides the basis for future trend analysis by country or by subgroups within country
PISA literacy scales					
Reading scales					
Retrieving information	2000	✓			
Interpreting texts	2000	✓			
Reflection and evaluation	2000	✓			
Mathematics scales					
Quantity	2003		✓		
Uncertainty	2003		✓		
Space & shape	2003	✓	✓		Established in 2003 and then applied to 2000 with a rescaling (no conditioning). Trends can be reported for countries, but are not optimal for subgroups within countries.
Change & relationships	2003	✓	✓		Established in 2003 and then applied to 2000 with a rescaling (no conditioning). Trends can be reported for countries, but are not optimal for subgroups within countries.
Science scales					
Explaining phenomena scientifically	2006			✓	
Identifying scientific Issues	2006			✓	
Using scientific evidence	2006			✓	
Physical systems	2006			✓	Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal.
Earth & space systems	2006			✓	Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal.
Living systems	2006			✓	Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal.
Special purpose scales					
Interim mathematics	2000	✓			
Interim science	2000	✓	✓		
Science trend 2003-2006	2006		✓	✓	Uses items that were common to PISA 2003 and 2006

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



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; and comments are made about the scale's appropriate use. In the text following, further details are provided on these scales.

PISA overall literacy scales

The primary PISA reporting scales are PISA reading, PISA mathematics and PISA science. These scales were established in the year in which the respective domain was the major domain, 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 scales

Across the three PISA assessments a total of 13 scales have been prepared and reported. In PISA 2000, three reading aspect-based scales were prepared; in PISA 2003, four mathematics content-based scales were prepared; and in 2006 a total of six science scales were prepared.[3]

The scales 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 scales. The one exception to this general practice is mathematics, for which the *space and shape* and *change and relationships* scales were reported for the PISA 2000 data as well as the PISA 2003 data. These scales, 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 scales where prepared using a methodology that permits trend analysis at the national level (or at the level of adjudicated regions), but the scales are not optimal for analysis at the level of student sub-groups.[4]

For science in PISA 2006, two alternative sets of scales were prepared. The first was a set of three process-based scales and the second was a set of three content-based scales. 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 scales 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" scale, with performance on the *using scientific evidence* scale. Furthermore the content-based scales 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 scales are suitable in addition for sub-group analyses.[5]

The metric of all of the PISA scales is set so that scales within a domain can be compared to each other and with the matching overall PISA reporting scale.[6]

Special purpose scales

There are three special purpose scales.

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 alternative 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 and the alternative 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 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.

OBSERVATIONS CONCERNING THE CONSTRUCTION OF THE PISA OVERALL LITERACY SCALES

A number of the PISA scales have been established to permit trend analyses. A review of the various links available and necessary to establish these scales is given below. Table 12.20 illustrates the six linkages of the PISA domains that are examined and discussed below. Links (1) and (2) are for reading 2000 to 2003 and 2003 to 2006 respectively, links (3) and (4) are for mathematics 2000 to 2003 and 2003 to 2006 respectively, links (5) and (6) are for science 2000 to 2003 and 2003 to 2006 respectively.

Table 12.20 also indicates in which data collections the domain was a major domain and on which occasions it was a minor domain. As a consequence one can note that on two occasions the links are major to minor (links (1) and (4)), on two occasions they are minor to minor (links (2) and (5)), and on two occasions they are minor to major (links (3) and (6)).

When a proficiency area is assessed as a major domain there are two key characteristics that distinguish it from a minor domain. First the framework for the area is fully developed and elaborated. Second the framework is comprehensively assessed since more assessment time is allocated to the major domain than is allocated to each of the minor domains.

Table 12.20
Linkage types among PISA domains 2000-2006

	2000		2003		2006
Reading	**Major**	(1) →	Minor	(2) →	Minor
Mathematics	Minor	(3) →	**Major**	(4) →	Minor
Science	Minor	(5) →	Minor	(6) →	**Major**

Framework development

For PISA 2000 a full and comprehensive framework was developed for reading to guide the assessment of reading as a major domain. Less fully articulated frameworks were developed to support the assessment of mathematics and science as minor domains.[7]

For PISA 2003, the mathematics framework was updated and fully developed to support a comprehensive assessment of mathematics. The reading and science frameworks were retained largely as they had been for PISA 2000.[8]



The key changes to the mathematics framework between 2000 and 2003 were:

- Addition of a theoretical underpinning of the mathematics assessment, expanding the rationale for the PISA emphasis on using mathematical knowledge and skills to solve problems encountered in life;
- Restructuring and expansion of domain content: expansion from two broad content areas (overarching ideas) to four; removal of all reference to mathematics curricular strands as a separate content categorisation (instead, definitions of the overarching ideas were expanded to include mention of the kinds of school mathematics topics associated with each);
- A more elaborated rational for the existing balance between realistic mathematicsand more traditional context-free items, in line with the literacy for lifenotion underlying OECD/PISA assessments;
- A redeveloped discussion of the relevant mathematical processes: a clearer and much enhanced link between the process referred to as mathematisation, the underlying mathematical competencies, and the competency clusters; and a better operationalisation of the competency classes through a more detailed description of the underlying proficiency demands they place on students;
- Considerable elaboration through addition of examples, including items from previous test administrations.

Clearly, the framework change involving an effective doubling of the mathematical content base of the study was of such significance that trend measures would be very seriously affected. Hence, only scale links to 2000 were possible, and the new framework provided the first comprehensive basis for the calculation of future trend estimates.

For PISA 2006, science was the major domain so the science framework was updated and fully developed to support a comprehensive assessment of science. The reading framework was retained largely as it had been for PISA 2000, and the mathematics framework as it had been for PISA 2003.[9] The key changes to the science framework between 2003 and 2006 as they relate to comparison in the science scales over time were:

- A clearer separation of knowledge about science as a form of human enquiry from knowledge of science, meaning knowledge of the natural world as articulated in the different scientific disciplines. In particular, PISA 2006 gives greater emphasis to knowledge about science as an aspect of science performance, through the addition of elements that underscore students' knowledge about the characteristic features of science and scientific endeavour; and
- The addition of new components on the relationship between science and technology.

Both of these changes carry the potential to disrupt links with the previous special purpose science scales: the interim science and trend science scales.

Testing time and item characteristics

In each of PISA 2000, 2003 and 2006 a total of 390 unique minutes of testing material was used.[10] The distribution of the testing minutes is given in Table 12.21. When a domain is assessed as a major domain then more minutes are devoted to it than for minor domains. For example 270 minutes were assigned to reading material in PISA 2000 to allow full coverage of the framework. Similarly, PISA 2003 included 210 minutes of mathematics material and PISA 2006 included 210 minutes of science material. When a domain is assessed as a minor domain the assessment is far less comprehensive and does not provide an in-depth assessment of the full framework that is developed when a domain is a major domain.



Table 12.21
Number of unique item minutes for each domain for each PISA assessment

	Reading	Mathematics	Science	Total
2000	270	60	60	390
2003	60	210	60	330[1]
2006	60	120	210	390

1. 60 minutes were devoted to Problem solving.

It is also important to recognise that given the PISA test design (see Chapter 2) the change of major domains over time means that the testing experience for the majority of students will be different in each cycle - it becomes dominated by the new major domain. For example, the design for PISA 2006 used 13 booklets. Eleven of them comprised at least 50% of science material. For four of these the other 50% comprised only mathematics material, four were completed with a mixture of reading and mathematics material, and for one booklet the other 50% comprised only reading material. Two booklets contained only science material.

The links in terms of numbers of items in common for successive pairs of assessments are shown in Table 12.22.

Table 12.22
Number of link items between successive PISA assessments

	Reading	Mathematics	Science
As Major Domain	129	84	108
Links 2000-2003	28	20	25
Link 2003-2006	28	48	22

Characteristics of each of the links

Reading 2000 to 2003

The PISA reading scale was established in 2000 on the basis of a fully developed and articulated framework and a comprehensive assessment of that framework. In PISA 2003 a subset of 28 of the 2000 reading items was selected and used. Equating procedures reported in OECD (2005) were then used to report the PISA 2003 data on the established PISA reading scale.

The trend results for the OECD countries that participated in both PISA 2000 and 2003 showed that of 32 countries, 10 had a significant decline in mean score and 5 had a significant rise in mean score (OECD 2004). This number of significant changes was regarded as somewhat surprising.

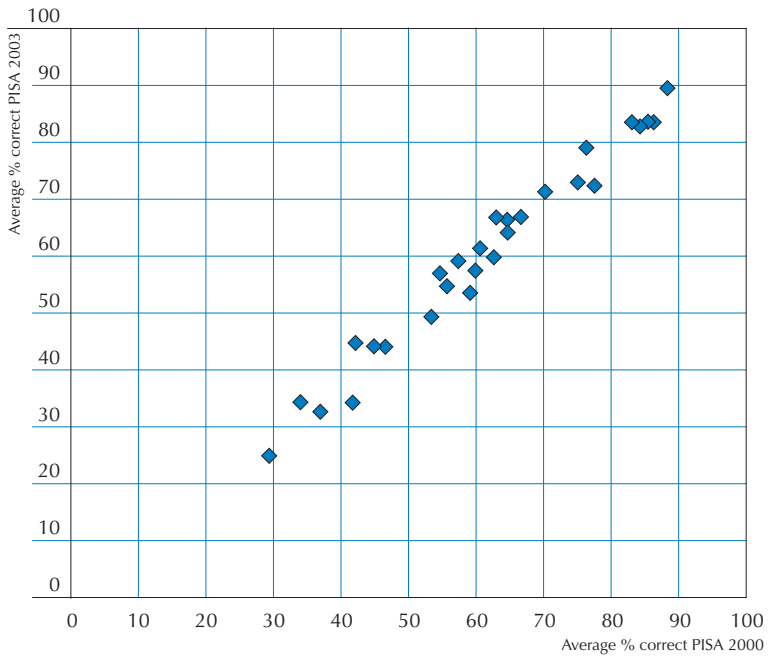
When reviewing the potential causes for this possible instability a number of potentially relevant issues where observed. First, there was a substantial test design change between PISA 2000 and 2003. The PISA 2003 design was fully balanced whereas the PISA 2000 design systematically placed minor domain items and some reading items at the end of the student booklets (see Adams & Wu, 2002). The complexity of the PISA 2000 design is such that the impact of this on the item parameter estimation and hence the equating is unclear. Second, the units that were selected from PISA 2000 for use in PISA 2003 were edited in minor ways. While none of the individual link items was edited, some items in the units were removed. As with the test design change, the impact of this change on the item parameter estimation and hence the equating is unclear. Third, the clusters of items that were used were not pre-existing clusters. In particular, units from PISA 2000 clusters one to seven were selected and reconstituted as two new clusters. Intact clusters of items could not be used from PISA 2000 since none of the individual pre-existing clusters provided an adequate coverage of the framework.



Table 12.23
Per cent correct for reading link items in PISA 2000 and PISA 2003

	% correct	
Item	2000	2003
R055Q01	84.4	82.9
R055Q02	53.4	49.1
R055Q03	62.7	59.8
R055Q05	77.7	72.5
R067Q01	88.5	89.7
R067Q04	54.7	57.0
R067Q05	62.9	67.1
R102Q04A	37.1	32.4
R102Q05	42.2	44.9
R102Q07	86.2	83.5
R104Q01	83.0	83.2
R104Q02	41.6	34.5
R104Q05	29.2	24.9
R111Q01	64.8	66.3
R111Q02B	34.2	34.0
R111Q06B	44.8	44.5
R219Q01	70.2	71.2
R219Q01E	57.4	59.3
R219Q02	76.5	78.8
R220Q01	46.8	44.4
R220Q02B	64.8	64.0
R220Q04	60.8	61.3
R220Q05	85.5	83.2
R220Q06	66.6	67.1
R227Q01	59.0	53.8
R227Q02	59.8	57.7
R227Q03	56.0	54.9
R227Q06	75.2	72.9

Figure 12.6
Scatter plot of per cent correct for reading link items in PISA 2000 and PISA 2003



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



The percentage correct on reading items that link PISA 2000 and PISA 2003 are given in Table 12.23, with the corresponding scatterplot in Figure 12.6. To compute the percentage correct, all students were included from countries that were included in trend analysis between PISA 2000 and PISA 2003. For this analysis 25 OECD countries were included. Excluded were the United Kingdom, the Netherlands, Luxembourg, the Slovak Republic and Turkey.

The mean of the differences (PISA 2000 minus PISA 2003) is 1.11, and the standard deviation of the differences is 2.82.

Reading 2003 to 2006

To link the PISA 2006 data to the PISA reading scale the same items (units and clusters) as were used in PISA 2003 were again used. The trend results for the OECD countries that participated in both PISA 2003 and 2006 showed that of the 38 countries which could be compared, five had a significant decline in mean score and two had a significant rise in mean score (OECD 2007). The number of significant changes is less than reported for the 2000-2003 link.

A number of reasons might be conjectured as possible explanations of this lack of consistency. First, presenting a large number of reading items with a small number of mathematics and science items interspersed, provides for a very different test-taking experience for students compared to a test with a majority of mathematics items, and a few reading, general problem solving and science items interspersed.

Table 12.24
Per cent correct for reading link items in PISA 2003 and PISA 2006

	% correct	
Item	2003	2006
R055Q01	81.4	80.9
R055Q02	47.9	46.8
R055Q03	58.2	57.2
R055Q05	72.6	71.0
R067Q01	89.5	88.2
R067Q04	56.1	55.6
R067Q05	66.4	65.9
R102Q04A	32.4	32.2
R102Q05	43.1	42.8
R102Q07	81.8	82.9
R104Q01	83.0	80.3
R104Q02	34.3	32.9
R104Q05	25.3	22.8
R111Q01	64.9	63.4
R111Q02B	32.9	33.4
R111Q06B	43.3	40.9
R219Q01	69.6	68.4
R219Q01E	57.5	57.4
R219Q02	78.1	78.8
R220Q01	43.2	42.5
R220Q02B	63.5	61.2
R220Q04	62.1	59.2
R220Q05	83.2	81.0
R220Q06	67.1	66.4
R227Q01	53.7	52.3
R227Q02	57.9	55.0
R227Q03	54.4	53.3
R227Q06	71.3	69.3

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

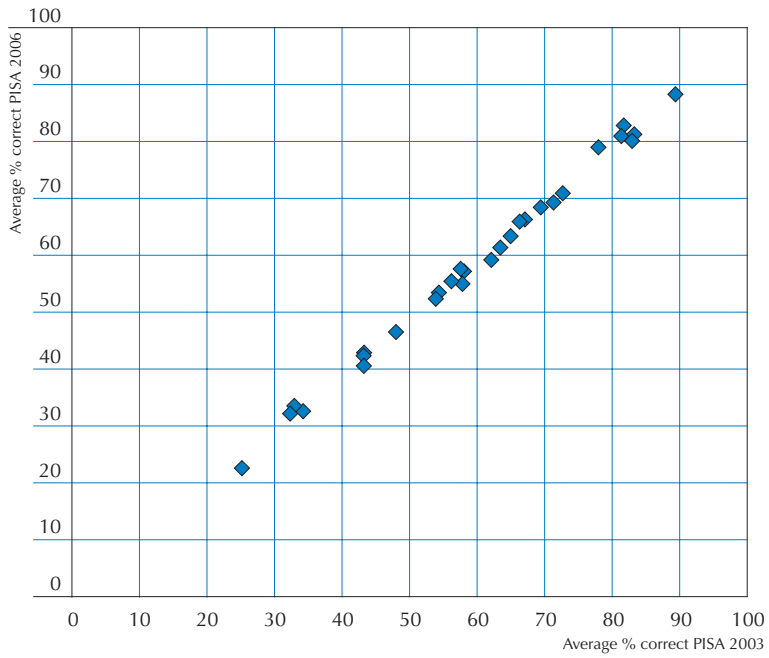


This may have impacted on the trend estimates. Second, the mix of reading items by aspect type was somewhat different between the two test administrations. In 2003 there was a larger proportion of score points in the reflection and evaluation aspect than had been the case for 2000.

The percentage correct on reading items that link PISA 2003 and PISA 2006 are given in Table 12.24, with the corresponding scatterplot in Figure 12.7. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 28 OECD countries were included. Excluded were the United Kingdom and the United States.

Figure 12.7

Scatter plot of per cent correct for reading link items in PISA 2003 and PISA 2006



The mean of the differences (PISA 2003 minus PISA 2006) is 1.17, and the standard deviation of the differences is 1.07. The standard deviation of this difference is much less than that for 2003 to 2006 and most likely due to the use of identical items in identical clusters for the two assessments.

Mathematics 2000 to 2003

The mathematics framework that was prepared for PISA 2000 was preliminary and the assessment was restricted to two of the so-called big ideas – *space and shape,* and *change and relationships.* For the PISA 2003 assessment, when mathematics was a major domain, the framework was fully develop and the assessment was broadened to cover the four overarching ideas – *quantity, uncertainty, space and shape*, and *change and relationships*.

Given that the mathematics framework was fully developed for PISA 2003, the PISA mathematics scale was developed at that point. As PISA 2000 had covered two of the scales, two scale trend scales were developed that permit comparison of performance between 2000 and 2003 for *space and shape*, and *change and relationships*.

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



Mathematics 2003 to 2006

A selection of 48 mathematics items was selected from PISA 2003 and used again in PISA 2006.[11] Hence the change from 2003 to 2006 involved reducing the number of items by almost half, and as was the case when reading changed from major to minor domain, it was not possible to make such a reduction whilst retaining intact clusters. Four new clusters were formed for PISA 2006 from the units retained from PISA 2003. The trend results for the OECD countries that participated in both PISA 2003 and 2006 showed that of the 39 countries which could be compared four had a significant decline in mean score and four had a significant rise in mean score (OECD 2007). The magnitude and number of these changes is consistent with the figures for reading 2003 to 2006 and with figures observed in TIMSS.

Table 12.25
Per cent correct for mathematics link items in PISA 2003 and PISA 2006

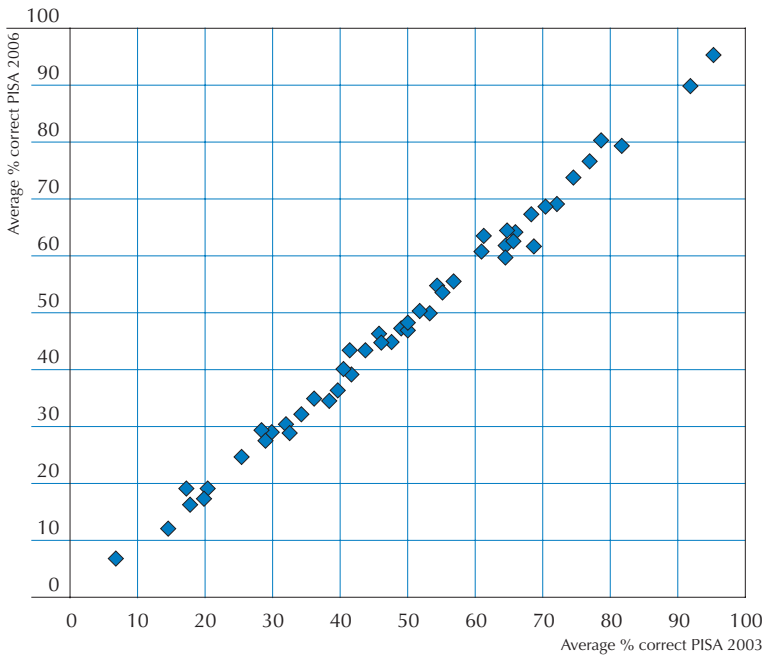
	% correct	
Item	2003	2006
M033Q01	77.0	76.8
M034Q01	43.6	43.5
M155Q01	64.9	64.6
M155Q02	61.0	60.8
M155Q03	17.0	19.1
M155Q04	56.7	55.7
M192Q01	40.7	40.3
M273Q01	55.1	53.7
M302Q01	95.3	95.4
M302Q02	78.6	80.4
M302Q03	29.9	28.9
M305Q01	64.5	61.7
M406Q01	29.1	27.7
M406Q02	19.7	17.2
M408Q01	41.5	43.4
M411Q01	51.8	50.5
M411Q02	46.3	44.8
M420Q01	49.9	48.2
M421Q01	65.8	62.8
M421Q02	17.8	16.3
M421Q03	38.5	34.4
M423Q01	81.5	79.6
M442Q02	41.8	39.3
M446Q01	68.3	67.1
M446Q02	6.9	7.0
M447Q01	70.5	68.6
M462Q01	14.5	12.1
M464Q01	25.4	24.9
M474Q01	74.6	73.7
M496Q01	53.3	50.1
M496Q02	66.0	64.1
M559Q01	61.3	63.5
M564Q01	49.9	47.1
M564Q02	46.0	46.3
M571Q01	49.0	47.3
M598Q01	64.4	59.9
M603Q01	47.7	45.0
M603Q02	36.2	35.1
M710Q01	34.3	32.5
M800Q01	91.9	89.5
M803Q01	28.3	29.7
M810Q01	68.6	61.7
M810Q02	72.3	69.1
M810Q03	20.4	19.2
M828Q01	39.8	36.5
M828Q02	54.5	54.7
M828Q03	32.5	29.1
M833Q01	31.8	30.2



It is interesting to contrast these results with those observed for reading. At the item level the consistency seems somewhat less for mathematics than for reading, whereas at the scale level the consistency is comparable. It is our conjecture that the item-level inconsistency is caused because by the change from mathematics as a major domain to mathematics as a minor domain. Two specific aspects of the change are likely to have contributed to the observed degree of consistency. One is the fact that it was necessary to select a subset of items and form new trend clusters. The rearrangement of items into new clusters appears to have a small impact on relative item difficulty. The second is the fact that the items were presented to students in a different context from previously; specifically that the items were no longer from the dominant domain, rather they represented a smaller set of items presented amongst a much larger number of science items.

The percentage correct on mathematics items that link PISA 2003 and PISA 2006 are given in Table 12.25, with the corresponding scatterplot in Figure 12.8. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 29 OECD countries were included. The United Kingdom was excluded because it was excluded from PISA 2003.

Figure 12.8
Scatter plot of per cent correct for mathematics link items in PISA 2003 and PISA 2006



The mean of the differences (PISA 2003 minus PISA 2006) is 1.40, and the standard deviation of the differences is 1.77. This standard deviation is less than that for reading between 2000 and 2003 but greater than that for reading between 2003 and 2006. This is consistent with the fact that 2003 and 2006 designs where both balanced but, unlike the reading items, the mathematics link items between 2003 and 2006 where not presented in the same clusters.



Science 2000 to 2003

Science was a minor domain in both PISA 2000 and 2003. As such the assessment on both of these occasions was less comprehensive than it was in 2006, when a more fully articulated framework and more testing time was available. There were 25 items that were common to both PISA 2000 and 2003. The trend results for the OECD countries that participated in both PISA 2000 and 2003 showed that of 32 countries, 5 had a significant decline in mean score and 13 a significant rise in mean score (OECD 2004). This number of significant changes was regarded as somewhat surprising.

The number of inconsistencies between 2000 and 2003 was greater than expected at both the item-level and at the scale level. When reviewing the potential causes for this possible instability a number of potentially relevant issues were observed. First, as mentioned above for reading, there was a substantial test design change between PISA 2000 and 2003. The complexity of the PISA 2000 design is such that impact of this on the item parameter estimation and hence the equating is unclear. Second, the units that were selected from PISA 2000 for use in PISA 2003 were edited in minor ways. As with reading, while none of the link items was edited some items in the units were removed. And as with the test design change, the impact of this on the item parameter estimation and hence the equating is unclear. Third the clusters of items that were used were not pre-existing clusters. The material retained from the two PISA 2000 clusters was supplemented with a small number of new units, and reconstituted as two new clusters. Fourth, there were just 25 link items between these two assessments, and unlike mathematics these items were spread across all aspects of the framework. This number was less than desirable and was a result of choices made concerning the release of items following the 2000 assessment to illustrate the nature of the PISA assessment to the public.

Table 12.26
Per cent correct for science link items in PISA 2000 and PISA 2003

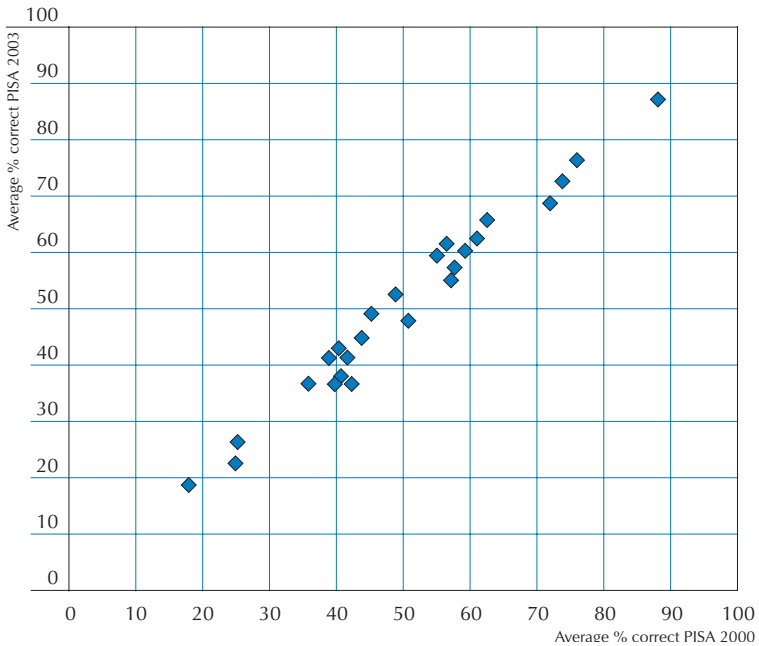
Item	% correct	
	2000	2003
S114Q03	57.3	55.0
S114Q04	39.8	36.8
S114Q05	24.9	22.7
S128Q01	62.6	65.7
S128Q02	45.2	49.0
S128Q03	61.2	62.5
S129Q01	38.8	41.6
S129Q02	17.9	19.0
S131Q02	50.9	47.9
S131Q04	25.2	26.5
S133Q01	56.7	61.6
S133Q03	42.3	36.6
S133Q04	43.8	44.7
S213Q01	40.3	43.2
S213Q02	76.1	76.6
S252Q01	48.8	52.8
S252Q02	72.2	68.6
S252Q03	55.0	59.2
S256Q01	88.3	87.3
S268Q01	73.7	72.4
S268Q02	40.8	38.1
S268Q06	57.9	57.4
S269Q01	59.2	60.2
S269Q03	41.8	41.6
S269Q04	35.9	36.5

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



Figure 12.9

Scatter plot of per cent correct for science link items in PISA 2000 and PISA 2003



The percentage correct on science items that link PISA 2000 and PISA 2003 are given in Table 12.26, with the corresponding scatterplot in Figure 12.9. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 25 OECD countries were included. The United Kingdom, the Netherlands, Luxembourg, the Slovak Republic and Turkey where excluded because they did not participate in either PISA 2000 or PISA 2003 or where excluded for quality assurance reasons from one of PISA 2000 or PISA 2003.

The mean of the differences (PISA 2000 minus PISA 2003) is –0.28, and the standard deviation of the differences is 2.79. This standard deviation is consistent with that observed for reading between 2000 and 2003.

Science 2003 to 2006

In PISA 2006, science was the major domain and as such it was comprehensively assessed on the basis of a newly developed and elaborated framework. As noted above there were quite substantial changes between the preliminary framework that had underpinned PISA 2000 and PISA 2003 assessments and the more fully developed framework used for PISA 2006. Note that in addition to the framework changes mentioned above there was an important change in the way science was assessed in PISA 2006, as compared with PISA 2003 and PISA 2000. First, to more clearly distinguish scientific literacy from reading literacy the PISA 2006 science test items required, on average, less reading than did the science items used in earlier PISA surveys. Second, as with each domain as it goes from a minor to a major domain the item pool (and therefore the testing experience for the majority of students) becomes dominated by the new major domain. For example, there were 108 science items used in PISA 2006, compared with 35 in PISA 2003; of these, just 22 items were common to PISA 2006 and PISA 2003 and 14 were common to PISA 2006 and PISA 2000.

So, as the first major assessment of science, the PISA 2006 assessment was used to establish the basis for the PISA science scale.

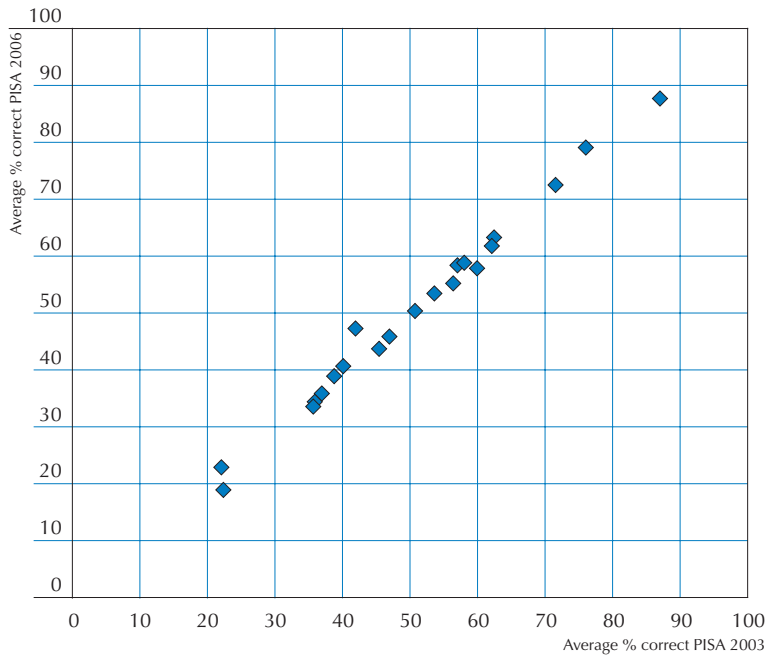


The percentage correct on science items that link PISA 2003 and PISA 2006 are given in Table 12.27, with the corresponding scatterplot and Figure 12.10. To compute the percentage correct, all students were included from countries that were included in these trend analyses. For percentage correct, 29 OECD countries were included. The United Kingdom was excluded because it was excluded from the PISA 2003 database.

Table 12.27
Per cent correct for science link items in PISA 2003 and PISA 2006

Item	% correct	
	2003	2006
S114Q03	53.6	53.6
S114Q04	35.9	34.4
S114Q05	22.4	18.8
S131Q02	46.9	46.2
S213Q01	41.9	47.4
S213Q02	76.2	79.2
S256Q01	87.0	87.5
S268Q01	71.7	72.5
S268Q02	36.9	36.1
S268Q06	56.6	55.4
S269Q01	60.0	57.9
S269Q03	40.1	40.7
S269Q04	35.6	33.8
S304Q01	45.5	43.8
S304Q02	62.0	62.1
S304Q03a	38.7	39.1
S304Q03b	50.7	50.6
S326Q01	58.2	58.7
S326Q02	62.6	63.4
S326Q03	57.2	58.3
S326Q04	22.2	22.8

Figure 12.10
Scatter plot of per cent correct for science link items in PISA 2003 and PISA 2006



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



The mean of the differences (PISA 2000 minus PISA 2003) is –0.01, and the standard deviation of the differences is 1.89. This standard deviation is less than for science 2000-2003 but greater than that for reading 2003-2006. As with the previous observations regarding the standard deviations of the differences this is consistent with PISA test design changes.

For the purposes of trend analysis an additional trend scale has been established that is based upon those items that were common to both PISA 2003 and 2006. Details on the construction of this trend scale are given below and international results are provided in the initial report (OECD, 2007; p.369-370).

On the science trends scale that was produced from these 39 countries that participated in both PISA 2003 and PISA 2006, one had a significant decline in mean score and five had a significant rise in mean score (OECD 2007).

TRANSFORMING THE PLAUSIBLE VALUES TO PISA SCALES

Reading

The reading plausible values were equated to the PISA 2000 scale. Since the same items were used in PISA 2003 as in PISA 2006, and in each case the mean of the item parameter estimates is set at zero, the transformation was exactly the same as in PISA 2003.

For female students:
PISA 2000 scale score = ((0.8739 * Logit + 0.0970 – 0.5076) / 1.1002) * 100 + 500

For male students:
PISA 2000 scale score =((0.8823 * Logit + 0.0204 – 0.5076) / 1.1002) * 100 + 500

For students with missing gender code:
PISA 2000 scale score =((0.8830 * Logit + 0.0552 – 0.5076) / 1.1002) * 100 + 500

For details about equating procedures in 2003, the reader is referred to the *PISA 2003 Technical Report* (OECD, 2005).

Mathematics

For mathematics, the PISA 2006 plausible values were equated to the PISA 2003 scale. A shift of 0.0405 of a logit was required to align the 2003 and 2006 scales. After applying this hift, the same standardisation was used as in PISA 2003 (where –0.1344 is the OECD mean and 1.2838 the OECD standard deviation).

PISA 2003 scale score = (((Logit–0.0405) + 0.1344) / 1.2838) * 100 + 500

Science

A new scale for science was established in PISA 2006. Therefore the only transformation to the plausible values was a standardisation to an OECD mean of 500 and OECD standard deviation of 100 (using an equally weighted, pooled database).

PISA 2006 scale score = ((Logit – 0.1797) / 1.0724) * 100 + 500

The same transformation parameters were used for the scales of science.

An additional set of plausible values was drawn for science link items only (in both 2003 and 2006) to provide estimates of trends in science. To equate the PISA 2006 abilities to the PISA 2003 scale, the following transformations was applied. After adding a shift that reflects the difference in mean item difficulty of the link



items in PISA 2003 and PISA 2006 (–0.1709), the same transformation was applied as in 2003. When the country means and the OECD average were computed, an upward trend was observed in most country means and the OECD average. To compensate for this shift, 13.0 PISA points were subtracted from the PISA 2006 country means to make the OECD average equal in both cycles (excluding the United Kingdom).

PISA 2003 scale score = (((1.0063 * (Logit–0.1709)–0.0155)+0.0933)/1.1085) * 100+500–13.0.

Attitudinal scales

The interest and support attitudinal scales were established in PISA 2006 as well, so the same methodology as for science was applied.

For interest in science:
PISA 2006 scale score = ((Logit – 0.1785) / 1.1190) * 100 + 500

For support of scientic enquiry
PISA 2006 scale score = ((Logit – 1.2694) / 0.8706) * 100 + 500

LINK ERROR

Link errors estimated using the methodology discussed in Chapter 9 for the following five links; PISA mathematics scale 2003 to 2006, PISA reading scale 2000 to 2003, PISA reading scale 2000 to 2006, PISA reading scale 2003 to 2006, and science trend scale 2003 to 2006, are given in Table 12.28. Note that the value of 4.474 given for the PISA reading scale 2000 to 2003 link is a little larger than the value of 3.744, as reported in OECD (2005). Similarly for the Interim science scale the new estimate of 3.112 is a little larger than the previously reported value of 2.959. The differences in these values is due to the improved link error estimation method used for PISA 2006.

Table 12.28
Link error estimates

	Link Error on PISA Scale
Mathematics scale 2003 to 2006	1.382
Reading scale 2000 to 2003	5.307
Reading scale 2000 to 2006	4.976
Reading scale 2003 to 2006	4.474
Interim Science scale 2000 to 2003	3.112
Interim Science trend scale 2003 to 2006	4.963



Notes

1. Note that the USA was not included in the correlations with reading.

2. 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.

3. For a description of the content of the scales see the PISA framework publication (OECD, 2006, *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*).

4. This is because conditioning variables were not used in the construction of the scales for the PISA 2000 data.

5. This is because gender was the only conditioning variable used in the construction of the content-bases scales.

6. Note, of course, that as mentioned above comparison across alternative scalings of the same domain are not appropriate.

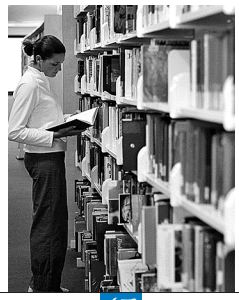
7. The PISA 2000 frameworks were published as OECD (1999) *Measuring Student Knowledge and Skills: A new Framework for Assessment.*

8. The PISA 2003 frameworks were published as OECD (2003) *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills.*

9. The PISA 2006 frameworks were published as OECD (2006) *Assessment Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006.*

10. In 2003 the total testing time was also 390 minutes, but 60 minutes of that testing time was allocated to an assessment of Problem Solving skills.

11. Representing 120 minutes of testing time.



13

Coding and Marker Reliability Studies

Homogeneity analyses 251
Multiple marking study outcomes (variance components) 254
- Generalisability coefficients 254

International coding review 261
- Background to changed procedures for PISA 2006 261
- ICR procedures 261
- Outcomes 264
- Cautions 270

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

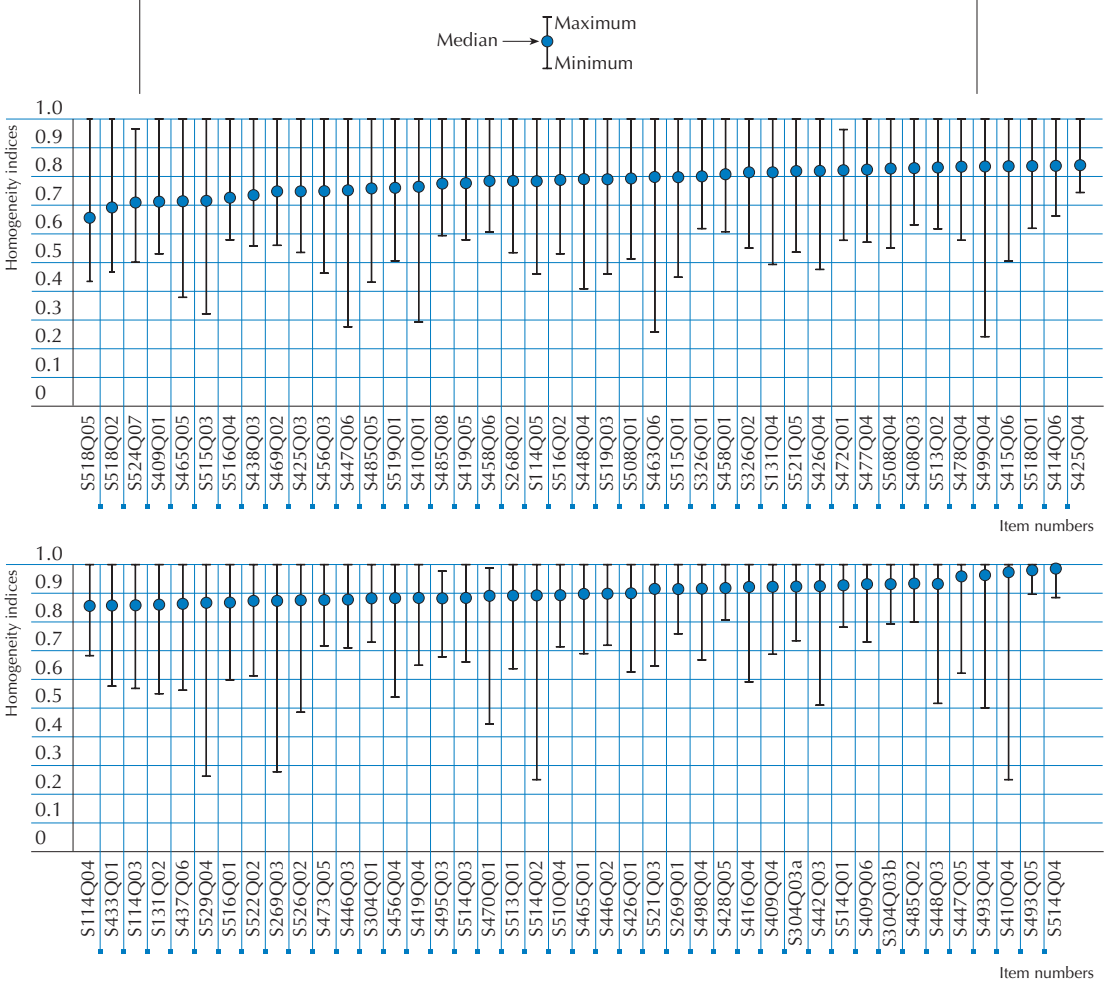


As explained in the first section of this report, on test design (see Chapter 2), a substantial proportion of the PISA 2006 items were open ended and required coding by trained personnel. It was important therefore that PISA implemented procedures that maximised the validity and consistency (both within and between countries) of this coding. Each country coded items on the basis of coding guides prepared by the consortium (see Chapter 2) using the design described in Chapter 6. Training sessions to train countries in the use of the coding guides were held prior to both the field trial and the main study.

This chapter describes the outcomes of three aspects of the coding and marking reliability studies undertaken in conjunction with the field trial and the main study. These are the homogeneity analyses undertaken with the field trial data to assist the test developers in constructing valid, reliable scoring rubrics; the variance component analyses undertaken with the main study data to examine within-country coder reliability; and an international coder review undertaken to examine the between-country consistency in applying the coding guides.

The methods used to compute the homogeneity indices and the variance components for PISA 2006 where the same as the methods used in PISA 2000 and PISA 2003. The methods for both homogeneity and variance components are fully discussed in Verhelst (2002).

Figure 13.1
Variability of the homogeneity indices for science items in field trial



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



HOMOGENEITY ANALYSES

Both in the field trial and the main study homogeneity analyses are used to estimate the level of agreement between coders of constructed-response items. In the field trial the primary purpose of the homogeneity analysis is to obtain data to inform the selection of items for the main study. In the field trial, many more items were tried than were used in the main study and one important purpose of the field trial was to select a subset of science items to be used in the main study. One obvious concern was to ensure that coders agreed to a reasonable degree in their categorisation of the answers.

For investigating the inter-coder agreement, the collected data were used to compute a homogeneity index by item and country. This coefficient theoretically can range from zero to one. A coefficient of one shows perfect agreement between coders. Figure 13.1 shows the distribution of the homogeneity indices for all science items in the field trial and for the selected science items for the main study.

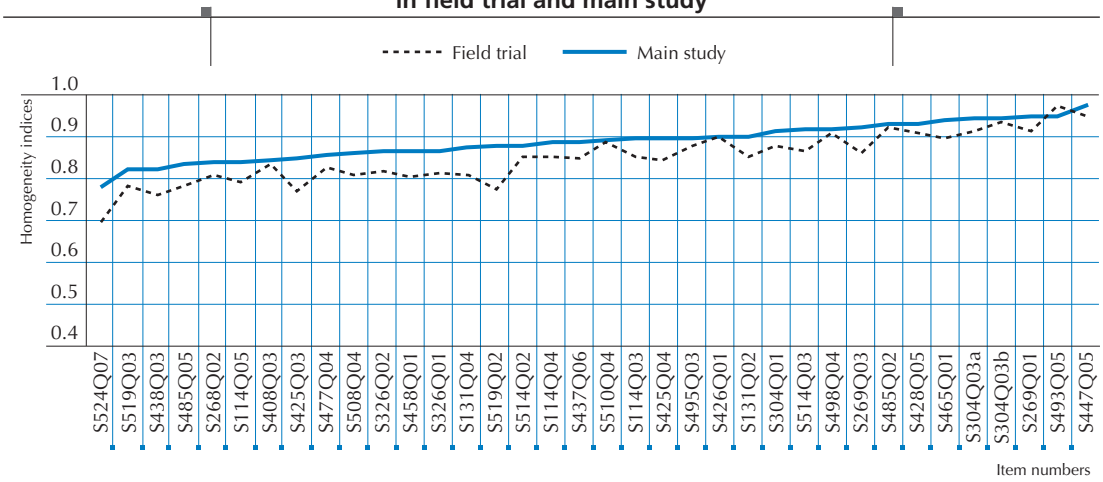
If an item had a weak homogeneity index in the field trial, this was a signal to the Science Expert Group and to the test developers either that the item should not to be retained for the main study or that the coding guide required clarification.

Figure 13.2 shows the average of the homogeneity indices per science item for the items included in the main study. In general the chart shows a marked improvement in the level of agreement between coders in the main study compared to the field trial. Changes to coding schemes contributed to this improvement in a number of cases – for example: in *S425Q03*, double-digit coding was replaced by single-digit coding; in *S465Q01*, partial credit was eliminated; and, in *S519Q01*, partial credit was introduced. However, for most items there was no change to the coding scheme between the field trial and the main study. In these cases, much of the improvement can be attributed to improvements to the coding guides – for example, in *S485Q01*, the level descriptors were refined; examples were added for the descriptors in *S447Q05*; and, in *S514Q03*, the descriptors were revised and additional examples were included. The addition of more workshop examples, the expanded coder query database, and the extra experience gained by coders in the field trial also would have contributed significantly to the general tendency for improvement. The small decrease in the homogeneity index for *S493Q05* can be attributed to the change from partial credit to double-digit coding for the main study.

Figure 13.3, Figure 13.4, and Figure 13.5 show the distribution of the national homogeneity indices per item in the main study.

Figure 13.2

Average of the homogeneity indices for science items in field trial and main study



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



Figure 13.3
Variability of the homogeneity indices for each science item in the main study

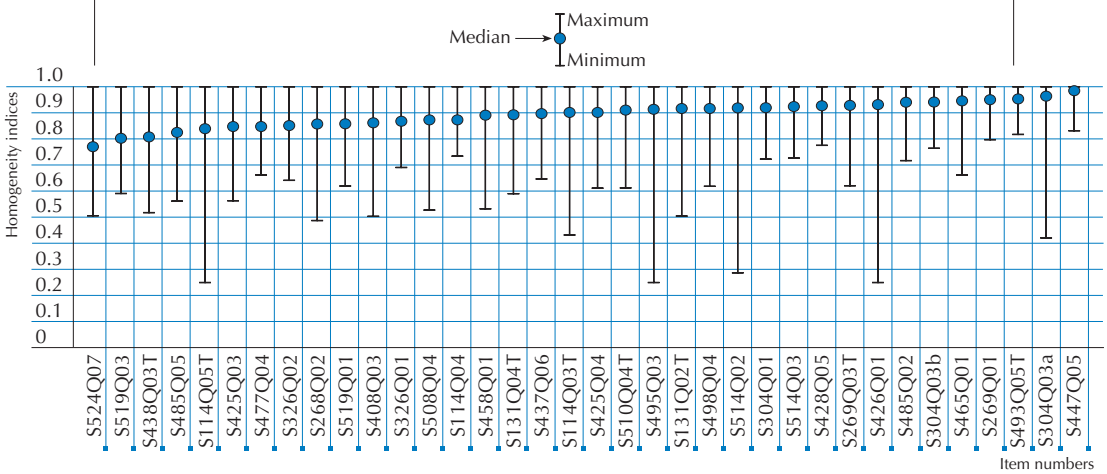


Figure 13.4
Variability of the homogeneity indices for each reading item in the main study

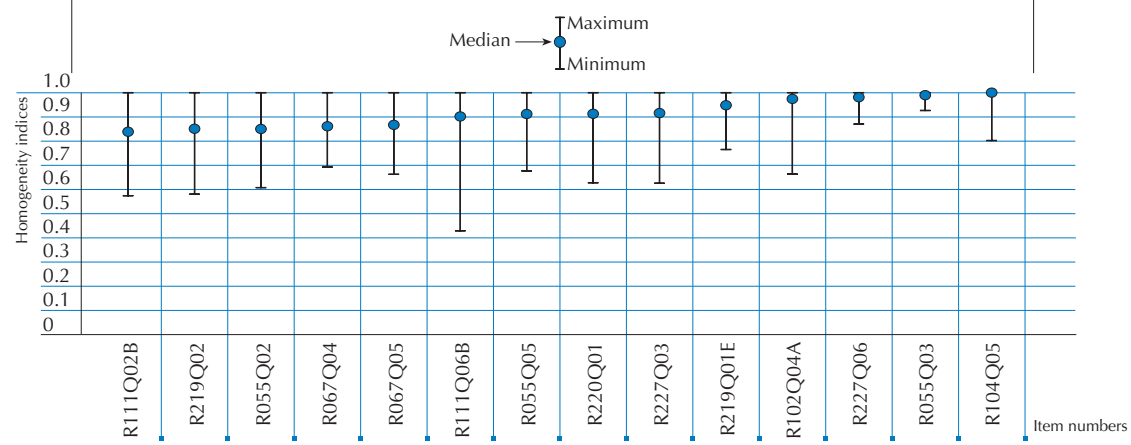


Figure 13.5
Variability of the homogeneity indices for each mathematics item in the main study

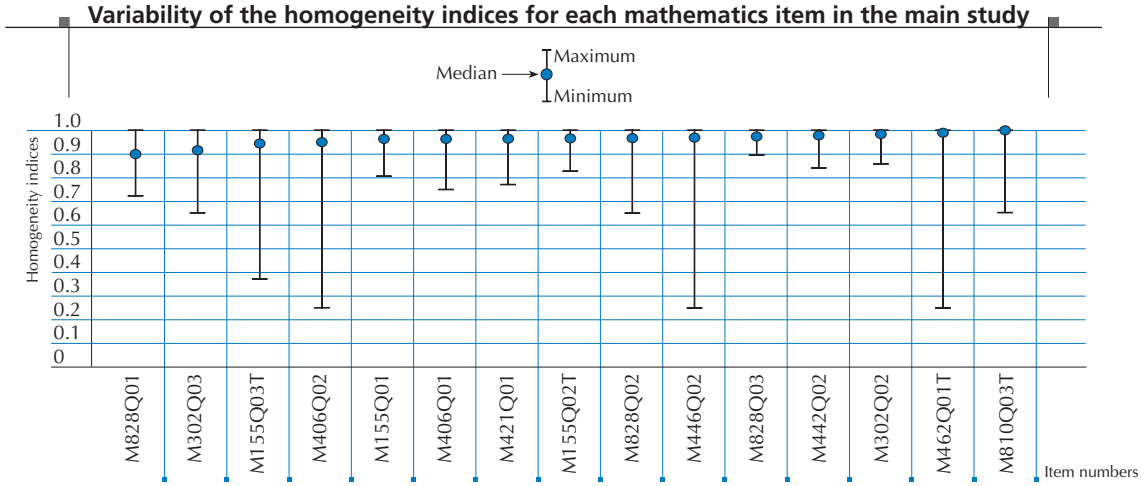
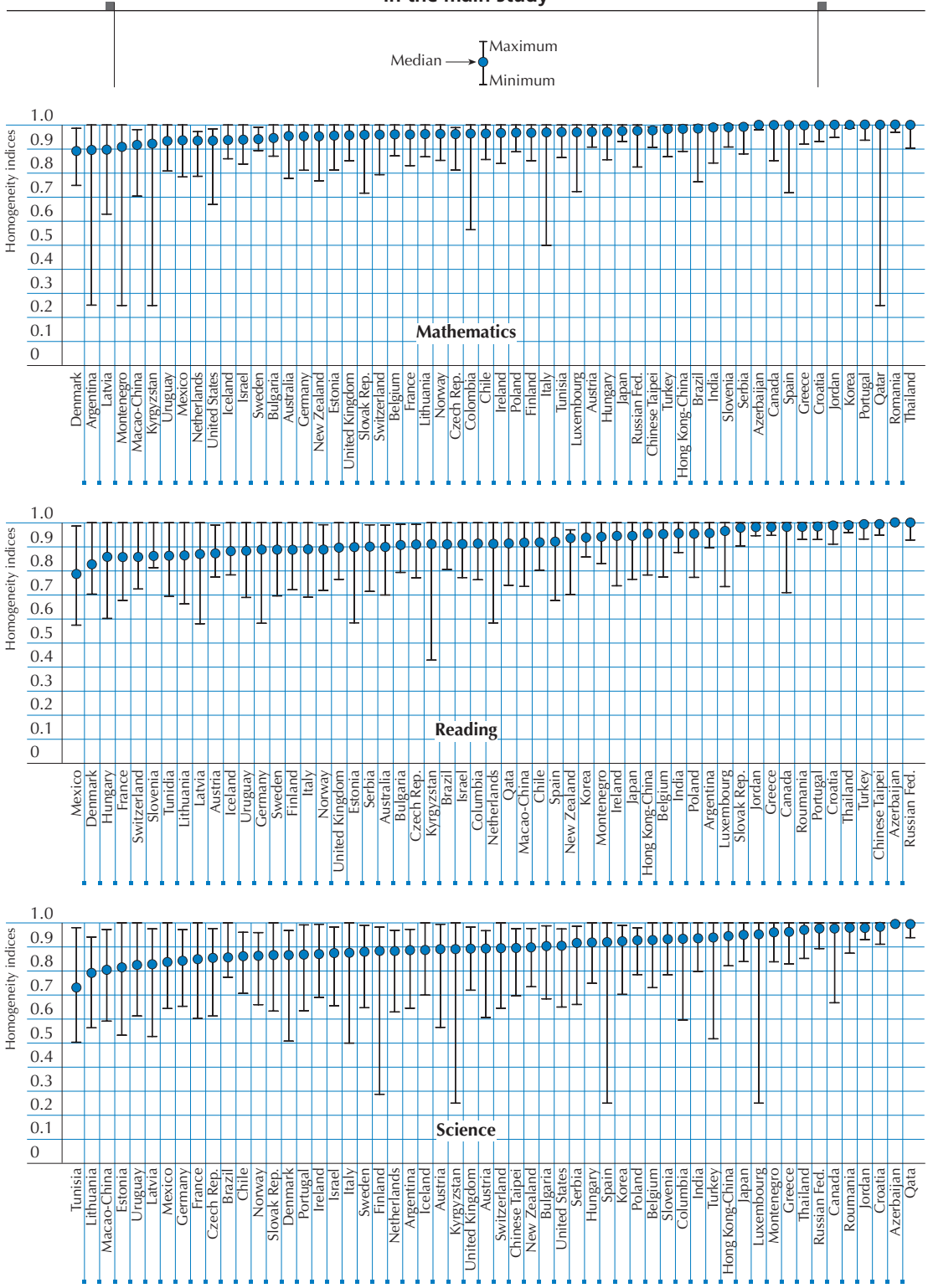




Figure 13.6
Variability of the homogeneity indices for the participating countries in the main study



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



For all items except one science item, *S524Q07,* the average index is greater than 0.80. Indices are higher for mathematics items which indicate that there is less disagreement between mathematics coders.

Figure 13.6 shows the distribution of homogeneity indices per domain and per country. There is more variability in the coding of reading and science than mathematics for most of the countries.

The results of the homogeneity analysis showed that the marking process of items is largely satisfactory and that on average countries are more or less reliable in the coding of the open-ended responses.

MULTIPLE MARKING STUDY OUTCOMES (VARIANCE COMPONENTS)

To obtain an estimate of the between-coder variability within each country, multiple coding was required for at least some student answers. Therefore, it was decided that multiple codings would be collected for open-ended items in both the field trial and the main study for a moderate number of students. In the main study, a selection of clusters from 600 students' booklets were multiply coded, with the full set of main study items requiring the judgement of a trained coder included in the exercise. The requirement was that the same four expert coders per domain (reading, mathematics and science) should code all items appearing together in the first two clusters of the test booklets 1, 3, 6, 8 and 10, and the first three clusters of booklet 5. A booklet 6 containing, for example, 14 reading items, would give a three-dimensional table for reading (100 students by 14 items by 4 markers), where each cell contains a single category. For each domain and each booklet, such a table was produced and processed in several analyses, which are described later. These data sets were required from each participating country.

Table 13.1 to Table 13.3 show the results of the variance components analysis for the multiply-marked items in mathematics, science, and reading, respectively. The variance components are each expressed as a percentage of their sum.

The tables show that those variance components associated with markers are small relative to the other components. This means that there are no significant systematic within-country marker effects.

Analyses of the type reported here can result in negative variance estimates. If the amount by which the component is negative is small, then this is a sign that the variance component is negligible (near zero). If the component is large and negative, then it is a sign that the analysis method is inappropriate for the data. In Table 13.1 to Table 13.3 countries with large inadmissible variance component estimates are indicated.

Generalisability coefficients

The generalisability coefficients are computed from the variance components using:

13.1

$$\rho_3(Y_{vg}, Y'_{vg}) = \frac{\sigma_A^2 + \frac{\sigma^2_{AB+E^*}}{I}}{\sigma_A^2 + \frac{\sigma^2_{AB+E^*}}{I} + \frac{\sigma^2_{ac}}{R} + \frac{\sigma^2_{abc+e^*}}{I \times R}}$$

and

13.2

$$\rho_3(Y_{vg}, Y'_{vg}) = \frac{\sigma_A^2 + \frac{\sigma^2_{AB}}{I}}{\sigma_A^2 + \frac{\sigma^2_{AB} + \sigma^2_{\varepsilon^*}}{I} + \frac{\sigma^2_{ac}}{R} + \frac{\sigma^2_{abc+e^*}}{I \times R}}$$

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



Table 13.1
Variance components for mathematics

	Student Component	Item Component	Marker Component	Student-item Interaction Component	Student-Marker Interaction Component	Item-Marker Interaction Component	Measurement Error component
Argentina	17.10	30.40	0.01	46.70	0.00	0.10	5.70
Australia	17.47	31.05	0.07	45.01	-0.02	0.10	6.32
Austria	25.21	19.34	0.00	51.76	0.02	0.07	3.60
Azerbaijan	9.53	27.04	0.00	63.31	0.00	0.00	0.11
Belgium (Dutch)	17.77	23.19	0.01	54.35	-0.09	0.03	4.73
Belgium (French)	23.82	17.55	0.03	54.09	0.17	0.02	4.32
Brazil	24.30	8.59	0.03	62.69	0.05	0.03	4.31
Bulgaria	16.86	17.02	0.00	59.11	-0.24	0.04	7.20
Canada (English)[1]	18.85	28.96	0.42	43.31	-20.00	-0.21	28.66
Canada (French)	11.73	30.86	0.01	52.52	0.03	0.12	4.72
Chile	17.58	21.00	0.02	55.57	-0.03	0.00	5.85
Colombia	14.69	21.93	0.00	59.18	-0.08	0.02	4.26
Croatia	13.84	23.03	0.00	62.20	0.01	0.01	0.91
Czech Republic	21.25	17.82	0.00	56.67	0.06	0.09	4.11
Denmark[1]	19.64	20.70	0.21	52.05	-4.87	0.15	12.13
Estonia (Estonian)[1]	10.71	30.09	0.01	52.19	-2.77	0.26	9.50
Estonia (Russian)	13.67	30.64	0.10	50.39	0.03	0.40	4.76
Finland	14.32	27.33	0.01	53.64	-0.06	0.08	4.69
France	23.78	17.25	0.02	53.40	0.05	0.09	5.42
Germany	18.72	21.24	0.00	53.14	-0.01	0.21	6.70
Greece	20.28	22.47	0.00	56.19	-0.01	0.00	1.06
Hong Kong-China	15.07	21.98	0.00	58.70	-0.06	0.10	4.21
Hungary	15.38	30.20	-0.01	51.08	0.04	0.03	3.28
Iceland	14.50	23.77	0.02	55.38	0.15	0.09	6.09
Indonesia	19.12	15.73	0.01	60.62	0.02	0.03	4.47
Ireland	16.38	29.41	0.01	48.39	-0.03	0.10	5.74
Israel	18.16	22.60	0.01	52.54	-0.04	0.10	6.63
Italy (German)	15.20	37.60	0.02	42.44	-0.06	0.09	4.71
Italy (Italian)	21.61	16.48	0.21	57.72	0.03	0.01	3.94
Japan	17.20	23.20	0.00	57.17	0.04	0.03	2.36
Jordan	13.09	18.00	0.00	67.75	0.00	0.01	1.15
Korea	20.66	18.43	0.00	60.36	-0.01	0.00	0.56
Kyrgyzstan (Kyrgyz)	6.12	6.31	-0.06	69.98	-0.42	0.64	17.44
Kyrgyzstan (Russian)	19.28	11.85	-0.02	63.56	-0.23	0.18	5.37
Latvia (Latvian)	16.37	19.34	0.30	49.87	0.08	1.21	12.83
Latvia (Russian)	13.47	26.62	0.33	46.52	0.42	0.60	12.04
Lithuania	18.69	21.88	0.01	54.41	-0.05	0.06	5.00
Luxembourg (French)	16.75	32.86	-0.02	44.01	-0.30	-0.04	6.74
Luxembourg (German)	23.12	19.92	0.00	54.45	-0.16	0.02	2.66
Macao-China	18.32	16.86	0.03	54.60	0.01	0.36	9.82
Mexico	14.35	19.35	0.04	56.47	0.07	0.13	9.58
Montenegro	17.89	11.30	0.06	58.26	-0.21	0.35	12.35
Netherlands	13.78	31.80	0.01	47.04	0.03	0.09	7.25
New Zealand	16.12	27.56	0.00	50.42	0.07	0.05	5.78
Norway	18.56	25.77	0.00	50.99	-0.06	0.02	4.72
Poland	24.57	13.30	0.00	57.94	0.05	0.04	4.10
Portugal	15.82	20.96	0.00	62.30	0.01	0.00	0.92
Qatar (Arabic)	14.44	9.16	0.00	74.83	-0.04	0.00	1.61
Qatar (English)	43.64	9.28	0.00	46.87	0.01	0.00	0.20
Romania	18.66	14.99	0.00	66.11	0.00	0.00	0.24
Russian Federation	20.30	25.91	0.02	50.33	0.00	0.08	3.37
Serbia	21.57	16.67	0.00	59.81	-0.03	0.00	1.99
Slovakia	22.10	21.58	0.00	50.22	0.00	0.07	6.03
Slovenia	15.72	18.08	0.00	64.36	0.43	0.01	1.41
Spain (Basque)	33.64	10.60	-0.01	53.17	0.00	-0.02	2.62
Spain (Catalan)	14.64	26.15	0.02	50.16	0.09	0.47	8.47
Spain (Galician)	14.83	30.01	0.06	48.90	-0.01	0.40	5.82
Spain (Spanish)	16.65	24.35	-0.05	54.24	0.05	0.30	4.44
Spain (Valencian)	5.70	36.88	0.14	46.23	-0.04	0.16	10.93
Sweden	16.05	27.62	-0.01	51.45	-0.03	0.04	4.87
Switzerland (French)	11.89	33.15	0.00	48.19	-0.02	0.08	6.71
Switzerland (German)	18.60	24.20	0.00	53.92	0.00	0.02	3.26
Chinese Taipei	20.13	15.33	0.00	61.05	-0.05	0.01	3.52
Thailand	20.52	18.17	0.00	60.05	0.05	0.01	1.21
Tunisia	16.04	10.82	0.01	68.03	-0.11	0.03	5.18
Turkey	27.17	9.63	0.00	60.26	0.00	0.02	2.93
United Kingdom (Scotland)	16.77	27.09	-0.01	51.35	-0.08	0.10	4.77
United Kingdom (The rest of)	17.02	32.82	0.01	44.69	-0.05	0.03	5.49
United States[1]	20.34	28.66	0.12	44.50	-5.78	0.03	12.13
Uruguay	16.42	20.70	0.01	56.24	-0.12	0.13	6.62

1. Countries with large inadmissible variance component estimates.



Table 13.2
Variance components for science

	Student Component	Item Component	Marker Component	Student-item Interaction Component	Student-Marker Interaction Component	Item-Marker Interaction Component	Measurement Error component
Argentina[1]	15.72	14.84	0.05	55.60	-3.30	0.20	16.89
Australia	17.26	23.19	0.00	47.53	0.02	0.43	11.56
Austria	17.37	20.17	0.00	50.23	-0.01	0.31	11.93
Azerbaijan	15.70	6.51	0.00	77.75	0.00	0.00	0.04
Belgium (Dutch)	13.78	28.44	0.02	49.48	0.00	0.17	8.12
Belgium (French)	17.39	22.53	0.02	54.44	0.04	0.04	5.54
Brazil	18.84	10.23	0.01	55.49	-0.08	0.65	14.86
Bulgaria	28.82	8.73	0.17	52.83	0.17	0.39	8.88
Canada (English)[1]	16.41	21.80	0.38	44.25	-10.49	0.46	27.19
Canada (French)	16.37	19.79	0.20	49.49	0.06	0.55	13.54
Chile	18.95	15.26	0.06	51.14	0.29	0.26	14.05
Colombia	15.28	13.22	0.01	61.50	0.01	0.07	9.91
Croatia	12.27	24.62	0.00	61.26	0.01	0.01	1.83
Czech Republic	16.80	21.08	0.02	48.07	-0.02	0.57	13.48
Denmark[1]	18.41	17.41	0.03	50.08	-1.98	0.27	15.78
Estonia (Estonian)[1]	16.41	26.43	0.10	42.93	-2.67	0.85	15.95
Estonia (Russian)	16.74	18.45	0.34	43.04	-0.14	1.37	20.20
Finland[1]	14.57	27.12	0.25	48.10	-1.58	0.36	11.18
France	16.37	24.24	0.05	46.27	0.05	0.43	12.58
Germany	16.08	18.59	0.09	50.13	0.15	0.80	14.15
Greece	18.55	19.32	0.00	59.00	0.02	0.02	3.07
Hong Kong-China	15.45	27.83	0.02	50.16	0.01	0.02	6.51
Hungary	16.06	15.43	0.01	59.70	0.13	0.12	8.56
Iceland	15.64	20.44	0.04	51.98	0.09	0.18	11.63
Indonesia	12.60	10.96	0.00	65.23	-0.93	0.56	11.57
Ireland	14.71	23.97	0.04	48.64	0.13	0.41	12.09
Israel	25.01	17.19	0.07	47.75	0.10	0.13	9.76
Italy (German)	16.11	21.08	-0.03	49.34	0.13	0.26	13.12
Italy (Italian)	16.19	15.99	0.63	56.47	0.00	0.14	10.57
Japan	19.37	22.93	0.01	54.02	0.03	0.03	3.61
Jordan	21.68	12.46	0.00	63.10	0.01	0.00	2.75
Korea	16.94	21.27	0.05	53.19	0.06	0.18	8.31
Kyrgyzstan (Kyrgyz)	10.79	7.64	0.01	65.64	0.28	0.35	15.30
Kyrgyzstan (Russian)	15.59	8.93	0.02	66.72	0.02	0.07	8.65
Latvia (Latvian)	13.92	19.55	0.10	48.34	0.12	1.10	16.87
Latvia (Russian)	16.15	22.47	-0.04	42.92	0.18	1.12	17.18
Lithuania	17.26	18.37	0.06	43.13	0.44	1.62	19.14
Luxembourg (French)	21.75	13.02	0.05	58.75	0.20	0.01	6.22
Luxembourg (German)	15.44	20.49	-0.02	56.92	0.10	0.27	6.80
Macao-China	12.76	23.01	0.44	44.02	0.07	1.39	18.31
Mexico	12.50	12.60	0.07	49.63	0.22	0.45	24.53
Montenegro	16.89	12.10	0.00	66.07	0.10	0.03	4.80
Netherlands	16.28	24.28	0.58	45.58	-0.31	0.73	12.87
New Zealand	18.50	19.56	0.08	50.95	0.06	0.12	10.73
Norway	17.80	14.33	0.09	52.65	0.04	0.50	14.59
Poland	14.72	23.42	0.01	54.92	0.02	0.03	6.87
Portugal	14.96	22.03	0.03	50.40	0.16	0.20	12.21
Qatar (Arabic)	17.95	14.35	0.00	66.09	0.03	0.00	1.59
Qatar (English)	21.19	15.59	0.00	61.83	-0.02	-0.01	1.41
Romania	18.44	10.98	0.00	68.08	-0.02	0.01	2.52
Russian Federation	15.99	16.22	0.00	65.18	0.00	0.00	2.60
Serbia	16.86	14.38	0.06	58.77	0.22	0.36	9.35
Slovakia	18.51	16.84	0.20	51.58	0.20	0.36	12.31
Slovenia	22.32	18.30	0.01	52.73	0.06	0.11	6.47
Spain (Basque)	13.59	21.27	0.04	57.83	-0.11	0.12	7.26
Spain (Catalan)	15.13	20.45	0.48	43.02	0.11	1.31	19.51
Spain (Galician)	11.88	23.02	0.13	50.36	0.14	0.47	13.99
Spain (Spanish)	14.73	21.99	0.43	52.56	0.02	0.27	10.00
Spain (Valencian)	17.16	6.92	0.55	49.05	-0.45	0.65	26.13
Sweden	17.52	19.97	0.00	51.49	0.07	0.20	10.76
Switzerland (French)	16.92	22.08	0.01	50.82	0.06	0.42	9.69
Switzerland (German)	20.69	19.54	0.05	50.05	0.09	0.23	9.36
Chinese Taipei	13.27	26.43	0.00	50.87	0.10	0.19	9.14
Thailand	15.72	17.45	0.01	62.73	-0.01	0.04	4.06
Tunisia	13.63	13.66	0.20	46.36	0.21	1.04	24.90
Turkey	17.33	11.62	0.25	59.48	0.17	0.26	10.89
United Kingdom (Scotland)	16.41	25.52	0.06	47.49	-0.04	0.20	10.35
United Kingdom (The rest of)	16.74	22.77	0.04	50.22	0.25	0.15	9.82
United States	20.67	17.06	0.01	51.45	0.06	0.15	10.60
Uruguay	15.82	15.23	0.04	53.34	0.09	0.75	14.73

1. Countries with large inadmissible variance component estimates.



Table 13.3
Variance components for reading

	Student Component	Item Component	Marker Component	Student-item Interaction Component	Student-Marker Interaction Component	Item-Marker Interaction Component	Measurement Error component
Argentina	21.35	20.82	0.00	54.35	0.01	0.03	3.44
Australia	23.78	23.57	0.01	41.80	0.05	0.19	10.60
Austria	20.50	13.19	0.20	52.75	0.02	0.52	12.81
Azerbaijan	25.28	8.64	0.00	66.08	0.00	0.00	0.00
Belgium (Dutch)	11.44	26.77	0.05	49.91	-0.09	0.25	11.66
Belgium (French)	21.50	14.83	0.00	59.21	0.18	0.00	4.28
Brazil	13.94	19.18	0.08	56.03	0.11	0.27	10.39
Bulgaria	31.00	13.90	0.00	48.38	0.03	0.02	6.67
Canada (English)[1]	16.86	26.80	0.01	45.22	-10.00	-0.20	21.30
Canada (French)	18.56	21.19	0.03	46.47	0.11	0.76	12.89
Chile	15.01	31.49	0.01	44.11	0.13	0.15	9.10
Colombia	14.58	21.06	-0.01	52.57	0.20	0.19	11.42
Croatia	15.40	20.46	0.02	61.03	0.02	0.03	3.04
Czech Republic	27.10	14.40	0.00	48.17	0.13	0.39	9.81
Denmark[1]	19.07	12.83	-0.02	46.26	-2.34	1.61	22.58
Estonia (Estonian)[1]	10.76	27.07	-0.01	51.22	-2.28	0.18	13.06
Estonia (Russian)	17.53	22.53	-0.10	40.40	-0.26	2.11	17.79
Finland	14.55	19.31	0.10	53.07	0.04	0.17	12.76
France	19.76	24.01	0.26	39.17	-0.10	1.37	15.54
Germany	21.68	14.11	0.00	51.31	-0.01	0.09	12.83
Greece	22.47	23.43	0.01	52.00	-0.02	0.00	2.10
Hong Kong-China	14.07	28.02	0.03	49.10	0.00	0.35	8.43
Hungary	22.87	16.36	0.16	43.00	0.57	0.52	16.52
Iceland	19.31	10.33	0.01	54.22	0.04	0.62	15.48
Indonesia	11.82	18.34	0.01	64.22	0.02	0.09	5.51
Ireland	22.66	21.22	0.06	45.78	0.07	0.14	10.07
Israel	16.79	22.92	0.08	49.54	0.07	0.24	10.36
Italy (German)	20.24	19.88	0.12	44.21	-0.15	0.12	15.58
Italy (Italian)	20.56	22.60	-0.11	46.78	-0.06	0.22	10.01
Japan	20.64	11.12	0.01	62.33	0.10	0.10	5.70
Jordan	15.02	16.27	0.00	66.46	0.01	0.00	2.25
Korea	16.14	27.33	0.02	51.90	0.04	0.04	4.52
Kyrgyzstan (Kyrgyz)	5.79	6.91	-0.06	56.07	-0.35	0.48	31.15
Kyrgyzstan (Russian)	28.85	11.87	-0.02	51.91	0.06	0.18	7.16
Latvia (Latvian)	16.00	19.52	0.22	44.78	0.20	1.08	18.21
Latvia (Russian)	16.01	24.25	0.29	43.32	0.03	1.15	14.95
Lithuania	20.54	17.10	0.07	43.69	0.06	1.62	16.93
Luxembourg (French)	20.87	15.50	-0.01	57.46	0.17	0.00	6.01
Luxembourg (German)	25.32	14.35	0.00	53.28	0.27	0.02	6.76
Macao-China	10.09	29.36	0.13	45.75	0.08	0.77	13.82
Mexico	13.26	23.70	0.64	36.90	0.32	2.19	22.99
Montenegro	13.68	11.56	-0.01	67.32	0.98	0.01	6.45
Netherlands	16.50	17.90	0.01	53.33	-0.01	0.17	12.11
New Zealand	25.16	22.05	0.10	43.06	0.05	0.12	9.46
Norway	27.00	11.67	0.02	50.09	0.07	0.33	10.82
Poland	18.49	26.01	0.01	47.84	-0.02	0.07	7.60
Portugal	10.31	34.21	0.00	52.04	0.18	-0.01	3.27
Qatar (Arabic)	12.54	13.76	-0.01	64.69	0.07	0.08	8.86
Qatar (English)	21.17	19.44	-0.01	49.55	0.14	0.06	9.66
Romania	17.43	16.05	0.00	64.56	-0.03	0.01	1.97
Russian Federation	20.09	22.07	0.00	56.71	0.00	0.00	1.13
Serbia	18.94	14.08	0.04	53.45	0.11	0.24	13.14
Slovakia	15.95	25.65	0.00	54.64	0.00	0.08	3.69
Slovenia	19.16	22.90	0.00	45.59	0.01	0.25	12.09
Spain (Basque)	24.16	14.96	-0.01	44.31	0.00	0.25	16.33
Spain (Catalan)	16.20	24.84	0.82	37.18	0.04	1.79	19.12
Spain (Galician)	15.20	24.82	0.06	40.97	-0.02	0.56	18.41
Spain (Spanish)	19.28	23.30	0.26	42.92	0.21	0.33	13.69
Spain (Valencian)	29.85	18.79	1.20	28.88	0.29	1.44	19.55
Sweden	23.24	13.35	0.01	49.16	0.09	0.29	13.86
Switzerland (French)	14.60	23.53	-0.04	50.96	0.12	0.60	10.23
Switzerland (German)	18.70	15.67	0.05	52.11	-0.02	0.03	13.47
Chinese Taipei	13.21	37.15	0.00	48.09	-0.02	0.00	1.57
Thailand	14.89	20.25	0.00	63.23	0.00	0.01	1.62
Tunisia	16.24	16.85	-0.04	51.22	0.12	0.44	15.17
Turkey	14.57	19.68	0.00	63.89	0.01	0.00	1.84
United Kingdom (Scotland)	22.87	23.01	0.01	44.53	-0.01	0.10	9.49
United Kingdom (The rest of)	21.10	25.92	-0.01	44.14	0.02	0.05	8.77
United States[1]	26.42	22.04	-0.05	42.17	-2.10	-0.01	11.53
Uruguay	17.15	22.85	0.03	49.88	0.12	0.24	9.72

1. Countries with large inadmissible variance component estimates.



Table 13.4
Generalisability estimates for mathematics

	I=8 M=1		I=16 M=1		I=24 M=1	
	p3	p4	p3	p4	p3	p4
Argentina	0.97	0.72	0.98	0.84	0.99	0.89
Australia	0.97	0.73	0.98	0.85	0.99	0.89
Austria	0.99	0.78	0.99	0.88	0.99	0.92
Azerbaijan	1.00	0.55	1.00	0.71	1.00	0.78
Belgium (Dutch)	0.98	0.71	0.99	0.83	1.00	0.88
Belgium (French)	0.98	0.76	0.98	0.86	0.99	0.90
Brazil	0.98	0.74	0.99	0.85	0.99	0.90
Bulgaria	0.97	0.68	0.99	0.81	1.00	0.87
Canada (English)						
Canada (French)	0.97	0.62	0.98	0.77	0.98	0.83
Chile	0.97	0.70	0.98	0.82	0.99	0.87
Colombia	0.98	0.65	0.99	0.79	0.99	0.85
Croatia	0.99	0.64	1.00	0.78	1.00	0.84
Czech Republic	0.98	0.74	0.99	0.85	0.99	0.89
Denmark						
Estonia (Estonian)						
Estonia (Russian)	0.97	0.66	0.98	0.80	0.99	0.85
Finland	0.98	0.66	0.99	0.80	0.99	0.86
France	0.98	0.76	0.99	0.87	0.99	0.91
Germany	0.97	0.72	0.98	0.83	0.99	0.88
Greece	1.00	0.74	1.00	0.85	1.00	0.90
Hong Kong-China	0.98	0.66	0.99	0.80	0.99	0.86
Hungary	0.98	0.69	0.99	0.82	0.99	0.87
Iceland	0.96	0.65	0.97	0.78	0.98	0.84
Indonesia	0.98	0.70	0.99	0.82	0.99	0.88
Ireland	0.97	0.71	0.98	0.83	0.99	0.88
Israel	0.97	0.71	0.98	0.83	0.99	0.88
Italy (German)	0.98	0.72	0.99	0.84	0.99	0.89
Italy (Italian)	0.98	0.74	0.99	0.85	0.99	0.89
Japan	0.99	0.70	0.99	0.82	0.99	0.87
Jordan	0.99	0.60	1.00	0.75	1.00	0.82
Korea	1.00	0.73	1.00	0.85	1.00	0.89
Kyrgyzstan (Kyrgyz)	0.89	0.37	0.94	0.55	0.97	0.66
Kyrgyzstan (Russian)	0.98	0.70	1.00	0.83	1.00	0.88
Latvia (Latvian)	0.93	0.67	0.96	0.80	0.97	0.86
Latvia (Russian)	0.91	0.64	0.93	0.77	0.94	0.83
Lithuania	0.98	0.72	0.99	0.84	0.99	0.89
Luxembourg (French)	0.98	0.74	0.99	0.85	1.00	0.90
Luxembourg (German)	0.99	0.77	1.00	0.87	1.00	0.91
Macao-China	0.95	0.69	0.97	0.82	0.98	0.87
Mexico	0.94	0.63	0.96	0.77	0.97	0.84
Montenegro	0.95	0.68	0.98	0.81	0.99	0.87
Netherlands	0.96	0.67	0.97	0.80	0.98	0.86
New Zealand	0.97	0.69	0.98	0.82	0.98	0.87
Norway	0.98	0.73	0.99	0.84	0.99	0.89
Poland	0.98	0.76	0.99	0.86	0.99	0.90
Portugal	1.00	0.67	1.00	0.80	1.00	0.86
Qatar (Arabic)	0.99	0.60	1.00	0.75	1.00	0.82
Qatar (English)	1.00	0.88	1.00	0.94	1.00	0.96
Romania	1.00	0.69	1.00	0.82	1.00	0.87
Russian Federation	0.98	0.75	0.99	0.86	0.99	0.90
Serbia	0.99	0.74	1.00	0.85	1.00	0.89
Slovakia	0.97	0.76	0.99	0.86	0.99	0.90
Slovenia	0.98	0.65	0.97	0.78	0.97	0.83
Spain (Basque)	0.99	0.83	1.00	0.91	1.00	0.94
Spain (Catalan)	0.95	0.66	0.97	0.80	0.97	0.85
Spain (Galician)	0.97	0.69	0.98	0.81	0.99	0.87
Spain (Spanish)	0.98	0.69	0.98	0.82	0.99	0.87
Spain (Valencian)	0.90	0.45	0.93	0.62	0.95	0.71
Sweden	0.98	0.70	0.99	0.82	0.99	0.87
Switzerland (French)	0.96	0.64	0.97	0.78	0.98	0.84
Switzerland (German)	0.98	0.72	0.99	0.84	0.99	0.89
Chinese Taipei	0.99	0.72	0.99	0.84	1.00	0.88
Thailand	0.99	0.73	1.00	0.84	1.00	0.89
Tunisia	0.98	0.64	0.99	0.78	0.99	0.85
Turkey	0.99	0.78	0.99	0.87	1.00	0.91
United Kingdom (Scotland)	0.98	0.71	0.99	0.83	0.99	0.88
United Kingdom (The rest of)	0.97	0.73	0.99	0.85	0.99	0.89
United States						
Uruguay	0.97	0.68	0.99	0.81	0.99	0.87

Note: Countries with no value are displayed, because they fall outside the acceptable [0,1] range.



Table 13.5
Generalisability estimates for science

	I=8 M=1		I=16 M=1		I=24 M=1	
	p3	p4	p3	p4	p3	p4
Argentina						
Australia	0.94	0.70	0.97	0.82	0.98	0.87
Austria	0.94	0.69	0.97	0.82	0.98	0.87
Azerbaijan	1.00	0.62	1.00	0.76	1.00	0.83
Belgium (Dutch)	0.95	0.66	0.97	0.79	0.98	0.85
Belgium (French)	0.97	0.70	0.98	0.82	0.99	0.87
Brazil	0.94	0.68	0.96	0.81	0.98	0.87
Bulgaria	0.97	0.79	0.98	0.88	0.98	0.91
Canada (English)						
Canada (French)	0.93	0.67	0.96	0.80	0.97	0.86
Chile	0.93	0.69	0.95	0.81	0.96	0.86
Colombia	0.95	0.63	0.97	0.77	0.98	0.84
Croatia	0.99	0.61	0.99	0.76	0.99	0.82
Czech Republic	0.93	0.69	0.96	0.81	0.97	0.87
Denmark						
Estonia (Estonian)						
Estonia (Russian)	0.90	0.68	0.95	0.81	0.96	0.87
Finland						
France	0.93	0.69	0.96	0.82	0.97	0.87
Germany	0.92	0.66	0.95	0.79	0.96	0.85
Greece	0.99	0.71	0.99	0.83	0.99	0.88
Hong Kong-China	0.96	0.69	0.98	0.81	0.98	0.87
Hungary	0.95	0.65	0.97	0.79	0.97	0.84
Iceland	0.94	0.66	0.96	0.79	0.97	0.85
Indonesia	0.98	0.59	1.00	0.77	1.00	0.85
Ireland	0.93	0.66	0.95	0.79	0.96	0.85
Israel	0.96	0.77	0.98	0.87	0.98	0.91
Italy (German)	0.93	0.67	0.95	0.80	0.96	0.86
Italy (Italian)	0.95	0.66	0.97	0.79	0.98	0.85
Japan	0.98	0.73	0.99	0.84	0.99	0.89
Jordan	0.99	0.73	0.99	0.84	1.00	0.89
Korea	0.96	0.69	0.97	0.81	0.98	0.87
Kyrgyzstan (Kyrgyz)	0.90	0.51	0.92	0.67	0.94	0.75
Kyrgyzstan (Russian)	0.96	0.62	0.97	0.77	0.98	0.83
Latvia (Latvian)	0.90	0.63	0.94	0.77	0.95	0.83
Latvia (Russian)	0.90	0.68	0.94	0.80	0.95	0.86
Lithuania	0.89	0.68	0.92	0.80	0.94	0.85
Luxembourg (French)	0.97	0.72	0.98	0.84	0.98	0.88
Luxembourg (German)	0.96	0.66	0.97	0.79	0.98	0.85
Macao-China	0.89	0.62	0.93	0.76	0.95	0.83
Mexico	0.85	0.57	0.90	0.72	0.92	0.79
Montenegro	0.97	0.65	0.98	0.79	0.99	0.85
Netherlands	0.94	0.70	0.98	0.83	0.99	0.89
New Zealand	0.95	0.70	0.97	0.83	0.98	0.88
Norway	0.93	0.68	0.96	0.81	0.97	0.86
Poland	0.96	0.66	0.98	0.79	0.98	0.85
Portugal	0.93	0.65	0.95	0.79	0.96	0.84
Qatar (Arabic)	0.99	0.68	0.99	0.81	1.00	0.86
Qatar (English)	1.00	0.73	1.00	0.84	1.00	0.89
Romania	0.99	0.68	0.99	0.81	1.00	0.86
Russian Federation	0.99	0.65	0.99	0.79	0.99	0.85
Serbia	0.95	0.66	0.96	0.79	0.97	0.85
Slovakia	0.94	0.69	0.96	0.82	0.97	0.87
Slovenia	0.97	0.75	0.98	0.86	0.99	0.90
Spain (Basque)	0.96	0.63	0.98	0.77	0.99	0.84
Spain (Catalan)	0.89	0.66	0.93	0.79	0.95	0.85
Spain (Galician)	0.91	0.59	0.94	0.74	0.95	0.81
Spain (Spanish)	0.94	0.65	0.97	0.79	0.98	0.85
Spain (Valencian)	0.89	0.66	0.95	0.80	0.97	0.87
Sweden	0.94	0.69	0.97	0.82	0.97	0.87
Switzerland (French)	0.95	0.69	0.97	0.82	0.98	0.87
Switzerland (German)	0.96	0.73	0.97	0.85	0.98	0.89
Chinese Taipei	0.94	0.64	0.96	0.78	0.97	0.84
Thailand	0.98	0.65	0.99	0.79	0.99	0.85
Tunisia	0.85	0.60	0.90	0.75	0.93	0.81
Turkey	0.94	0.66	0.96	0.79	0.97	0.85
United Kingdom (Scotland)	0.95	0.70	0.97	0.82	0.98	0.87
United Kingdom (The rest of)	0.94	0.68	0.96	0.81	0.97	0.86
United States	0.95	0.73	0.97	0.84	0.98	0.89
Uruguay	0.92	0.65	0.95	0.79	0.96	0.84

Note: Countries with no value are displayed, because they fall outside the acceptable [0,1] range.



Table 13.6
Generalisability estimates for reading

	I=8 M=1		I=16 M=1		I=24 M=1	
	p3	p4	p3	p4	p3	p4
Argentina	0.99	0.75	0.99	0.86	0.99	0.90
Australia	0.96	0.78	0.97	0.88	0.98	0.91
Austria	0.94	0.71	0.97	0.83	0.98	0.88
Azerbaijan	1.00	0.75	1.00	0.86	1.00	0.90
Belgium (Dutch)	0.93	0.60	0.96	0.75	0.97	0.82
Belgium (French)	0.98	0.73	0.98	0.84	0.99	0.88
Brazil	0.94	0.62	0.96	0.77	0.97	0.83
Bulgaria	0.98	0.82	0.99	0.90	0.99	0.93
Canada (English)						
Canada (French)	0.93	0.71	0.96	0.83	0.97	0.88
Chile	0.94	0.69	0.96	0.81	0.97	0.87
Colombia	0.93	0.64	0.95	0.78	0.96	0.84
Croatia	0.98	0.66	0.99	0.79	0.99	0.85
Czech Republic	0.96	0.79	0.98	0.88	0.98	0.91
Denmark						
Estonia (Estonian)						
Estonia (Russian)	0.92	0.71	0.96	0.84	0.98	0.89
Finland	0.93	0.64	0.96	0.78	0.97	0.84
France	0.93	0.75	0.96	0.86	0.98	0.90
Germany	0.95	0.73	0.97	0.84	0.98	0.89
Greece	0.99	0.77	1.00	0.87	1.00	0.91
Hong Kong-China	0.95	0.66	0.97	0.80	0.98	0.85
Hungary	0.92	0.74	0.94	0.84	0.95	0.88
Iceland	0.93	0.69	0.96	0.82	0.97	0.87
Indonesia	0.97	0.58	0.98	0.73	0.98	0.80
Ireland	0.96	0.76	0.97	0.86	0.98	0.90
Israel	0.94	0.69	0.97	0.82	0.97	0.87
Italy (German)	0.94	0.73	0.97	0.85	0.98	0.90
Italy (Italian)	0.96	0.75	0.98	0.86	0.98	0.90
Japan	0.97	0.71	0.98	0.83	0.99	0.88
Jordan	0.99	0.64	0.99	0.78	0.99	0.84
Korea	0.97	0.70	0.98	0.82	0.99	0.87
Kyrgyzstan (Kyrgyz)	0.78	0.35	0.85	0.53	0.90	0.64
Kyrgyzstan (Russian)	0.97	0.80	0.98	0.89	0.99	0.92
Latvia (Latvian)	0.90	0.67	0.93	0.80	0.95	0.85
Latvia (Russian)	0.92	0.69	0.95	0.81	0.97	0.87
Lithuania	0.92	0.73	0.95	0.84	0.97	0.89
Luxembourg (French)	0.97	0.72	0.98	0.84	0.98	0.88
Luxembourg (German)	0.97	0.77	0.98	0.86	0.98	0.90
Macao-China	0.90	0.57	0.93	0.73	0.95	0.80
Mexico	0.85	0.63	0.90	0.77	0.92	0.83
Montenegro	0.93	0.57	0.93	0.71	0.93	0.77
Netherlands	0.94	0.67	0.96	0.80	0.97	0.86
New Zealand	0.96	0.79	0.98	0.88	0.98	0.92
Norway	0.96	0.78	0.98	0.87	0.98	0.91
Poland	0.96	0.73	0.98	0.84	0.99	0.89
Portugal	0.97	0.59	0.97	0.74	0.98	0.81
Qatar (Arabic)	0.95	0.58	0.96	0.73	0.97	0.80
Qatar (English)	0.95	0.74	0.97	0.85	0.98	0.89
Romania	0.99	0.68	1.00	0.81	1.00	0.86
Russian Federation	1.00	0.74	1.00	0.85	1.00	0.89
Serbia	0.94	0.69	0.96	0.82	0.97	0.87
Slovakia	0.98	0.69	0.99	0.81	0.99	0.87
Slovenia	0.94	0.73	0.97	0.84	0.98	0.89
Spain (Basque)	0.94	0.76	0.96	0.86	0.98	0.91
Spain (Catalan)	0.90	0.70	0.94	0.82	0.96	0.87
Spain (Galician)	0.90	0.67	0.94	0.81	0.96	0.86
Spain (Spanish)	0.93	0.73	0.95	0.84	0.96	0.88
Spain (Valencian)	0.92	0.83	0.95	0.90	0.97	0.93
Sweden	0.94	0.75	0.97	0.85	0.97	0.90
Switzerland (French)	0.94	0.65	0.96	0.79	0.97	0.85
Switzerland (German)	0.94	0.70	0.96	0.82	0.98	0.87
Chinese Taipei	0.99	0.68	1.00	0.81	1.00	0.87
Thailand	0.99	0.65	1.00	0.79	1.00	0.85
Tunisia	0.92	0.66	0.95	0.79	0.96	0.85
Turkey	0.99	0.64	0.99	0.78	1.00	0.84
United Kingdom (Scotland)	0.96	0.77	0.98	0.87	0.99	0.91
United Kingdom (The rest of)	0.96	0.76	0.98	0.86	0.98	0.91
United States						
Uruguay	0.95	0.69	0.97	0.82	0.97	0.87

Note: Countries with no value are displayed, because they fall outside the acceptable [0,1] range.



They provide an index of reliability for the multiple marking in each country. *I* denotes the number of items and *M* the number of markers. By using different values for *I* and *M*, one obtains a generalisation of the Spearman-Brown formula for test-lengthening. In Table 13.4 to Table 13.6 the formula is evaluated for the three combinations of $I = \{8, 16, 24\}$ and $M = 1$, using the variance component estimates from the corresponding tables presented above. For some countries, no values are displayed, because they fall outside the acceptable (0,1) range.

INTERNATIONAL CODING REVIEW

An international coding review (ICR) was conducted as one of the PISA 2006 quality control procedures in order to investigate the possibility of systematic differences among countries in the coding of open-ended items. The objective of this study was to estimate potential bias (either leniency or harshness) in each country's PISA results, and to express this potential bias in the same units as are used to report country performance on the PISA scales.

The need for the ICR arises because the manual coding of student responses to certain test items is performed by coders trained at the national level. This introduces the possibility of national-level bias in the resulting PISA scores. Coders in country A may interpret and apply the coding instructions more or less leniently than coders in country B.

The data used for the ICR were generated from the multiple coding study. That study, described above, had been implemented earlier to test consistency among coders within each country, and to compare that degree of consistency across countries. Some of the student responses and their multiple codes were selected from the multiple coding study for inclusion in the ICR. These responses, which had already been coded by four national coders, were coded a fifth time by an independent verifier (and in some cases were coded a sixth time by an international adjudicator) to enable estimation of a potential bias.

Background to changed procedures for PISA 2006

Similar ICR studies had been conducted as part of PISA 2000 and PISA 2003 surveys. However, during 2005 and 2006, a review of procedures that had been used previously suggested that improvements and efficiencies could be achieved. The main conclusions from the first two survey cycles were that on the basis of analyses using percentage of agreement among coders, verifiers and adjudicators, there was little evidence of any systematic problems with the application of coding standards; that the relatively small number of problems observed seemed to apply only to particular items (for example only some of the more difficult items) and to only one or two coders in particular national centres. The most useful outcomes of the process, therefore, had been in providing quite specific and detailed information to national centres that would assist them in their own review of coder training procedures, relating either to individual items or to individual coders.

The ICR review called for a simplification of procedures, and most importantly called for the addition of a new element – a way of quantifying the potential impact of any evidence of discrepant coding at the national level on a country's performance. Specifically, a potential bias (degree of harshness or leniency of the coding in each country) expressed in PISA score units, was seen as the most useful way of describing the outcomes of any future ICR.

ICR procedures

Revised procedures designed to estimate national-level bias in coding were developed during the latter part of 2006 and implemented during 2007, achieving simplification and improving effectiveness and efficiency in comparison with procedures used previously. Preliminary planning for the ICR saw the consortium identify a set of booklet types and a set of items for inclusion in the study. Three booklets were chosen: booklet 5 (from which 15 science items were selected, of the 42 science items in total requiring manual coding), booklet 6



(from which 14 of the available 17 manually coded reading items were selected), and booklet 8 (from which 9 mathematics items were selected, of the 20 mathematics items altogether requiring manual coding).

These booklets and items were also amongst those used previously in the multiple coding study. A random selection was made of 60 of these booklets for each domain from each distinct coding centre within all adjudicated PISA entities (and selecting a representative proportion of each language involved). This meant that 900 responses to science items, 840 responses to reading items, and 540 responses to mathematics items were available from each national coding centre for examination in the ICR. The codes that had been assigned to the student responses to these items by the four national coders involved previously in the multiple coder study were extracted. Coding of each student response a fifth time was then carried out by a member of a team of independent reviewers who had been trained specifically for this task. These independent reviewers had been involved as part of the international translation verification team. The code assigned by the independent reviewer was referred to as the verifier code.

The ICR analysis procedures were carried out in two related but independent parts. The first part was aimed at identifying countries in which evidence of coder bias exists, and estimating the magnitude of that bias. The second part was aimed at identifying particular items, student responses, and coders, that tended to generate coding discrepancies.

Part 1: Flagging countries

The main goal of the analysis of the ICR data was to express leniency or harshness of national coders as an effect on countries' mean performance in each PISA domain. For some countries, where national coding was performed by different teams each having responsibility for student responses in different languages, results were analysed separately for language-based subgroups. To perform this analysis, the domain-ability (using weighted likelihood estimates, or WLEs) of each of the 60 selected students was estimated twice: once using the original reported score on all items from that domain in the relevant booklet; and once with the verifier codes substituted for each item response from that booklet that had been included in the ICR. The scores for items not included in the ICR stayed unchanged in the two estimations. The reported scores for each student were derived from a mixture of about 25% of codes from each of the four national coders involved in the Multiple Coder Study. The abilities were transformed to the PISA scale. This resulted in a maximum of 60 pairs of ability estimates, from which 60 differences were calculated. The average of the differences in each country was an indication of the bias in country mean performance for that domain. In fact a 95% confidence interval was constructed around the mean difference, and if that interval did not contain the value zero then potential bias was indicated.

A *t*-test was then performed on the paired ability estimates to test for significance of the difference in country mean performance. If the country mean performance that was based on the verifier codes differed significantly from the mean performance based on the reported scores, the country was flagged as having a potential bias in their average score for that domain. Before confirming this potential bias, the consortium implemented one final quality check: a review to judge the quality of the verifier codes. This final review is referred to as adjudication.

Nineteen responses were randomly selected for each flagged country by domain (by language) combination for adjudication. Before selecting these responses, cases with perfect agreement amongst the five coders were excluded, because it is highly likely that the adjudicator would agree with the verifier in these cases. The 19 responses that were selected were sent to an international adjudicator, along with the five previously assigned codes. This review and adjudication was carried out by the consortium staff member responsible



for leading the relevant domain. The adjudicator provided a single definitive code to each of the sampled student responses, which had been back-translated into English for this purpose.

The overall percentage of agreement between verifier and adjudicator for one domain in one country was estimated based on their coding of the 19 responses. Two assumptions had to be made for this estimation: (1) that the percentage of agreement between verifier and adjudicator would have been 100% for the excluded responses that had perfect agreement among the first five coders, and (2) that the percentage of agreement on the 19 responses could be generalised to the responses that were randomly not selected for adjudication.

The percentage agreement, $\hat{P}$, between verifier and adjudicator was therefore estimated as follows:

13.3

$$\hat{P} = \frac{[n + (N - n)Z]100}{N}$$

where n is the number of responses for which there was perfect agreement among verifier and all four national coders, Z is the observed proportion of adjudicated responses for which the adjudicator and verifier agreed, and N is the total number of responses (usually 60).

The estimated percentage of agreement between verifier and adjudicator was used to assess the quality of the verifier codes. If the percentage was 90 or above, the coding from the verifier was deemed to be correct and the estimated national bias was reported. If the percentage was below 90, the verifier codes were deemed to be not sufficiently reliable to justify confirmation of the observed difference in country mean.

Part 2: Flagging responses

The second part of the ICR procedure for PISA 2006 aimed to give a more in-depth picture of differences between national coders and international verifiers by country, language, domain and item, in order to support evaluation and improvement processes within countries.

After international verifiers completed their coding of the 900 science, 840 reading and 540 mathematics responses for each country, their codes were compared to the four codes given by the national coders. Two types of inconsistencies between national codes and verifier codes were flagged:

- When the verifier code was compared with each of the four national codes in turn, fewer than two matches were observed;
- The average raw score of the 4 coders was at least 0.5 points higher or lower than the score based on the verifier code.

Examples of flagged cases are given in Table 13.6.

Table 13.7
Examples of flagged cases

CNT	Student ID	Question	Coder1	Coder2	Coder3	Coder4	Verifier	Flag (Y/N)
xxx	Xxxxx00001	R067Q04	0	1	1	1	1	N
xxx	Xxxxx00012	R067Q04	1	1	1	1	0	Y
xxx	Xxxxx00031	R067Q04	1	1	1	0	0	Y
xxx	Xxxxx00014	R067Q04	0	1	1	2	0	Y
xxx	Xxxxx00020	R067Q04	1	0	2	1	2	Y
xxx	Xxxxx00025	R067Q04	2	0	2	0	2	Y

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



In addition to flagging cases of discrepancy between national coders and verifier, the individual items figuring more frequently in these discrepancies were also identified for each country. The difference between the mean raw score from the four national codes and the raw score from the verifier code was calculated item by item. The 60 differences per item (in case of one test language) were averaged. A positive difference for a particular item was an indication of leniency of national coders for that item, a negative difference an indicator of harshness of national coders. The number and percentages of flagged responses and mean differences per item were reported back to national centres as described later in this chapter.

Outcomes

Sixty-seven units of analysis were involved in the ICR study for PISA 2006, each comprising a country or a language-based group within a country. Each unit was analysed for the three assessment domains of science, reading and mathematics. Of these 67 units, in the first stage of the analysis (Part 1: Flagging countries), 26 were flagged for adjudication in mathematics, 41 in reading and 29 in science. These are summarised in Table 13.8.

Table 13.8
Count of analysis groups showing potential bias, by domain

Potential difference indicated	Mathematics	Reading	Science	Total (%)
Harshness in national coding	9	13	14	36 (17.9%)
No significant difference	41	26	38	105 (52.2%
Leniency in national coding	17	28	15	60 (29.9%)
Total Analysis Groups	**67**	**67**	**67**	**201 (100%)**

In order to confirm the potential bias indicated by this flagging process, the overall consistency of the adjudicator and verifier codes was checked. Table 13.9 shows an overall summary of this comparison. In over 60% of the individual cases (across the three domains) the adjudicator agreed with the code assigned by the verifier.

Table 13.9
Comparison of codes assigned by verifier and adjudicator

Difference (Verifier-Adjudicator)	Number of Cases	Percent
–2	58	3.7
–1	293	18.6
0	952	60.5
1	241	15.3
2	30	1.9
Total Cases	**1574**	**100.0**

After adjudication, differences between mean performance for the 67 units of analysis using the reported codes and the verifier codes were judged to be significant in 22 units for mathematics, 20 for reading and 13 for science. The units are listed in Table 13.10. The '+' symbol indicates that the difference was positive, suggesting potential lenience in the national coding. The '–' symbol indicates that the difference was negative, suggesting potential harshness in the national coding. Blank cells indicate either no evidence of bias, or that evidence of bias was not confirmed by the adjudicator. Of the 55 units in which the difference was confirmed, 30 cases indicated positive bias (leniency in national coding) and 25 cases indicated negative bias (harshness in national coding).

In total, 25 cases of harshness in the standards applied in national coding centres were detected, alongside 30 cases of lenient coding at national level.



Table 13.10
Outcomes of ICR analysis part 1

	Reading	Mathematics	Science
Argentina		+	
Australia	+		
Austria			–
Azerbaijan	+	+	
Belgium (FLA)	–	+	
Belgium (FRA)		–	
Brazil			
Bulgaria			+
Canada (ENG)			
Canada (FRA)			
Chile		+	
Colombia			
Croatia	–		
Czech Republic			+
Denmark		–	
Estonia (EST)		+	+
Estonia (RUS)			
Finland			+
France		–	
Germany			
Greece	–		+
Hong Kong-China			–
Hungary		+	
Iceland		+	
Indonesia			
Ireland	+		
Israel			+
Italy			+
Japan			
Jordan			
Korea			–
Kyrgyzstan (KIR)	+		
Kyrgyzstan (RUS)	+		
Latvia (LVA)	–		
Latvia (RUS)	+	+	
Lithuania			
Luxembourg	+		
Macao-China	–	–	
Mexico			
Montenegro	–	–	
Netherlands			
New Zealand			
Norway			
Poland			
Portugal		–	
Qatar (ARA)	+	+	–
Qatar (ENG)	+	+	
Romania	–		
Russian Federation		–	
Serbia	–		
Slovak Republic			+
Slovenia			
Spain (BAQ)			
Spain (CAT)	–		
Spain (GLG)	–		
Spain (SPA)	–		
Sweden			
Switzerland (FRE)			
Switzerland (GER)	–		
Chinese Taipei		+	
Thailand			
Tunisia			
Turkey	–	+	
UK. England. Wales. N. Ireland			
UK. Scotland		–	+
Uruguay			
United States			
Count harsh ("-")	**13**	**8**	**4**
Count lenient ("+")	**9**	**12**	**9**
Count no difference	**45**	**47**	**54**

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



Table 13.11 [Part 1/3]
ICR outcomes by country and domain

		PISA score difference (reported-verifier)				PISA scores		
	Domain	Sign	CI_lo	CI_hi	Agree (%)	Ver	Rep	Adj
Argentina	Mathematics	ns	–2.72	4.55				
	Reading	+	5.06	13.84	97.40	15.90	17.10	15.90
	Science	+	1.16	6.54	94.80	21.40	21.90	21.50
Australia	Mathematics	+	0.88	8.58	97.40	17.50	17.60	17.50
	Reading	ns	–10.92	4.01				
	Science	ns	–3.49	1.97				
Austria	Mathematics	ns	–2.39	3.66				
	Reading	ns	–11.51	0.33				
	Science	–	–4.16	–0.02	95.80	38.30	37.80	38.20
Azerbaijan	Mathematics	+	7.28	13.46	98.10	10.20	11.60	10.60
	Reading	+	10.35	30.28	98.00	13.80	16.20	13.80
	Science	–	–5.88	–0.05	95.40	20.30	19.30	19.80
Belgium (FRE)	Mathematics	ns	–4.23	1.36				
	Reading	–	–20.67	–0.79	95.20	22.20	20.90	22.20
	Science	ns	–1.01	2.94				
Belgium (DUT)	Mathematics	–	–6.90	–0.06	95.20	18.00	17.60	17.70
	Reading	+	11.26	22.34	96.20	21.60	23.40	21.80
	Science	ns	–0.67	2.23				
Bulgaria	Mathematics	ns	–2.38	4.31				
	Reading	+	4.04	19.61	90.60	13.20	14.10	13.20
	Science	+	6.30	12.53	98.50	28.70	30.80	28.60
Brazil	Mathematics	ns	–5.76	1.20				
	Reading	ns	–3.30	10.81				
	Science	ns	–2.60	3.17				
Canada (ENG)	Mathematics	ns	0.99	11.33				
	Reading	ns	–3.78	5.76				
	Science	–	–5.92	1.46	90.40	37.70	37.60	37.80
Canada (FRE)	Mathematics	ns	–9.69	3.39				
	Reading	ns	–13.67	8.35				
	Science	–	–11.48	–0.61	87.50	31.30	30.00	31.00
Chile	Mathematics	–	–9.08	–1.69	94.20	11.30	10.80	10.80
	Reading	+	0.17	9.08	95.00	17.70	18.10	18.30
	Science	ns	–4.13	0.40				
Colombia	Mathematics	ns	–0.13	5.13				
	Reading	ns	–9.03	3.15				
	Science	ns	–2.27	1.94				
Croatia	Mathematics	–	–8.60	–1.16	96.50	13.40	12.70	12.90
	Reading	ns	–0.44	10.70				
	Science	ns	–2.26	1.63				
Czech Republic	Mathematics	ns	–2.27	3.05				
	Reading	+	3.75	15.54	91.10	19.90	20.50	20.20
	Science	+	0.94	7.42	93.30	43.90	44.60	44.30
Denmark	Mathematics	ns	–5.39	1.93				
	Reading	–	–15.45	–3.60	94.00	22.90	21.30	22.70
	Science	ns	–3.17	0.81				
Estonia	Mathematics	ns	–3.68	2.72				
	Reading	+	3.81	17.07	95.20	24.20	25.50	24.60
	Science	+	3.10	11.30	100.00	34.80	36.10	34.80
Estonia (RUS)	Mathematics	ns	–1.25	3.40				
	Reading	–	–11.99	7.61	81.50	21.00	20.80	21.50
	Science	–	–6.71	6.57	92.90	31.50	31.00	31.20
Finland	Mathematics	ns	–1.55	5.26				
	Reading	ns	–11.65	4.97				
	Science	+	1.43	5.71	95.30	42.60	42.90	42.80
France	Mathematics	ns	–6.67	0.55				
	Reading	–	–13.09	–0.43	92.30	23.80	23.40	23.70
	Science	ns	–1.21	3.92				
Germany	Mathematics	ns	–4.64	1.26				
	Reading	ns	–5.67	4.58				
	Science	ns	–4.93	0.72				
Greece	Mathematics	–	–5.58	–0.59	97.30	13.50	13.10	13.10
	Reading	ns	–8.82	0.42				
	Science	+	1.71	5.87	98.00	34.00	35.10	34.30
Hong Kong-China	Mathematics	ns	–2.79	3.36				
	Reading	ns	–5.32	6.82				
	Science	–	–5.64	–0.48	97.50	41.00	40.70	41.00
Hungary	Mathematics	ns	–0.16	6.67				
	Reading	+	3.86	18.76	93.10	21.60	22.50	21.90
	Science	+	1.69	6.19	93.00	37.10	37.50	37.40



Table 13.11 [Part 2/3]
ICR outcomes by country and domain

		PISA score difference (reported-verifier)				PISA scores		
	Domain	Sign	CI_lo	CI_hi	Agree (%)	Ver	Rep	Adj
Iceland	Mathematics	ns	–6.24	0.41				
	Reading	+	2.54	14.49	93.90	19.80	21.00	19.90
	Science	ns	–2.90	0.19				
Indonesia	Mathematics	ns	–0.60	11.63				
	Reading	–	–15.36	–2.25	88.40	12.10	11.60	11.80
	Science	+	2.15	7.16	86.40	21.70	22.50	21.30
Ireland	Mathematics	+	0.33	6.17	97.70	14.20	14.80	14.30
	Reading	+	1.67	11.91	90.70	22.80	23.20	23.00
	Science	ns	–1.98	3.04				
Israel	Mathematics	ns	–2.50	5.53				
	Reading	ns	–8.81	4.54				
	Science	+	0.31	5.17	94.50	35.10	35.60	35.40
Italy	Mathematics	ns	–6.87	0.14				
	Reading	ns	–4.37	3.66				
	Science	+	0.52	5.13	96.10	37.10	37.50	37.80
Jordan	Mathematics	ns	–7.84	2.71				
	Reading	ns	–0.28	9.94				
	Science	+	0.77	6.04	94.30	26.40	27.10	26.60
Japan	Mathematics	ns	–5.52	0.53				
	Reading	+	16.77	30.32	87.00	22.50	24.90	23.30
	Science	ns	–2.84	1.36				
Korea	Mathematics	ns	–3.85	3.27				
	Reading	+	16.33	27.12	90.50	24.00	26.20	24.90
	Science	–	–4.71	–0.78	94.70	37.90	37.70	38.00
Kyrgyzstan (KIR)	Mathematics	+	–1.10	7.76	99.50	4.90	5.10	4.80
	Reading	ns	–1.28	8.39				
	Science	–	–5.59	0.45	96.80	14.10	13.60	13.80
Kyrgyzstan (RUS))	Mathematics	+	–1.15	10.96	100.00	9.70	10.00	9.70
	Reading	ns	–11.09	19.11				
	Science	–	–7.79	2.70	92.90	17.70	17.60	18.00
Latvia (LVA)	Mathematics	–	–14.89	–5.63	94.00	14.40	14.00	14.40
	Reading	+	–5.23	7.51	89.10	23.00	22.70	23.50
	Science	ns	–6.02	0.01				
Latvia (RUS)	Mathematics	+	–3.44	14.04	95.70	15.20	14.60	15.40
	Reading	+	13.30	33.71	92.30	19.00	20.30	19.30
	Science	ns	–5.67	6.52				
Lithuania	Mathematics	ns	–4.13	1.67				
	Reading	–	–9.71	–1.43	92.40	19.20	19.20	19.90
	Science	ns	–5.01	1.04				
Luxembourg	Mathematics	+	1.93	7.85	96.60	13.60	14.30	13.90
	Reading	ns	–8.30	2.03				
	Science	ns	–2.36	1.48				
Macao-China	Mathematics	–	–7.50	–0.57	97.90	15.70	15.30	15.70
	Reading	–	–12.71	–0.22	94.60	20.10	19.60	20.10
	Science	ns	–4.64	1.11				
Mexico	Mathematics	–	–11.54	–3.57	93.20	11.40	10.60	11.10
	Reading	ns	–5.78	7.95				
	Science	–	–12.87	–8.45	87.90	26.90	25.60	26.60
Montenegro	Mathematics	–	–10.47	–1.37	98.70	11.10	10.60	10.90
	Reading	–	–17.56	–1.41	98.10	14.70	13.30	14.50
	Science	ns	–2.02	2.48				
Netherlands	Mathematics	ns	–2.72	6.15				
	Reading	+	0.79	15.65	79.60	21.40	22.20	22.10
	Science	+	1.36	8.22	80.60	38.10	39.20	38.60
New Zealand	Mathematics	ns	–1.45	4.86				
	Reading	ns	–0.01	11.38				
	Science	ns	–3.43	1.86				
Norway	Mathematics	ns	–0.72	4.59				
	Reading	+	17.46	30.65	92.50	19.50	21.20	20.00
	Science	ns	–3.80	0.41				
Poland	Mathematics	ns	–0.05	5.78				
	Reading	ns	–2.21	8.75				
	Science	ns	–3.48	0.91				
Portugal	Mathematics	–	–11.73	–3.65	90.90	15.30	14.30	14.70
	Reading	–	–28.44	–15.39	94.30	21.90	19.50	21.70
	Science	–	–14.93	–8.79	90.30	33.20	31.20	32.90



Table 13.11 [Part 3/3]
ICR outcomes by country and domain

		PISA score difference (reported-verifier)				PISA scores		
	Domain	Sign	CI_lo	CI_hi	Agree (%)	Ver	Rep	Adj
Qatar (ARA)	Mathematics	+	0.54	15.16	98.80	6.00	6.90	6.60
	Reading	+	5.91	14.89	97.40	12.30	12.80	12.40
	Science	–	–5.32	–0.18	98.70	24.90	24.10	24.70
Qatar (ENG)	Mathematics	+	–0.95	15.25	99.20	11.90	12.90	13.20
	Reading	+	–1.83	26.91	97.40	23.60	24.60	23.80
	Science	–	–8.35	2.90	92.30	32.40	31.60	31.80
Romania	Mathematics	–	–6.30	–0.64	98.10	8.90	8.20	8.50
	Reading	ns	–13.55	0.38				
	Science	ns	–5.20	1.21				
Russian Federation	Mathematics	ns	–1.87	4.01				
	Reading	–	–26.37	–15.21	94.20	21.10	19.40	21.10
	Science	ns	–0.61	3.96				
Serbia	Mathematics	–	–10.13	–3.19	95.90	13.60	12.80	13.10
	Reading	ns	–8.39	1.48				
	Science	–	–5.70	–1.33	92.40	30.50	29.40	30.30
Scotland	Mathematics	ns	–3.91	2.76				
	Reading	–	–14.08	–2.32	92.90	22.80	22.40	23.00
	Science	+	0.96	6.87	95.70	39.60	40.30	42.60
Slovak Republic	Mathematics	ns	–4.58	2.25				
	Reading	+	6.02	15.37	91.90	18.50	19.90	19.30
	Science	+	1.49	5.78	94.40	38.60	39.10	38.90
Slovenia	Mathematics	ns	–3.66	3.17				
	Reading	+	2.62	14.24	91.50	20.90	21.10	21.10
	Science	+	2.24	7.16	93.40	39.30	39.90	39.50
Spain (BAQ)	Mathematics	ns	–13.48	3.21				
	Reading	ns	–6.97	19.02				
	Science	ns	–10.01	2.33				
Spain (CAT)	Mathematics	–	–10.18	–2.45	95.50	15.10	14.40	14.80
	Reading	ns	–12.06	0.17				
	Science	ns	–3.83	1.03				
Spain (GLG)	Mathematics	–	–10.45	–1.64	97.10	14.50	13.80	14.20
	Reading	+	–1.15	18.98	85.10	18.00	19.30	18.70
	Science	ns	–6.96	0.41				
Spain (SPA)	Mathematics	–	–4.46	–0.76	97.70	17.00	16.70	16.80
	Reading	ns	–5.50	3.88				
	Science	ns	–0.59	3.07				
Sweden	Mathematics	ns	–4.69	2.08				
	Reading	+	14.05	29.10	91.20	22.10	24.10	22.40
	Science	ns	–0.61	3.82				
Switzerland (FRE)	Mathematics	–	–11.10	6.86	92.20	16.60	16.20	16.80
	Reading	ns	–23.92	2.49		15.00	13.00	15.00
	Science	ns	–2.75	9.74				
Switzerland (GER)	Mathematics	–	–11.45	–2.55	95.90	18.30	17.90	18.00
	Reading	–	–22.28	–4.11	89.90	25.10	23.90	25.20
	Science	ns	–5.04	1.51				
Chinese Taipei	Mathematics	ns	–5.05	1.48				
	Reading	+	2.51	12.48	98.30	23.70	24.50	23.90
	Science	ns	–3.72	0.95				
Thailand	Mathematics	ns	–4.25	0.33				
	Reading	–	–16.39	–5.24	94.20	19.30	18.20	19.30
	Science	ns	–3.69	0.06				
Tunisia	Mathematics	ns	–4.77	0.95				
	Reading	+	5.94	19.20	91.20	12.20	13.00	12.30
	Science	ns	–2.85	1.82				
Turkey	Mathematics	–	–10.53	–2.66	96.10	13.40	12.60	12.80
	Reading	+	9.44	22.44	96.70	18.00	20.20	18.10
	Science	ns	–0.78	5.44				
United Kingdom	Mathematics	ns	–2.58	6.32				
	Reading	+	0.79	10.93	87.60	20.30	20.90	20.80
	Science	ns	–3.84	0.45				
Uruguay	Mathematics	ns	–6.10	1.34				
	Reading	+	2.18	13.74	88.00	19.10	19.80	19.50
	Science	ns	–2.66	2.13				
United States	Mathematics	ns	–0.06	5.79				
	Reading	+	1.33	10.42	89.30	23.90	24.30	24.50
	Science	ns	–0.71	5.55				

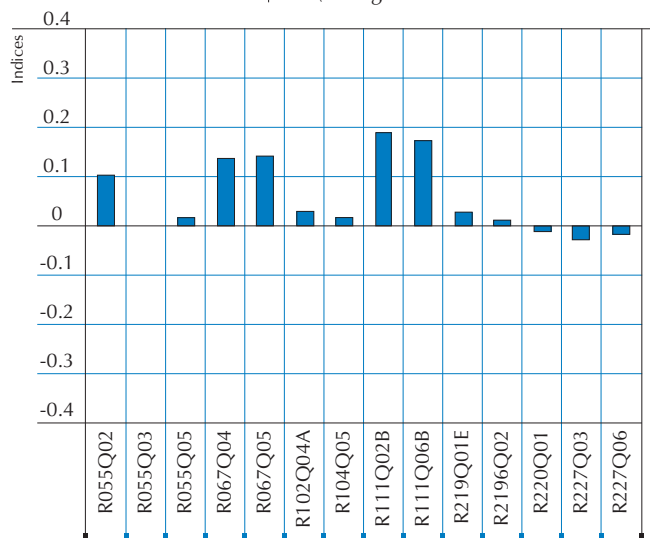


In Table 13.11 the outcomes of the ICR process are summarised for each country and by language group (where appropriate) and domain. In columns 3–5 of that table, information is reported about the estimated bias in the national score for the domain, in PISA score units, based on the difference observed when the score is calculated from national scores, and when calculated using the verifier score. The sign of any difference is reported, with the "+" symbol indicating leniency at the national level, "–" indicating harshness at the national level, and "ns" indicating no significant difference. The 95% confidence interval around the mean difference is reported in the next two columns. The column headed "Agree (%)" displays the estimated level of agreement between the adjudicator and the verifier, calculated according to the formula given earlier. And finally, three estimated PISA scores are given – those based on the codes given by the verifier, the country codes, and the adjudicator codes respectively.

At the conclusion of the ICR, a report was sent to each participant country summarising the outcomes of the international coding review for each test domain. The report contained several elements. One was a graph showing the discrepancies item by item within each domain between the average raw score based on codes given by the four national coders, and the raw score from the verifier's code, hence providing a fine-grained report at the item level of average discrepancies of national coders relative to an independent benchmark. The report also showed the number and the percentage of individual student responses that had been flagged in Part 2 of the ICR analysis. Finally, the report showed whether there was statistical evidence of bias in national coding, and the estimate of the extent of the bias in PISA score units. National centres were therefore given information that they could use to review their coding operation, and to inform planning for the recruitment and training of coders for future surveys.

An example of an ICR country report is provided in Figure 13.7. Looking at this example, the graph indicates a marked positive average difference between the mean of the four national coders' scores and the verifier score for five of the 14 reading items. Differences for the other nine reading items were much smaller, or non-existent. This provides evidence of leniency in the standards applied by coders in this country in the coding of five of the reading items. This information may be useful input to the coder training for the next PISA survey cycle.

Figure 13.7
Example of ICR report (reading)



Total number of flagged responses	**56**
Total per cent of flagged responses	**6.67%**
Confidence interval for estimated difference in PISA score (reported – verified)	**[5.06 , 13.84]**

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



To the right of this graph, the total number and percentage of flagged responses are given for this domain. In this example, 56 of the 840 reading item responses that were included in the ICR study from this country were flagged. That is, for about 6% of the student responses reviewed, differences were observed between the coding standards applied by the national coders and those applied by the international verifier.

The final element of the report is the estimated bias in the average reading score for this country expressed as a range of values, in PISA score units. The values are the 95% confidence interval about the mean estimate. This information is reported only in cases where the final adjudication process confirms the differences found by the international verifier.

The difference is calculated between the country's reported average reading score, and the score that would be calculated had the codes awarded by the international verifier been used in the scaling, but based only on the reading items in the test booklet used in the ICR. For this country, the degree of leniency estimated lies between about 5 and 14 points on the PISA reading scale.

Cautions

In interpreting the results of the international coder review, it should be borne in mind that the study gives only an indication of possible bias in national results.

First, only some of the manually coded items in each domain were included in the ICR, and the items selected for inclusion were not intended as a random sample of all manually coded items. The selection was made largely on practical and logistical grounds designed to minimise work for participating countries, namely, what was a selection of a small number of booklets that contained as many suitable items as possible. The behaviour of national coders on these items may not be an accurate representation of their behaviour in coding all items.

Related to this, the estimation of the magnitude of observed bias uses mean national ability estimates that are based only on one booklet for each domain, whereas reported PISA outcomes are based on a rotated design involving all 13 booklets. It is well known that positioning of items within test booklets has an impact on the calculation of item difficulty estimates, and therefore also student ability estimates. This further exacerbates the potential unreliability of the bias estimates.



14

Data Adjudication

Introduction 272
- Implementing the standards – quality assurance 272
- Information available for adjudication 273
- Data adjudication process 273

General outcomes 274
- Overview of response rate issues 274
- Detailed country comments 275

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



INTRODUCTION

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

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

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

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

Implementing the standards – quality assurance

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



Information available for adjudication

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

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

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

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

Data adjudication process

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

Step 1: Quality control and quality monitoring data were collected throughout the survey cycle.

Step 2: Data collected from both quality control and quality monitoring activities were entered into a single quality assurance database.

Step 3: Experts compiled country-by-country reports that contained quality assurance data for key areas of project implementation.

Step 4: Experts considered the quality assurance data that were collected from both the quality control and quality monitoring activities, to make a judgement. In this phase the experts collaborated with the project director and other consortium staff to address any identified areas of concern. Where necessary, the relevant NPM was contacted through the project director. At the end of this phase experts constructed, for each adjudicated dataset, a summary detailing how the PISA technical standards had been met.

Step 5: The consortium and the Technical Advisory Group reviewed the reports and made a determination with regard to the quality of the data.

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

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



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

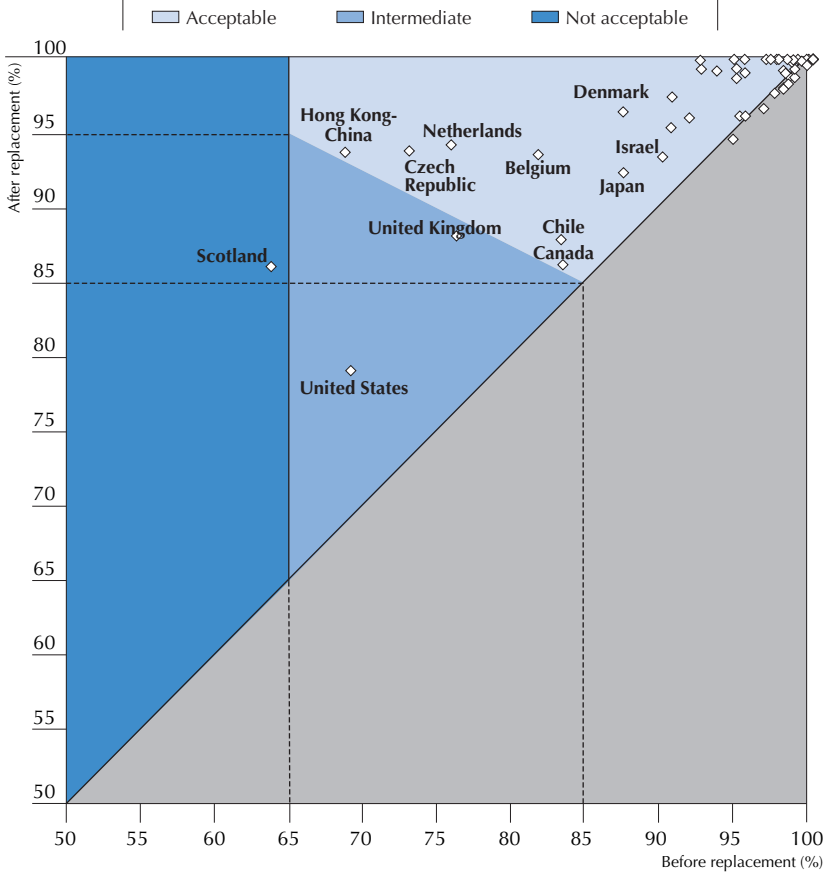
Throughout PISA 2006 the consortium concentrated its quality control activities to ensure that the highest scientific standards were met. However during data adjudication a wider definition of quality was used especially when considering data that were at risk. In particular the underlying criterion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

GENERAL OUTCOMES

Overview of response rate issues

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

Figure 14.1
Attained school response rates



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



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

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

Detailed country comments

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

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

Departures from standard procedures in the national data collection plan

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

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

Departures from standards arising from implementation

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



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

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

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

Argentina

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

Azerbaijan

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

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



Belgium

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

- ***Flanders***

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

Brazil

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

Canada

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

Chile

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

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

Colombia

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

Czech Republic

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

Denmark

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

Estonia

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

Finland

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



France

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

Hungary

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

Iceland

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

Ireland

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

Israel

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

Italy

- ***Provincia Sicilia***
  With a sample size of 1335, Sicilia had fewer than the required 1500 assessed students.
- ***Provincia Sardegna***
  With a sample size of 1390, Sardegna had fewer than the required 1500 assessed students.
- ***Provincia Campania***
  With a sample size of 1406, Campania had fewer than the required 1500 assessed students.
- ***Provincia Lombardia***
  Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were at 2.12%.

Japan

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

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



Latvia

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

Luxembourg

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

Macao-China

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

Montenegro

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

New Zealand

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

Norway

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

Qatar

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

Slovak Republic

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

Spain

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

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

- ***Asturias***

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



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

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

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

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

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

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

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

Sweden

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

Chinese Taipei

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

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

Tunisia

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



United Kingdom

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

- ***Scotland***

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

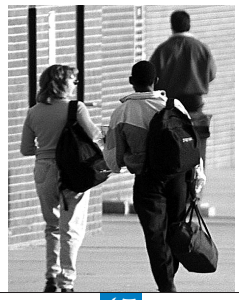
  The school response rate was 63.6% before replacement and therefore was in the not acceptable range. The national centre provided a detailed analysis of school non-response bias, which indicated no evidence of substantial bias resulting from school non-response. The response rate after replacement was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

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

United States

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

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



15

Proficiency Scale Construction

Introduction 284
Development of the described scales 285
- Stage 1: Identifying possible scales 285
- Stage 2: Assigning items to scales 286
- Stage 3: Skills audit 286
- Stage 4: Analysing field trial data 286
- Stage 5: Defining the dimensions 287
- Stage 6: Revising and refining with main study data 287
- Stage 7: Validating 287

Defining proficiency levels 287
Reporting the results for PISA science 290
- Building an item map 290
- Levels of scientific literacy 292
- Interpreting the scientific literacy levels 299

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



INTRODUCTION

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.

The relative ability of students taking a particular test can be estimated by considering the proportion of test items they get correct. The relative difficulty of items in a test can be estimated by considering the proportion of test takers getting each item correct. 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 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 literacy variable they demonstrate, and it is possible to estimate the location of individual test items, thereby seeing how much of the literacy variable each item embodies. This continuum is referred to as the overall PISA literacy scale in the relevant test domain of reading, mathematics or science.

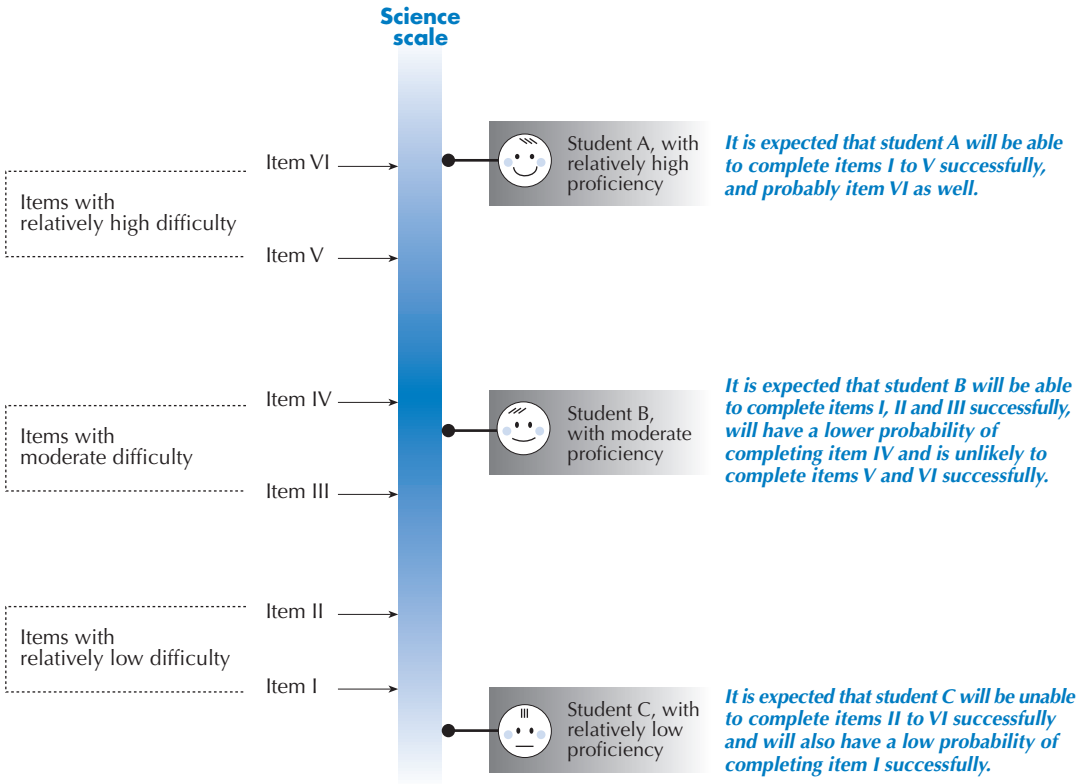
PISA assesses students, and uses the outcomes of that assessment to produce estimates of students' proficiency in relation to a number of literacy variables. These variables are defined in the relevant PISA literacy framework (OECD, 2006). For each of these literacy variables, one or more scales are defined, which stretch from very low levels of literacy through to very high levels. When thinking about what such a scale means in terms of student proficiency, it can be observed that a student 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.

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 a method for 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 survey. Essentially the same methodology has again been used to develop proficiency descriptions for PISA 2006. Given the volume and breadth of data that were available from the PISA 2006 assessment, development of more detailed descriptions of scientific literacy became possible. The detailed proficiency descriptions that had been developed for the reading domain in PISA 2000 were used again with the reduced data available from PISA 2003 and 2006. The detailed descriptions used for mathematics in 2003 were used again in 2006.



Figure 15.1
The relationship between items and students on a proficiency scale



The SEG worked with the consortium to develop sets of described proficiency scales for PISA science. Consultations regarding these described scales with the PGB, the science forum, NPMs and the PISA TAG took place over several stages before their final adoption by the PGB.

This chapter discusses the methodology used to develop those scales and to describe a number of levels of proficiency in the different PISA literacy variables, and presents the outcomes of that development process.

DEVELOPMENT OF THE DESCRIBED SCALES

The development of described proficiency scales for PISA was 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. 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 a scale for *retrieving information* a second scale for *interpreting texts* and a third for *reflection and evaluation*[1].

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



In the case of mathematics, a single proficiency scale was developed for PISA 2000, but with the additional data available in the 2003 survey cycle, when mathematics was the major test domain, the possibility of reporting according to the four overarching ideas or the three competency clusters described in the PISA mathematics framework were both considered.

For science, given the small number of items in PISA 2000 and 2003, a single overall proficiency scale was developed to report results. However, 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 in 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 phyics, chemistry, biology, Earth and space science, and technology; and 'knowledge about science' involving the central processes underpinning in 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 primary database were the three competency based scales: *identifying scientific issues*, *explaining phenomena scientifically* and *using scientific evidence*.

Wherever multiple scales 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. Because of 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.

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 scales under consideration. Science experts (including members of the expert group, the test developers and consortium staff) judged the characteristics of each test item against the relevant framework categories. Later, statistical analysis of item scores from the field trial was used to obtain a more objective measure of fit of each item to its assigned scale.

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 sub-scale 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 size 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 study data

When the main study 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 by the working groups, and other interested parties. The preliminary descriptions and levels were then reviewed and revised in the light of further technical information that was provided by the TAG, and the approach to defining levels and associating students with those levels that had been used in the reporting of PISA 2000 and PISA 2003 results was applied.

Stage 7: Validating

Two major approaches to validation were then considered by the science working groups. One method was to provide knowledgeable experts (*e.g.* teachers, or members of the subject matter expert groups) with material that enabled them to judge PISA items against the described levels, or against a set of indicators that underpinned the described levels. Second, the described scales were subjected to an extensive consultation process involving all PISA countries through their NPMs. This approach to validation rests on the extent to which users of the described scales find them informative.

DEFINING 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 – there is some probability that a particular student can correctly do 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 in order to locate a student on the same scale on which the items are laid out. When placing a student at a particular point on the scale, what probability of success should we insist on 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 with 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, an approach to defining performance levels, and to associating students with those levels, is needed. The methodology that was developed and used for PISA 2000 and 2003 was essentially retained for PISA 2006.



Defining proficiency levels for PISA 2000 progressed in two broad phases. The first, which came after the development of the described scales, was based on a substantive analysis of PISA items in relation to the aspects of literacy that underpinned each test domain. This produced descriptions of increasing proficiency that reflected observations of student performance and a detailed analysis of the cognitive demands of PISA items. The second phase involved decisions about where to set cut-off points for levels and how to associate students with each level. 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.

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:

- A common understanding of the meaning of levels should be developed and promoted. 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 to solve tasks at that level than to fail them. By implication, it must be expected that they would get at least half of the items correct on a test composed of items uniformly spread across that level, which 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 much higher success rate. At the top end of the bandwidth of a level would be the students who are masters of that level. These students would be likely to solve about 80% 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. 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. 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.
- A way of implementing these principles was developed for PISA 2000 and used again in PISA 2003 and 2006. This method 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").

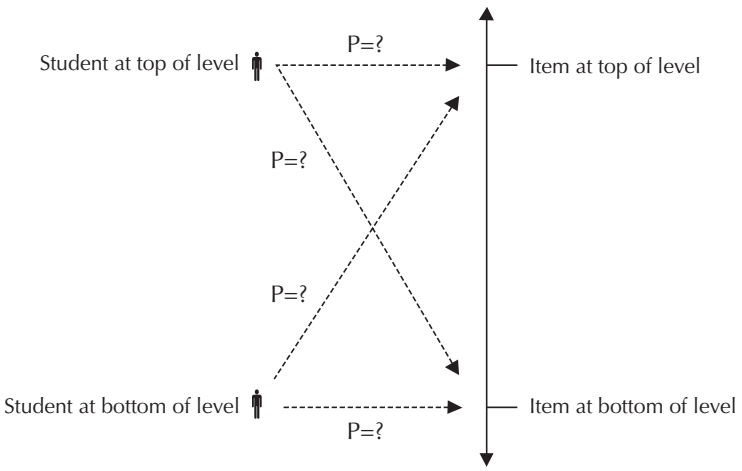
Figure 15.2 summarises the relationship among these three mathematically linked variables. It shows a vertical line representing a part of the scale being defined, one of the bounded levels on the scale, a student at both the top and the bottom of the level, and reference to an item at the top and an item at the bottom of the level. Dotted lines connecting the students and items are labelled "P=?" to indicate the probability associated with that student correctly responding to that item.

PISA 2000 implemented the following solution: start with the substantively determined range of 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 – 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 at that level).

After doing this, the exact average percentage of correct answers on a test composed of items at a level could vary slightly among the different domains, but will always be at least 50% at the bottom of the level except for the lowest described level.

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.

Figure 15.2
What it means to be at a level



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



As levels 2, 3, 4 and 5 (within a domain) will be equally broad, it was proposed that the floor of the lowest described level be placed at this breadth below the upper boundary of level 1 (that is, the cut-off between levels 1 and 2). Student performance below this level is lower than that which PISA can reliably assess and, more importantly, describe.

REPORTING THE RESULTS FOR PISA SCIENCE

In this section, the way in which levels of scientific literacy are defined, described and reported will be discussed. Levels of performance on the PISA scientific literacy scale will be established and described, and they will be exemplified using a number of items from the PISA 2006 assessment.

Building an item map

The data from the PISA science assessment were processed to generate a set of item difficulty measures for the 103 items included in the assessment. 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, a total of 109 item difficulty estimates were generated.

During the process of item development, experts undertook a qualitative analysis of each item, and developed descriptions of aspects of the cognitive demands of each item (and each individual item step in the case of partial credit items that were scored polytomously). This analysis included judgements about the aspects of the PISA science framework that were relevant to each item. For example, each item was analysed to determine which competency and type of knowledge (of or about) was involved in a correct response. Similarly, the situation (context) in which the stimulus and question were located, and to which the competencies, knowledge and attitudes were related, was identified. This included identifying whether the science involved was of personal, social or global interest. As well as these broad categorisations, a short description was developed that attempted to capture the most important demands placed on students by each particular item.

Following data analysis and the resultant generation of difficulty estimates for each of the 109 items steps, 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 2006 test. 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, 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. Many of those factors reflect variables that are central to constructs used in the science framework's discussion of scientific literacy. Patterns emerge that make it possible to describe aspects of scientific literacy that are consistently associated with various locations along the difficulty continuum shown by the map. For example, at a very general level it can be seen that the easiest items are predominately in the *explaining phenomena scientifically* competence and lie in the *knowledge of science* area. These items are similar in that they require little interpretation, the recall of relatively straight forward factual knowledge, and the application of that knowledge in simple familiar situations. This pattern in not repeated above the mid-point of level two (defined in Table 15.1) in the sense that a specific competence is dominant. Above this level the distribution of competencies and knowledge areas is more even. This observation applies equally well to the full set of science items. However, it is possible to see growth in a number of dimensions as student student ability increases on the scientific literacy scale.



Figure 15.3
A map for selected science items

Code	Item name	Item difficulty on PISA scale	Item demands	Competency			Knowledge						Focus		
							of				about				
				Identifying scientific issues	*Explaining phenomena scientifically*	*Using scientific evidence*	*Physical systems*	*Living systems*	*Earth and space systems*	*Technology systems*	*Scientific enquiry*	*Scientific explanations*	Personal	Social	Global
S485Q05(2)	**ACID RAIN**	717	The reason for a control in an investigation is understood and explicitly recognised. An ability to understand the modelling in the investigation is a pre-requisite.	•							•		•		
S114Q05	**GREENHOUSE**	709	There is a pre-requisite to understand the need to control variables. Knowledge of factors contributing to the greenhouse effect is then applied in determining a variable to be controlled.		•				•						•
S114Q04(2)	**GREENHOUSE**	659	Given a conclusion can compare two graphs and locate corresponding areas that are at odds with that conclusion and accurately describe that difference.			•						•			•
S447Q05	**SUNSCREENS**	616	Correctly interprets a dataset expressed diagrammatically and provides an explanation that summarises the data.			•						•	•		
S447Q02	**SUNSCREENS**	588	The control 'aspects' of an investigation are recognised.	•							•		•		
S493Q05	**PHYSICAL EXERCISE**	583	Recognition that increased exercise results in increased respiration and thus the need for more oxygen and/or removal of more carbon dioxide.		•			•					•		
S114Q04(1)	**GREENHOUSE**	568	Recognises differences in two graphs relating to a phenomenon but cannot provide a clear explanation as to why the differences are at odds with a given conclusion.			•						•			•
S213Q01	**CLOTHES**	567	Can apply knowledge of the features of a scientific investigation to decisions about whether specific issues are scientifically investigatable.	•							•			•	
S493Q01	**PHYSICAL EXERCISE**	545	Can identify some features of physical exercise that are advantageous to health – cardiovascular system, bodyweight.		•			•					•		
S114Q03	**GREENHOUSE**	529	Shows an understanding of what two graphs relating to a phenomenon are depicting and can compare them for similarities.			•						•			•
S485Q05(1)	**ACID RAIN**	513	Recognises that a comparison is being made between two tests but is unable to articulate the purpose of the control.	•							•		•		
S477Q04	**MARY MONTAGU**	507	Recognises that the immune systems of young and old people are less resistant to viruses than those of the general population.		•			•						•	
S447Q03	**SUNSCREENS**	499	Can recognise the change and measured variables from a description of an investigation and as a consequence identify the question motivating the investigation.	•							•		•		
S426Q07	**GRAND CANYON**	485	Can recognise issues in which scientific measurement can be applied to answwering a question.	•							•			•	
S485Q03	**ACID RAIN**	460	Recognises that the loss of gas in a chemical reaction results in a reduction of mass for the products left behind.			•	•						•		
S426Q03	**GRAND CANYON**	451	Applies knowledge that water increases in volume as it changes from liquid to solid.		•				•					•	
S477Q03	**MARY MONTAGU**	431	Recalls knowledge of the role of antibodies in immunity.		•			•						•	
S508Q03	**GENETICALLY MODIFIED CROPS**	421	Understands that a fair test involves finding out if an outcome is affected by a range of extraneous conditions.	•							•			•	
S213Q02	**CLOTHES**	399	Can select the correct apparatus to measure an electric current.		•					•			•		
S493Q03	**PHYSICAL EXERCISE**	386	Rejects the notion that fats are formed in the muscles and knows that the rate of flow of blood increases during exercise.		•			•					•		

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



Based on the patterns observed when the full question set is reviewed against the three proficiency scales, it is possible to characterise the increase in the levels of complexity of competencies measured. This can be done by referring to the ways in which science competencies are associated with questions located at different points ranging from the bottom to the top of the scale. The ascending difficulty of science questions in PISA 2006 is associated with the following characteristics, which require all three competencies but which shift in emphasis as students progress from the identification of issues to the use of evidence to communicate an answer, decision or solution.

The degree to which the transfer and application of knowledge is required: At the lowest levels the application of knowledge is simple and direct. The requirement can often be fulfilled with simple recall of single facts. At higher levels of the scale, individuals are required to identify multiple fundamental concepts and combine categories of knowledge in order to respond correctly.

The degree of cognitive demand required to analyse the presented situation and synthesise an appropriate answer: This centres on features such as the depth of scientific understanding required, the range of scientific understandings required and the proximity of the situation to the students' life. At the highest level this is characterised by in-depth understanding, an ability to apply a range of scientific understandings and to apply these in broad or global contexts.

The degree of analysis needed to answer the question: This includes the demands arising from the requirement to discriminate among issues presented in the situation under analysis, identify the appropriate knowledge domain *(Knowledge of science* and *Knowledge about science),* and use appropriate evidence for claims or conclusions. The analysis may include the extent to which the scientific or technological demands of the situation are clearly apparent or to which students must differentiate among components of the situation to clarify the scientific issues as opposed to other, non-scientific issues.

The degree of complexity needed to solve the problem presented: The complexity may range from a single step where students identify the scientific issue, apply a single fact or concept, and present a conclusion to multi-step problems requiring a search for advanced scientific knowledge, complex decision making, information processing and ability to form an argument.

The degree of synthesis needed to answer the question: The synthesis may range from a single piece of evidence where no real construction of justification or argument is required to situations requiring students to apply multiple sources of evidence and compare competing lines of evidence and different explanations to adequately argue a position.

Levels of scientific literacy

The approach to reporting used by the OECD has been defined in previous cycles of PISA and is based on the definition of a number of bands or 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 2006 outcomes for science.

For PISA science, student scores have been transformed to the PISA scale, with a mean of 500 and a standard deviation of 100, and six levels of proficiency have been defined and described. The continuum of increasing scientific literacy that is represented in Figure 15.4 has been divided into five bands, each of equal width, and two unbounded regions, one at each end of the continuum. The band definitions on the PISA scale are given in Table 15.1.



The information about the items in each band has been used to develop summary descriptions of the kinds of scientific competencies associated with different levels of proficiency. These summary descriptions can then be used to encapsulate typical scientific proficiency of students associated with each level. As a set, the descriptions encapsulate a representation of growth in scientific literacy.

To develop the summary descriptions, growth in scientific competence was first considered separately in relation to items from each of the competencies. Three sets of descriptions were developed. These are presented in following sections, in Figure 15.5. The three sets of descriptions were combined to produce meta-descriptions of six levels of overall scientific literacy, presented here in Figure 15.4.

Table 15.1
Scientific literacy performance band definitions on the PISA scale

Level	Score points on the PISA scale
6	Above 707.9
5	633.3 to 707.9
4	558.7 to 633.3
3	484.1 to 558.7
2	409.5 to 484.1
1	334.9 to 409.5

A clear progression through these levels is apparent in the way in which the individual scientific competencies specified in the PISA scientific literacy framework play out as literacy levels increase.

For example, the competency *identifying scientific issues* is observed to follow a progression through two related dimensions described in combination in Figure 15.5. These dimensions are:

Understanding the methodology of science: At the lowest level students can usually identify variables that are not open to scientific measurement but can do little more than that. Around the middle of the range, levels 3 and 4, there is identification of the independent and dependent variables in an investigation and a developing understanding of both the reason for a control (referent) and the need to account for extraneous variables. Examples of this can be found in SUNSCREENS Q02 and ACID RAIN Q05(1)[2]. As demonstrated by SUNSCREENS Q03, students at these levels can usually identify the question motivating the investigation. At higher levels students are able to view an investigation in its totality and show an awareness of the range of issues that need to be accounted for if meaning is to be ascribed to the outcomes of a testing regime.

Designing an investigation: At the lowest levels students are able to ask questions that elicit relevant information about straightforward scientific issues within familiar contexts. They are able to suggest comparisons to be made given simple cause and effect relationships. Around the middle of the literacy scale students show the capacity to produce simple designs to investigate direct or concrete relationships that are set in relatively familiar contexts. They exhibit an awareness of features that they need to control or account for in their designs. At the highest levels students are able to design ways of investigating questions that involve abstract ideas within the scope of their conceptual knowledge.

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



Figure 15.4
Summary descriptions of the six proficiency levels on the science scale

Level	Lower score limit	Percentage of students able to perform tasks at each level or above (OECD average)	What students can typically do
6	707.9	1.3% of students across the OECD can perform tasks at Level 6 on the science scale	At Level 6, students can consistently identify, explain and apply scientific knowledge and *knowledge about science* in a variety of complex life situations. They can link different information sources and explanations and use evidence from those sources to justify decisions. They clearly and consistently demonstrate advanced scientific thinking and reasoning, and they demonstrate willingness to use their scientific understanding in support of solutions to unfamiliar scientific and technological situations. Students at this level can use scientific knowledge and develop arguments in support of recommendations and decisions that centre on personal, social or global situations.
5	633.3	9.0% of students across the OECD can perform tasks at least at Level 5 on the science scale	At Level 5, students can identify the scientific components of many complex life situations, apply both scientific concepts and *knowledge about science* to these situations, and can compare, select and evaluate appropriate scientific evidence for responding to life situations. Students at this level can use well-developed inquiry abilities, link knowledge appropriately and bring critical insights to situations. They can construct explanations based on evidence and arguments based on their critical analysis.
4	558.7	29.3% of students across the OECD can perform tasks at least at Level 4 on the science scale	At Level 4, students can work effectively with situations and issues that may involve explicit phenomena requiring them to make inferences about the role of science or technology. They can select and integrate explanations from different disciplines of science or technology and link those explanations directly to aspects of life situations. Students at this level can reflect on their actions and they can communicate decisions using scientific knowledge and evidence.
3	484.1	56.7% of students across the OECD can perform tasks at least at Level 3 on the science scale	At Level 3, students can identify clearly described scientific issues in a range of contexts. They can select facts and knowledge to explain phenomena and apply simple models or inquiry strategies. Students at this level can interpret and use scientific concepts from different disciplines and can apply them directly. They can develop short statements using facts and make decisions based on scientific knowledge.
2	409.5	80.8% of students across the OECD can perform tasks at least at Level 2 on the science scale	At Level 2, students have adequate scientific knowledge to provide possible explanations in familiar contexts or draw conclusions based on simple investigations. They are capable of direct reasoning and making literal interpretations of the results of scientific inquiry or technological problem solving.
1	334.9	94.8% of students across the OECD can perform tasks at least at Level 1 on the science scale	At Level 1, students have such a limited scientific knowledge that it can only be applied to a few, familiar situations. They can present scientific explanations that are obvious and that follow explicitly from given evidence.



Figure 15.5 [Part 1/2]

Summary descriptions of the six proficiency levels in *identifying scientific issues*

General proficiencies students should have at each level	Tasks a student should be able to do	Examples from released questions
LEVEL 6 1.3% of all students across the OECD area can perform tasks at Level 6 on the *identifying scientific issues* scale.		
Students at this level demonstrate an ability to understand and articulate the complex modelling inherent in the design of an investigation.	▪ Articulate the aspects of a given experimental design that meet the intent of the scientific question being addressed. ▪ Design an investigation to adequately meet the demands of a specific scientific question. ▪ Identify variables that need to be controlled in an investigation and articulate methods to achieve that control.	**ACID RAIN** Question 5
LEVEL 5 8.4% of all students across the OECD area can perform tasks at least at Level 5 on the *identifying scientific issues* scale.		
Students at this level understand the essential elements of a scientific investigation and thus can determine if scientific methods can be applied in a variety of quite complex, and often abstract contexts. Alternatively, by analysing a given experiment can identify the question being investigated and explain how the methodology relates to that question.	▪ Identify the variables to be changed and measured in an investigation of a wide variety of contexts. ▪ Understand the need to control all variables extraneous to an investigation but impinging on it. ▪ Ask a scientific question relevant to a given issue.	
LEVEL 4 28.4% of all students across the OECD area can perform tasks at least at Level 4 on the *identifying scientific issues* scale.		
Students at this level can identify the change and measured variables in an investigation and at least one variable that is being controlled. They can suggest appropriate ways of controlling that variable. The question being investigated in straightforward investigations can be articulated.	▪ Distinguish the control against which experimental results are to be compared. ▪ Design investigations in which the elements involve straightforward relationships and lack appreciable abstractness. ▪ Show an awareness of the effects of uncontrolled variables and attempt to take this into account in investigations.	**SUNSCREENS** Questions 2 and 4 **CLOTHES** Question 1

• • •

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



Figure 15.5 [Part 2/2]

Summary descriptions of the six proficiency levels in *identifying scientific issues*

General proficiencies students should have at each level	Tasks a student should be able to do	Examples from released questions
LEVEL 3 56.7% of all students across the OECD area can perform tasks at least at Level 3 on the *identifying scientific issues* scale.		
Students at this level are able to make judgements about whether an issue is open to scientific measurement and, consequently, to scientific investigation. Given a description of an investigation can identify the change and measured variables.	▪ Identify the quantities able to be scientifically measured in an investigation. ▪ Distinguish between the change and measured variables in simple experiments. ▪ Recognise when comparisons are being made between two tests (but are unable to articulate the purpose of a control).	**ACID RAIN** Question 5 (Partial) **SUNSCREENS** Question 3
LEVEL 2 81.3% of all students across the OECD area can perform tasks at least at level 2 on the *identifying scientific issues* scale.		
Students at this level can determine if scientific measurement can be applied to a given variable in an investigation. They can recognise the variable being manipulated (changed) by the investigator. Students can appreciate the relationship between a simple model and the phenomenon it is modelling. In researching topics students can select appropriate key words for a search.	▪ Identify a relevant feature being modelled in an investigation. ▪ Show an understanding of what can and cannot be measured by scientific instruments. ▪ Select the most appropriate stated aims for an experiment from a given selection. ▪ Recognise what is being changed (the cause) in an experiment. ▪ Select a best set of Internet search words on a topic from several given sets.	**GENETICALLY MODIFIED CROPS** Question 3
LEVEL 1 94.9% of all students across the OECD area can perform tasks at least at Level 1 on the *identifying scientific issues* scale.		
Students at this level can suggest appropriate sources of information on scientific topics. They can identify a quantity that is undergoing variation in an experiment. In specific contexts they can recognise whether that variable can be measured using familiar measuring tools or not.	▪ Select some appropriate sources from a given number of sources of potential information on a scientific topic. ▪ Identify a quantity that is undergoing change, given a specific but simple scenario. ▪ Recognise when a device can be used to measure a variable (within the scope of the student's familiarity with measuring devices).	

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



Progression in the *explaining phenomena scientifically* competency can be seen along three dimensions. Descriptions applicable to the various levels can be found in Figure 15.6.

Breadth and depth of scientific knowledge: At the lowest levels students can recall singular scientific facts either learned in a school environment or experienced in daily life in giving simple explanations. Examples of this can be found in CLOTHES Q02 and MARY MONTAGUE Q02. Around the middle of the scale students are able to apply several related pieces of information to an explanation of a phenomenon. In MARY MONTAGUE Q04 students were required to bring knowledge of vaccination, immunity systems and differential resistance in human populations to the question. The knowledge utilised is distinguishable from that of lower levels of literacy by its breadth and the inclusion of an abstract concept where applicable. At the highest levels students can draw upon a broad range of abstract scientific concepts in developing explanations of a phenomenon such as in GREENHOUSE Q05.

Figure 15.6 [Part 1/2]

Summary descriptions of the six proficiency levels in *explaining phenomena scientifically*

General proficiencies students should have at each level	Tasks a student should be able to do	Examples from released questions
LEVEL 6 1.8% of all students across the OECD area can perform tasks at Level 6 on the *explaining phenomena scientifically* scale.		
Students at this level draw on a range of abstract scientific knowledge and concepts and the relationships between these in developing explanations of processes within systems.	▪ Demonstrate an understanding of a variety of complex, abstract physical, biological or environmental systems. ▪ In explaining processes, articulate the relationships between a number of discrete elements or concepts.	**GREENHOUSE** Question 5
LEVEL 5 9.8% of all students across the OECD area can perform tasks at least at Level 5 on the *explaining phenomena scientifically* scale.		
Students at this level draw on knowledge of two or three scientific concepts and identify the relationship between them in developing an explanation of a contextual phenomenon.	▪ Take a scenario, identify its major component features, whether conceptual or factual, and use the relationships between these features in providing an explanation of a phenomenon. ▪ Synthesise two or three central scientific ideas in a given context in developing an explanation for, or a prediction of, an outcome.	
LEVEL 4 29.4% of all students across the OECD area can perform tasks at least at Level 4 on the *explaining phenomena scientifically* scale.		
Students at this level have an understanding of scientific ideas, including scientific models, with a significant level of abstraction. They can apply a general, scientific concept containing such ideas in the development of an explanation of a phenomenon.	▪ Understand a number of abstract scientific models and can select an appropriate one from which to draw inferences in explaining a phenomenon in a specific context (*e.g.* the particle model, planetary models, models of biological systems). ▪ Link two or more pieces of specific knowledge, including from an abstract source in an explanation (*e.g.* increased exercise leads to increased metabolism in muscle cells, this in turn requires an increased exchange of gases in the blood supply which is achieved by an increased rate of breathing).	**PHYSICAL EXERCISE** Question 5

• • •

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



Figure 15.6 [Part 2/2]

Summary descriptions of the six proficiency levels in *explaining phenomena scientifically*

General proficiencies students should have at each level	Tasks a student should be able to do	Examples from released questions
LEVEL 3 56.4% of all students across the OECD area can perform tasks at least at Level 3 on the *explaining phenomena scientifically* scale.		
Students at this level can apply one or more concrete or tangible scientific ideas/concepts in the development of an explanation of a phenomenon. This is enhanced when there are specific cues given or options available from which to choose. When developing an explanation, cause and effect relationships are recognised and simple, explicit scientific models may be drawn upon.	▪ Understand the central feature(s) of a scientific system and, in concrete terms, can predict outcomes from changes in that system (*e.g.* the effect of a weakening of the immune system in a human). ▪ In a simple and clearly defined context, recall several relevant, tangible facts and apply these in developing an explanation of the phenomenon.	**MARY MONTAGU** Question 4 **ACID RAIN** Question 2 **PHYSICAL EXERCISE** Question 1
LEVEL 2 80.4% of all students across the OECD area can perform tasks at least at Level 2 on the *explaining phenomena scientifically* scale.		
Students at this level can recall an appropriate, tangible, scientific fact applicable in a simple and straightforward context and can use it to explain or predict an outcome.	▪ Given a specific outcome in a simple context, indicate, in a number of cases and with appropriate cues the scientific fact or process that has caused that outcome (*e.g.* water expands when it freezes and opens cracks in rocks, land containing marine fossils was once under the sea). ▪ Recall specific scientific facts with general currency in the public domain (*e.g.* vaccination provides protection against viruses that cause disease).	**GRAND CANYON** Question 3 **MARY MONTAGU** Questions 2 and 3 **GRAND CANYON** Question 5
LEVEL 1 94.6% of all students across the OECD area can perform tasks at least at Level 1 on the *explaining phenomena scientifically* scale.		
Students at this level can recognise simple cause and effect relationships given relevant cues. The knowledge drawn upon is a singular scientific fact that is drawn from experience or has widespread popular currency.	▪ Choose a suitable response from among several responses, given the context is a simple one and that recall of a single scientific fact is involved (*e.g.* ammeters are used to measure electric current). ▪ Given sufficient cues, recognise simple cause and effect relationships (*e.g.* Do muscles get an increased flow of blood during exercise? Yes or No).	**PHYSICAL EXERCISE** Question 3 **CLOTHES** Question 2

System complexity: At the lowest levels students are able to deal with very simple contexts involving cause and effect relationships as illustrated by PHYSICAL EXERCISE Q03 where the effect of increased exercise is an increase in the flow of blood. Those in the middle ranges of the scientific literacy scale are beginning to view phenomena from a system viewpoint, increasingly extending and recognising the relationships that bear on the phenomenon. The models they understand and use in developing explanations start to deal with abstract scientific ideas and a degree of complexity. PHYSICAL EXERCISE Q05 involves students in drawing on knowledge of the human respiratory system. At the highest levels students show a capacity to develop explanations for contexts with a high degree of complexity involving abstract ideas and sub-systems drawn from a variety of scientific disciplines.

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



Synthesis: The ability to bring together relevant concepts and to understand the relationships between them in constructing an explanation of a phenomenon is another dimension that shows progression over the levels of scientific literacy. At level two this is demonstrated in ACID RAIN Q03 where the requirement was to recognise that water would turn to ice below 0°C, know that water expands as it turns to ice and that there is a relationship between that expansion and the breaking down of rock. This involves the synthesis of concrete facts and ideas. By level 4 a progression in this dimension can be seen in PHYSICAL EXERCISE Q05. There, the requirement is to bring into relationship abstract ideas about the need of the muscles of the body for an increased rate of gaseous exchange in the lungs during physical exercise.

Progression in two dimensions is evident in the *using scientific evidence* competency. Descriptions relating to this progression in scientific literacy in this area can be found in Figure 15.5.

Complexity of the data used: At a low level of literacy the student can make comparisons between rows in a simple table or make a conclusion from a simple change in a single variable. A level 2 example of this can be found in GRAND CANYON Q03 where the requirement was to recognise that the loss of gas in a chemical reaction results in a loss of mass for the products left behind. Around the middle of the literacy scale students can utilise data presented in the form of line graphs in making inferences, make simple comparisons between graphs and describe patterns in increasingly complex tables of data. Examples of this can be found in GREENHOUSE Q03 and Q04(1). At higher levels students are able to describe patterns in complex data, summarise that data and suggest explanations for the patterns.

Comparative skills and critical abilities applied to conclusions: Given options or clues, students at the lower levels of this competency can identify a conclusion that is supported by a simple data set. At levels around the middle of the scale students can make judgements about the merit of a conclusion by identifying evidence that is consistent with the conclusion and evidence that does not support it. At the highest levels students can comment on whether evidence is consistent with a given hypothesis and describe the limitations that are inherent in conclusions.

Interpreting the scientific literacy levels

The proficiency levels defined and described in the preceding sections require one more set of technical decisions before they can be used to summarise and report the performance of particular students. The scale of PISA scientific literacy is a continuous scale. The use of performance bands, or levels of proficiency, involves an essentially arbitrary division of that continuous scale into discrete parts. The number of divisions and the location of the cut-points that mark the boundaries of the divisions are two matters that must be determined. For PISA science, the scale has been divided into seven regions, including 5 bounded regions labelled levels 1 to 5, an unbounded region below level 1, and an unbounded upper region (labelled level 6). The cutpoints that mark the boundaries between these regions were given in Table 15.1 .

The creation of these performance bands leads to a situation where a range of values on the continuous scale is grouped together into each single band. Given that range of performances within each level, how do we assign individual students to the levels, and what meaning do we ascribe to being at a level? In the context of the OECD reporting of PISA 2000 results, a common sense interpretation of the meaning of being at a level was developed and adopted. That is, students are assigned to the highest level for which they would be expected to correctly answer the majority of assessment items. If we could imagine a test composed of items spread uniformly across a level, a student near the bottom of the level will be expected to correctly answer at least half of the test questions from that level. Students at progressively higher points in that level would be expected to correctly answer progressively more of the questions in that level. It should be remembered that the relationship between students and items is probabilistic – it is possible to estimate the probability that a student at a particular location on the scale will get an item at a particular location on the

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



scale correct. Students assigned to a particular level will be expected to successfully complete some items from the next higher level, and it is only when that expectation reaches the threshold of 'at least half of the items' in the next higher level that the student would be placed in the next higher level. Mathematically, the probability level used to assign students to the scale to achieve this common-sense interpretation of being at a level is 0.62. Students are placed on the scale at the point where they have a 62% chance of correctly answering test questions located at the same point.

The same meaning has been applied in the reporting of PISA 2000, 2003 and 2006 results. Such an approach makes it possible to summarise aspects of student proficiency by describing the things related to PISA scientific literacy that students can be expected to do at different locations on the scale.

Figure 15.7 [Part 1/2]

Summary descriptions of the six proficiency levels in *using scientific evidence*

General proficiencies students should have at each level	Tasks a student should be able to do	Examples from released questions
LEVEL 6 2.4% of all students across the OECD area can perform tasks at Level 6 on the *using scientific evidence* scale.		
Students at this level demonstrate an ability to compare and differentiate among competing explanations by examining supporting evidence. They can formulate arguments by synthesising evidence from multiple sources.	▪ Recognise that alternative hypotheses can be formed from the same set of evidence. ▪ Test competing hypotheses against available evidence. ▪ Construct a logical argument for an hypothesis by using data from a number of sources.	
LEVEL 5 11.8% of all students across the OECD area can perform tasks at Level 5 on the *using scientific evidence* scale.		
Students at this level are able to interpret data from related datasets presented in various formats. They can identify and explain differences and similarities in the datasets and draw conclusions based on the combined evidence presented in those datasets.	▪ Compare and discuss the characteristics of different datasets graphed on the one set of axes. ▪ Recognise and discuss relationships between datasets (graphical and otherwise) in which the measured variable differs. ▪ Based on an analysis of the sufficiency of the data, make judgements about the validity of conclusions.	**GREENHOUSE** Question 4
LEVEL 4 31.6% of all students across the OECD area can perform tasks at Level 4 on the *using scientific evidence* scale.		
Students at this level can interpret a dataset expressed in a number of formats, such as tabular, graphic and diagrammatic, by summarising the data and explaining relevant patterns. They can use the data to draw relevant conclusions. Students can also determine whether the data support assertions about a phenomenon.	▪ Locate relevant parts of graphs and compare these in response to specific questions. ▪ Understand how to use a control in analysing the results of an investigation and developing a conclusion. ▪ Interpret a table that contains two measured variables and suggest credible relationships between those variables. ▪ Identify the characteristics of a straightforward technical device by reference to diagrammatic representations and general scientific concepts and thus form conclusions about its method of operation.	**SUNSCREENS** Question 5 **GREENHOUSE** Question 4 (Partial)

• • •

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



Figure 15.7 [Part 2/2]

Summary descriptions of the six proficiency levels in *using scientific evidence*

General proficiencies students should have at each level	Tasks a student should be able to do	Examples from released questions
LEVEL 3 56.3% of all students across the OECD area can perform tasks at Level 3 on the *using scientific evidence* scale.		
Students at this level are able to select a piece of relevant information from data in answering a question or in providing support for or against a given conclusion. They can draw a conclusion from an uncomplicated or simple pattern in a dataset. Students can also determine, in simple cases, if enough information is present to support a given conclusion.	▪ Given a specific question, locate relevant scientific information in a body of text. ▪ Given specific evidence/data, choose between appropriate and inappropriate conclusions. ▪ Apply a simple set of criteria in a given context in order to draw a conclusion or make a prediction about an outcome. ▪ Given a set of functions, determine if they are applicable to a specific machine.	**GREENHOUSE** Question 3
LEVEL 2 78.1% of all students across the OECD area can perform tasks at Level 2 on the *using scientific evidence* scale.		
Students at this level are able to recognise the general features of a graph if they are given appropriate cues and can point to an obvious feature in a graph or simple table in support of a given statement. They are able to recognise if a set of given characteristics apply to the function of everyday artifacts in making choices about their use.	▪ Compare two columns in a simple table of measurements and indicate differences. ▪ State a trend in a set of measurements or simple line or bar graph. ▪ Given a common artifact can determine some characteristics or properties pertaining to the artifact from among a list of properties.	**ACID RAIN** Question 3
LEVEL 1 92.1% of all students across the OECD area can perform tasks at Level 1 on the *using scientific evidence* scale.		
In response to a question, students at this level can extract information from a fact sheet or diagram pertinent to a common context. They can extract information from bar graphs where the requirement is simple comparisons of bar heights. In common, experienced contexts students at this level can attribute an effect to a cause.	▪ In response to a specific question pertaining to a bar graph, make comparisons of the height of bars and give meaning to the difference observed. ▪ Given variation in a natural phenomenon can, in some cases, indicate an appropriate cause (*e.g.* fluctuations in the output of wind turbines may be attributed to changes in wind strength).	

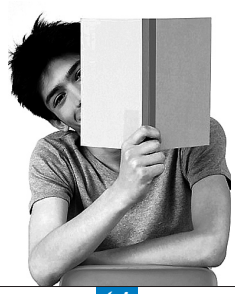
PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009



Notes

1. While strictly speaking the scales based on aspects of reading are sub-scales of the combined reading literacy scale, for simplicity they are mostly referred to as 'scales' rather than 'sub-scales' in this report.

2. Examples referred to are reproduced in Volume 1 of *PISA 2006: Science Competencies for Tomorrow's World*.



16

Scaling Procedures and Construct Validation of Context Questionnaire Data

Overview 304
Simple questionnaire indices 304
- Student questionnaire indices 304
- School questionnaire indices 307
- Parent questionnaire indices 309

Scaling methodology and construct validation 310
- Scaling procedures 310
- Construct validation 312
- Describing questionnaire scale indices 314

Questionnaire scale indices 315
- Student scale indices 315
- School questionnaire scale indices 340
- Parent questionnaire scale indices 342
- The PISA index of economic, social and cultural status (ESCS) 346

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



OVERVIEW

The PISA 2006 context questionnaires included numerous items on student characteristics, student family background, student perceptions, school characteristics and perceptions of school principals. In 16 countries (optional) parent questionnaires were administered to the parents of the tested students.

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 so as to measure latent constructs that cannot be observed directly. For these items, transformations or scaling procedures are needed to construct meaningful indices.

This chapter describes how student, school and parent questionnaire indices 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;
- Scale indices: These indices were constructed through the scaling of items. Typically, scale scores for these indices are estimates of latent traits derived through 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 index on economic, social and cultural status (ESCS), including a discussion of some modifications from the PISA 2003 ESCS index. Some indices had already been used in previous PISA surveys and are constructed based on a similar scaling methodology (see Schulz, 2002; and OECD 2005). Most indices, however, were based on the elaboration of a questionnaire framework and are related to science as the major domain of the third PISA survey (see Chapter 3).

SIMPLE QUESTIONNAIRE INDICES

Student 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 know 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

$$AGE = (100 + T_y - S_y) + \frac{(T_m - S_m)}{12}$$

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 2006 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 will be included in a separate variable (*PROGN*) where the first three digits are the ISO code for a country, the next two digits are the sub-national category, and the last two digits

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



are the nationally specific programme code. All study programmes were classified using the international standard classification of education (ISCED) (OECD, 1999). The following indices are derived from the data on study programmes: programme level (*ISCDL*) indicating whether students are on the lower or upper secondary level (ISCDE 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 is 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 response were coded to four-digit ISCO codes (ILO,1990) and then mapped to the international socio-economic index of occupational status (*ISEI*) (Ganzeboom *et al.*, 1992). Three indices were obtained from these scores: father's occupational status (*BFMJ*); mother's occupational status (*BMMJ*); 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 a more relevant influence on a student's outcomes than is 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 educational outcomes than has been done on occupational status. The core 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 immigration, difficulties in the national mapping of parental qualifications gained abroad.

Parental education is classified using ISCED (OECD,1999). Indices on parental education are 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 (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 on the highest educational level of parents (*HISCED*) corresponds to the higher ISCED level of either parent.

The index scores for highest educational level of parents were also recoded into estimated years of schooling (*PARED*). A mapping of ISCED levels of years of schooling is provided in Appendix 5.

Immigration background

As in PISA 2000 and PISA 2003, information on the country of birth of the students and their parents was collected. Included in the database are three country-specific variables relating to the country of birth of the student, mother, and father (*CTNUMS, CTNUMM,* and *CTNUMF*). Also, the items ST11Q01, ST11Q02 and ST11Q03 have been recoded for the database into the following categories: (1) country of birth is same as country of assessment, and (2) otherwise.



The index on immigrant background (*IMMIG*) is calculated from these variables, and has the following categories: (1) native students (those students who had at least one parent born in the country), (2) first-generation students (those students born outside the country of assessment and whose parents were also born in another country). and (3) second generation' students (those born in the country of assessment but whose parent(s) were born in another country), Students with missing responses for either the student or for both parents have been given missing values for this variable.

Language spoken at home

Similar to PISA 2003, students also indicated what language they usually spoke at home, and the database includes a variable (*LANGN*) containing country-specific codes for each language. In addition, the item ST12Q01 has be recoded for the international database into the following categories: (1) language at home is same as the language of assessment for that student, (2) language at home is a national language of the country but the student was assessed in a different language, and (3) language at home is another (foreign) language.

Expected occupational status

As in PISA 2000 and 2003, students were asked to report their expected occupation at age 30 and a description of this job. The responses were coded to four-digit ISCO codes (ILO, 1990) and then mapped to the *ISEI* index (Ganzeboom *et al.*, 1992). Recoding of ISCO codes into *ISEI* index results in scores for the students' expected occupational status (*BSMJ*), where higher scores of *ISEI* indicate higher levels of expected occupational status.

Blue-collar/white-collar parental occupation

As in 2003, the ISCO codes of parents were recoded into 4 categories: (1) white collar high skilled, (2) white collar low skilled, (3) blue collar high skilled, and (4) blue collar low skilled. Three variables are included, one indicating the mother's employment category (*MSECATEG*), another indicating father's employment category (*FSECATEG*), and another indicating the highest employment category of either parent (*HSECATEG*).

Table 16.1
ISCO major group white-collar/blue-collar classification

ISCO Major Group	White-collar/blue-collar classification
1	White-collar high-skilled
2	White-collar high-skilled
3	White-collar high-skilled
4	White-collar low-skilled
5	White-collar low-skilled
6	Blue-collar high-skilled
7	Blue-collar high-skilled
8	Blue-collar low-skilled
9	Blue-collar low-skilled

Science-related occupations for parents and students

The ISCO data were used to compute four variables indicating whether or not the student expects to have a science-related career at age 30 (*SCIS5*), whether their mother (*SCIM1*) or father (*SCIM2*) are in a science career, or whether either or both parents are in a science related career (*SCIH12*). Values of 1 on these indicate "yes", while values of 0 indicate "no or undetermined".

To reduce the amount of missing data for parents' career status, parents with the following responses for occupations were recoded to "no/undetermined': home makers, social beneficiaries and students.



Furthermore, to reduce the amount of missing data on students' expected career status at age 30, students indicating "don't know" were recoded from missing to "no/undetermined". Also, students who responded to the items immediately subsequent to this question, but who did not respond to expected job at 30 were recoded to "no/undetermined'.

Since the ISCO coding scheme is rather broad for this purpose (*e.g.* some teaching professionals may be in a science-related career, but the scheme does not distinguish between teachers in different subject areas and disciplines), these science-related career variables should be interpreted as broad indicators rather than precise classifications. The ISCO occupation categories that were classified as science-related occupations are shown in Table 16.2.

Table 16.2
ISCO occupation categories classified as science-related occupations

ISCO Group Number	Occupation Category
1236	Computing services department managers
1237	Research and development department managers
211	Physicists, chemists and related professionals
2122	Statisticians
213	Computing professionals
214	Architects, engineers etc, professionals
221	Life science professionals
222	Health professionals except nursing
223	Nursing and midwifery professionals
2442	Sociologists, anthropologists etc, professionals
2445	Psychologists
2446	Social work professionals
311	Physical and engineering science associate professionals
313	Optical and electronic equipment operators
3143	Aircraft pilots etc, associate professionals
3144	Air traffic controllers
3145	Air traffic safety technicians
315	Safety and quality inspectors
321	Life science etc, associate professionals
322	Modern health professionals except nursing
323	Nursing and midwifery associate professionals

School questionnaire indices

School size

As in previous surveys, the PISA 2006 index of school size (*SCHLSIZE*) 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.

Class size

The average class size (*CLSIZE*) is derived from one of nine possible categories, ranging from "15 students or fewer" to "More than 50 students'. *CLSIZE* takes the midpoint of each response category, a value of 13 for the lowest category, and a value of 53 for the highest.

Proportion of girls enrolled at school

As in previous surveys, the PISA 2006 index on the proportion of girls at school (*PCGIRLS*) is based on the enrolment data provided by the school principal, dividing the number of girls by the total of girls and boys at a school.

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



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: (1) public schools controlled and managed by a public education authority or agency, (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) 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.[1]

Availability of computers

As in PISA 2000 and PISA 2003, school principals were asked to report the number of computers available at school. However, the question wording was modified for 2006 where principles were asked to report on the total number of computers, the number of computers available for instruction and the number of computers connected to the internet. The index of availability of computers (*RATCOMP*) is obtained by dividing the number of computers at school by the number of students at school. The overall ratio of computers to school size (*IRATCOMP*) was obtained by dividing the number of computers available for instruction at school by the number of students at school. The proportion of computers connected to the Internet (*COMPWEB*) was obtained by dividing the total number of computers connected to the Web by the total number of computers.

Quantity of teaching staff at school

As in previous PISA surveys, principles were asked to report the number of full-time and part-time teachers at school. However, the number of items was reduced in 2006 to capture only teachers in total, certified teachers, and teachers with an ISCED 5A qualification.

The student-teacher ratio (*STRATIO*) was obtained by dividing the school size by the total number of teachers. The number of part-time teachers is weighted by 0.5 and the number of full-time teachers is weighted by 1.0. 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 who have an ISCED 5A qualification (*PROP5A*) was calculated by dividing the number of these kinds of teachers by the total number of teachers.

School selectivity

As in previous surveys, school principals were asked about admittance policies at their school. Among these policies, principles were asked how much consideration was given to the following factors when students are admitted to the school, based on a scale with the categories "not considered", "considered", "high priority", and "pre-requisite": students' academic record (including placement tests) and the recommendation of feeder schools.

An index of school selectivity (*SELECT*) was computed by assigning schools to four different categories: (1) schools where none of these factors is considered for student admittance; (2) schools considering at least one of these factors; (3) schools giving high priority to at least one of these factors; and (4) schools where at least one of these factors is a pre-requisite for student admittance.

Ability grouping

School principals were asked to report the extent to which their school organises instruction differently for student with different abilities. PISA 2003 included a similar question with two additional items which focused on mathematics classes. In 2006, this has been reduced to two items which ask about subject



grouping in a more general sense. One item asked about the occurrence of ability grouping into different classes and the other regarding ability grouping within classes (with the response categories "For all subjects", "For some subjects" and "Not for any subject").

An index of ability grouping between or within classes (*ABGROUP*) was derived from the two items by assigning schools to three categories: (1) schools with no ability grouping for any subjects, (2) schools with at least one of these forms of ability grouping for some subjects and (3) schools with at least one of these two forms of ability grouping for all subjects.

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 measuring the school principals' report on who has considerable responsibility for tasks regarding school management of 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 principal or teachers to "yes" responses for central educational authority. Higher values on the scale indicate 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).[2]

School responsibility for curriculum and assessment

An index of the relative level of responsibility of school staff in issues relating to curriculum and assessment (*RESPCURR*) was computed from four items measuring the school principal's report concerning who had responsibility for curriculum and assessment ("Establishing student assessment policies", "Choosing which textbooks are used", "Determining course content", "Deciding which courses are offered"). The index was calculated on the basis of the ratio of "yes" responses for principal or teachers to "yes" responses for central education authorities. Higher values indicate relatively higher levels of school responsibility in this area. The index was standardised to having an OECD mean of zero and a standard deviation of one (for the pooled data with equally weighted country samples).[3]

Parent questionnaire indices

Educational level of parents

Administration of this instrument in PISA 2006 provided the opportunity to collect data on parental education directly from the parents in addition to the data provided by the student questionnaire. Similar to the student questionnaire data, parental education 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 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 (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

Occupational data for both the student's father and student's mother were obtained by asking open-ended questions in a manner similar to the questions asked of students. The responses were coded to four-digit



ISCO codes (ILO, 1990) and then mapped to the SEI index (Ganzeboom, de Graaf & Treiman, 1992). Three SEI indices were computed from these scores.

Recoding of ISCO codes into SEI gives scores for the Mother's occupational status (*PQBMMJ*) and Father's occupational status (*PQBFMJ*). The highest occupational level of parents (*PQHISEI*) is the higher SEI score of either parent or to the only available parent's SEI score. Higher scores of SEI will indicate higher level of occupational status.

Similar to the science-related career variables derived from the student questionnaire, three indicators were derived from the parent data: whether the mother (*SCIM3*) or father (*SCIF4*) is in a science-related career, and whether either or both of the parents is in a science-related career (*SCIH34*).

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

$$P_i(\theta) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

where $P_i(\theta)$ 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

$$P_{x_i}(\theta) = \frac{\exp \sum_{k=0}^{x} (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^{h} (\theta_n - \delta_i + \tau_{ik})} \quad x_i = 0, 1, \ldots, m_i$$

where $P_{xi}(\theta)$ denotes the probability of person *n* to score *x* on item *i*. 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.

Item fit was assessed using the weighted mean-square statistic (infit), which is a residual based fit statistic. Weighted infit statistics were reviewed both for item and step parameters. The ACER *ConQuest*® software (Wu, Adams and Wilson, 1997) was used for the estimation of item parameters and the analysis of item fit.

International item parameters were obtained from calibration samples consisting of randomly selected sub-samples:

- For the calibration of student item parameters, sub-samples of 500 students were randomly selected within each OECD country sample. 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 student within explicit sampling strata. The final calibration sample included data from 15,000 students;



- For the calibration of school item parameters, 100 schools were randomly selected within each OECD country sample. The random selection was based on school level weights in order to ensure that a representative sample of schools was selected from each country. School data from Luxembourg were not included due to of the small number of schools. Data from France were not available because the school questionnaire was not administered in France. The final calibration sample included data from 2 800 school principals.

Once international item parameter had been estimated from the calibration sample, weighted likelihood estimation was used to obtain individual student scores. WLEs can be computed by minimising the equation

16.4

$$\sum_{i\in\Omega}\left[\left(r_x+\frac{J_n}{2I_n}\right)-\sum_{j=1}^{k}\frac{\exp\left(\sum_{j=0}^{x}\theta_n-\delta_i+\tau_{ij}\right)}{\sum_{h=0}^{m_i}\exp\sum_{k=1}^{k}\left(\theta_n-\delta_i+\tau_{ij}\right)}\right]=0$$

for each case n, where r_x is the sum score obtained from a set of k items with j categories. This can be achieved by applying the Newton-Raphson method. The term $J_n/2I_n$ (with I_n being the information function for student n and J_n being its derivative with respect to θ) is used as a weight function to account for the bias inherent to maximum likelihood estimation (*see* Warm, 1989). IRT scores were derived using ACER *ConQuest*® with pre-calibrated item parameters.

Table 16.3
OECD means and standard deviations of WL estimates

Student-level indices	Mean	Standard deviation
CARINFO	–0.14	2.12
CARPREP	1.33	2.19
CULTPOSS	0.30	1.64
ENVAWARE	0.30	1.39
ENVOPT	–0.92	1.39
ENVPERC	1.77	1.42
GENSCIE	1.65	1.65
HEDRES	2.67	1.52
HIGHCONF	1.33	1.36
HOMEPOS	1.57	1.11
INSTSCIE	0.65	3.19
INTCONF	2.52	1.29
INTSCIE	–0.09	1.35
INTUSE	0.24	0.88
JOYSCIE	0.42	3.29
PERSCIE	0.58	1.80
PRGUSE	–0.53	1.04
RESPDEV	1.52	1.45
SCAPPLY	–0.20	1.63
SCHANDS	–0.73	1.64
SCIEACT	–2.04	1.68
SCIEEFF	0.45	1.31
SCIEFUT	–1.52	3.16
SCINTACT	–0.07	1.56
SCINVEST	–1.58	1.64
SCSCIE	0.23	3.04
WEALTH	1.28	1.46
School-level indices		
ENVLRN	–1.87	1.54
SCIPROM	0.95	1.53
SCMATEDU	0.24	1.55
TCSHORT	0.62	1.40

Note: Means and standard deviations for equally weighted OECD data.

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



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

16.5

$$\theta'_n = \frac{\theta_n - \bar{\theta}_{OECD}}{\sigma_{\theta(OECD)}}$$

where θ'_n are the scores in the international metric, θ_n the original WLE in logits, and $\bar{\theta}_{OECD}$ is the OECD mean of logit scores with equally weighted country sub-samples. $\sigma_{\theta(OECD)}$ is the corresponding OECD standard deviation of the original WL estimates. Means and standard deviations used for the transformation into the international metric are shown in Table 16.3.

Construct validation

As in previous PISA surveys, it was important to develop comparable measures of student background, attitudes and perceptions. There are different methodological approaches for validating questionnaire constructs, each with their advantages, limitations and problems. Cross-country validity of these constructs is of particular importance as measures derived from questionnaires are often used to explain differences in student performance within and across countries and are, thus, potential sources of policy-relevant information about ways of improving educational systems.

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 (1) consistent dimensionality and (2) consistent construct validity across participating countries. This means that, once the measurement stability for each scale is confirmed, the multidimensional relationship between these constructs should be reviewed as well (see Wilson, 1994; Schulz 2006a; Walker 2006). It should be noted, however, that between-country differences in the strength of relationships between constructs do not necessarily indicate a lack of consistency as they may be due to differences between national contexts (for example, different educational systems or learning practices).

Confirmatory factor analysis

Structural Equation Modelling (SEM) was used to confirm theoretically expected dimensions and, if necessary, to re-specify the dimensional structure (Kaplan, 2000). Using Confirmatory Factor Analysis (CFA) requires a theoretical model of item dimensionality, which can be tested using the collected data.

Fit indices measure the extent to which a model based on the a-priori structure as postulated by the researcher fits the data. In the PISA 2006 analysis, model fit was assessed using the root-mean square error of approximation (RMSEA), the root mean square residual (RMR), the comparative fit index (CFI) and the non-normed fit index (NNFI) (see Bollen and Long, 1993). RMSEA values over 0.10 are usually interpreted as a sign of unacceptable model fit whereas values below 0.05 indicate a close model fit. RMR values should be less than 0.05. Both CFI and NNFI are bound between 0 and 1 and values between 0.90 and 0.95 indicate an acceptable model fit, with values greater than 0.95 indicating a close model fit.

For the results presented in this chapter, maximum likelihood estimation and covariance matrices were used for the analyses of the (categorical) Likert-type items, that is, the items were treated as if they were continuous. Confirmatory factor analyses of student data were based on the international calibration sample in order to have comparable (sub-)sample sizes across OECD countries. For the comparative analysis of item dimensionality the use of random OECD sub-samples was deemed appropriate.

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



The SAS® CALIS procedure and the *LISREL* program were used to estimate the models based on Likert-type items. In order to assess cross-country validity of item dimensionality and constructs models were estimated both for the pooled OECD calibration sample (with 500 students per country) and for each country calibration sub-sample separately. CFA were carried out only for the student questionnaire data.

In the case of dichotomous items, weighted least squares (WLS) estimation with polychoric correlations was used (see Jöreskog and Sörbom, 1993). As the unadjusted WLS estimator requires very large sample sizes, a mean- and variance- adjusted WLS estimator (WLSMV) was used, which is available in the *Mplus* software program, (see Muthén, du Toit, and Spisic, 1997). Confirmatory factor analyses for dichotomous student-level items were only estimated for the pooled international calibration sample.

Between-school variance of student-level indices

The structure of the national PISA samples includes students that are nested within schools. Consequently, the variation in variables collected from students can either be between or within schools. Analyses of cognitive data tend to show that depending on the structure of educational systems in some countries a considerable amount of variation is found between schools.

Table 16.4 shows the median, maximum and minimum percentages of between-school variance for student questionnaire indices. For most of the student-level indices the average proportion of between-school variance is below 10%. However, for some indices there is a considerable variance between schools. Notably, home background indices like *WEALTH*, or *CULTPOSS* have relatively high intra-class correlations in many countries.

Table 16.4
Median, minimum and maximum percentages of between-school variance for student-level indices across countries

	OECD countries			Partner countries and economies		
Index	Median	Max	Min	Median	Max	Min
CULTPOSS	11	24	3	11	21	5
WEALTH	10	40	5	20	44	0
ENVAWARE	8	16	2	10	19	4
SCINVEST	8	17	3	9	22	2
HEDRES	8	30	2	16	39	3
INTCONF	7	24	2	16	48	6
SCIEEFF	6	14	1	6	13	4
SCINTACT	6	10	2	6	11	2
SCHANDS	6	10	2	6	11	2
CARPREP	6	18	1	4	11	2
INTSCIE	5	13	2	5	14	2
JOYSCIE	5	13	2	6	16	2
SCAPPLY	5	13	1	6	13	0
INTUSE	5	12	1	7	22	0
INSTSCIE	5	14	1	6	12	0
SCIEFUT	5	12	0	6	18	1
SCSCIE	4	14	2	4	12	0
GENSCIE	4	9	2	4	8	2
SCIEACT	4	8	1	6	24	3
RESPDEV	4	11	2	4	8	1
PRGUSE	4	10	2	4	12	0
PERSCIE	3	8	1	4	9	1
CARINFO	3	12	0	4	9	1
HIGHCONF	3	9	2	8	35	0
ENVOPT	3	9	0	5	17	1
ENVPERC	2	11	0	3	9	0

Note: Results from multi-level analysis with random intercepts only.

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



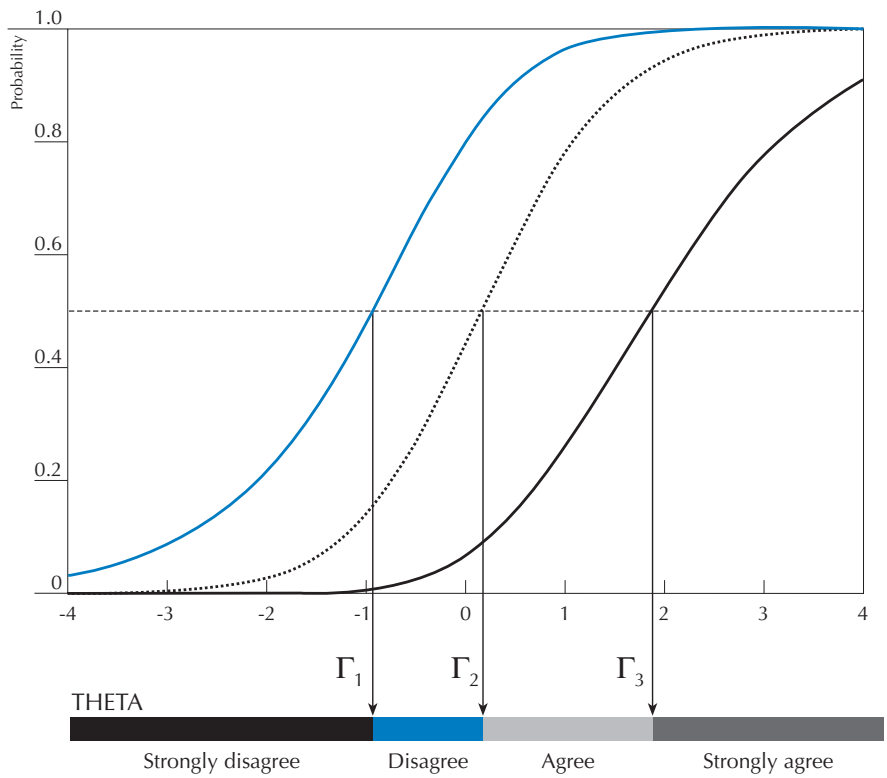
Describing questionnaire scale indices

As in previous PISA surveys, in PISA 2006 categorical items from the context questionnaires were scaled using IRT modelling. Weighted likelihood estimates (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.[4]

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.

Figure 16.1
Summed category probabilities for fictitious item



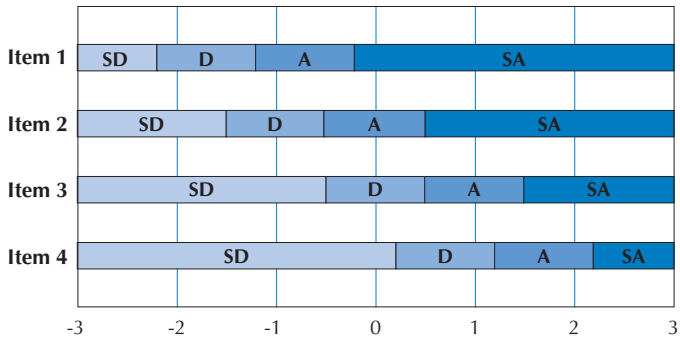
PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009



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 for each item category those points on a scale, at which respondents have a .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 50% chance to agree with the item.

Figure 16.2
Fictitious example of an item map



The fictitious example in Figure 16.2 illustrates the interpretation of an item map for a fictitious scale with four different Likert-type items:

- 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 more than 0.5 probability 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;
- Likewise, students with a score 1 (one standard deviation above the OECD average) would have more than a 0.5 probability to strongly agree with the first two items, but still have less than 0.5 probability 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 household possessions as indicators of family wealth has received much attention in international studies in the field of education (Buchmann, 2000). Household assets are believed to capture wealth better than income because they reflect a more stable source of wealth.



In PISA 2006, students reported the availability of 13 different household items at home. In addition, countries added three specific household items that were seen as appropriate measures of family wealth within the country's context. Appendix 6 includes a list of the country-specific household items.

Four different indices were derived from these items: (i) family wealth possessions (*WEALTH*), (ii) cultural possessions (*CULTPOSS*), (iii) home educational resources (*HEDRES*) and (iiii) home possessions (*HOMEPOS*). The last index is a summary index of all household items and also included the variable indicating the number of books at home, but recoded into three categories: (0) 0-25 books, (1) 26-100 books, and (2) 101 or more books. *HOMEPOS* was also one of three components in the construction of the index on economic, social and cultural status (ESCS, see the section on ESCS index construction below). Table 16.5 shows the wording of items and their allocation to the four indices.

A confirmatory factor analysis using polychoric correlations with a WLSMV estimator showed a reasonable model fit for the international calibration sample of OECD countries (RMSEA = 0.080, CFI = 0.88, NNFI = 0.92). The estimated latent correlations between these constructs were 0.80 between *WEALTH* and *HEDRES*, 0.25 between *WEALTH* and *CULTPOSS*, and 0.52 between *CULTPOSS* and *HEDRES*.[5]

Analysis of differential item functioning (DIF) showed a considerable amount of between-country variation in the item parameters. It was decided to use nationally defined item parameters for scaling instead of using parameters estimated for the combined OECD sample (as done in previous cycles).

Table 16.5
Household possessions and home background indices

		Item is used to measure index			
Item		WEALTH	CULTPOSS	HEDRES	HOMEPOS
ST13	In your home, do you have:				
ST13Q01	A desk to study at			X	X
ST13Q02	A room of your own	X			X
ST13Q03	A quiet place to study			X	X
ST13Q04	A computer you can use for school work			X	
ST13Q05	Educational software			X	X
ST13Q06	A link to the Internet	X			X
ST13Q07	Your own calculator			X	X
ST13Q08	Classic literature (*e.g.* <Shakespeare>)		X		X
ST13Q09	Books of poetry		X		X
ST13Q10	Works of art (*e.g.* paintings)		X		X
ST13Q11	Books to help with your school work			X	X
ST13Q12	A dictionary			X	X
ST13Q13	A dishwasher (country-specific)	X			X
ST13Q14	A <DVD or VCR> player (country-specific)	X			X
ST13Q15	<Country-specific wealth item 1>	X			X
ST13Q16	<Country-specific wealth item 2>	X			X
ST13Q17	<Country-specific wealth item 3>	X			X
ST14	How many of these are there at your home?				
ST14Q01	Cellular phones	X			X
ST14Q02	Televisions	X			X
ST14Q03	Computers	X			X
ST14Q04	Cars	X			X
ST15	How many books are there in your home				X

Note: Item categories were "yes" (1) and "no" (2) for ST13, "None", "One", "Two" and "Three or more" for ST14, The categories for ST15 ("0-10 books", "11-25 books", "26-100 books", 101-200 books", "201-500 books" and "More than 500 books") were recoded into three categories ("0-25 books", 26-100 books" and "More than 100 books" ; Items in ST13 for were inverted for scaling and the first two categories of ST14Q01 and ST14Q02 were collapsed into one for scaling.

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



The *WEALTH* and *HOMEPOS* scales were constructed in two stages. A basket of common items was chosen (ST13Q02, ST13Q06, ST14Q01, ST14Q02, ST14Q03 and ST14Q04 for *WEALTH*, and in addition to these ST13Q01, ST13Q03, ST13Q05 to ST13Q12 and ST15Q01 for *HOMEPOS*) and item parameters were estimated for each country based on this item set. The sum of the set's item parameters was constrained to zero for each country. Next, these item parameters were anchored. The remaining country-specific items were added, and each country was scaled separately.

The other two scales derived from household possession items, *CULTPOSS* and *HEDRES*, were scaled in one step but the item parameters were allowed to vary by country.

Table 16.6 shows the scale reliabilities in OECD countries for all four scales, Table 16.7 those in partner countries. *HEDRES* has notably lower scale reliabilities when compared with the three indices. Similar results were already found for this index in PISA 2000 (see Schulz, 2002, p. 214) and PISA 2003 (see OECD, 2005, p. 284).

When comparing OECD and partner countries it appears that scale reliabilities for *WEALTH*, *HEDRES* and *HOMEPOS* are generally higher in partner countries. This may be due to the higher degree of accessibility of household items for larger proportions of the population in developed countries: In more developed countries there are very high percentages of students reporting the existence of many of the household items which makes them less appropriate as indicators of wealth.

Table 16.6
Scale reliabilities for home possession indices in OECD countries

	WEALTH	HEDRES	CULTPOSS	HOMEPOS
Australia	0.60	0.60	0.61	0.62
Austria	0.61	0.41	0.60	0.63
Belgium	0.62	0.47	0.62	0.61
Canada	0.61	0.53	0.63	0.64
Czech Republic	0.67	0.51	0.59	0.65
Denmark	0.60	0.43	0.64	0.58
Finland	0.59	0.44	0.67	0.61
France	0.65	0.46	0.64	0.64
Germany	0.64	0.47	0.61	0.62
Greece	0.66	0.42	0.53	0.65
Hungary	0.70	0.50	0.62	0.73
Iceland	0.55	0.44	0.61	0.59
Ireland	0.56	0.56	0.62	0.61
Italy	0.61	0.47	0.57	0.64
Japan	0.60	0.46	0.61	0.65
Korea	0.64	0.50	0.62	0.73
Luxembourg	0.61	0.49	0.64	0.64
Mexico	0.83	0.60	0.58	0.77
Netherlands	0.57	0.42	0.57	0.56
New Zealand	0.67	0.57	0.61	0.69
Norway	0.58	0.57	0.67	0.61
Poland	0.72	0.60	0.58	0.74
Portugal	0.73	0.51	0.67	0.75
Slovak Republic	0.68	0.62	0.62	0.71
Spain	0.64	0.45	0.58	0.64
Sweden	0.58	0.53	0.63	0.63
Switzerland	0.60	0.42	0.58	0.59
Turkey	0.78	0.66	0.54	0.76
United Kingdom	0.60	0.58	0.66	0.63
United States	0.70	0.65	0.66	0.74
Median	**0.62**	**0.49**	**0.61**	**0.64**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

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



Table 16.7
Scale reliabilities for home possession indices in partner countries/economies

	WEALTH	HEDRES	CULTPOSS	HOMEPOS
Argentina	0.77	0.59	0.52	0.75
Azerbaijan	0.81	0.57	0.59	0.64
Brazil	0.80	0.58	0.46	0.72
Bulgaria	0.74	0.66	0.63	0.74
Chile	0.77	0.60	0.52	0.78
Colombia	0.77	0.64	0.56	0.74
Croatia	0.68	0.44	0.65	0.68
Estonia	0.69	0.45	0.57	0.69
Hong Kong-China	0.61	0.47	0.55	0.71
Indonesia[1]	0.78	0.55	0.48	0.65
Israel	0.75	0.62	0.68	0.65
Jordan	0.80	0.71	0.52	0.72
Kyrgyzstan	0.73	0.47	0.46	0.62
Latvia	0.69	0.51	0.57	0.70
Liechtenstein	0.59	0.33	0.66	0.57
Lithuania	0.70	0.50	0.70	0.73
Macao-China	0.70	0.50	0.56	0.72
Montenegro	0.72	0.52	0.58	0.66
Qatar	0.78	0.70	0.55	0.75
Romania	0.79	0.69	0.51	0.80
Russian Federation	0.68	0.56	0.46	0.72
Serbia	0.71	0.54	0.65	0.70
Slovenia	0.61	0.42	0.65	0.63
Chinese Taipei	0.56	0.55	0.68	0.68
Thailand	0.82	0.63	0.54	0.80
Tunisia	0.84	0.72	0.56	0.73
Uruguay	0.79	0.58	0.58	0.74
Median	**0.74**	**0.56**	**0.56**	**0.72**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.
1. Indonesia had omitted item ST13Q13 ("Dishwasher") from their national questionnaire and reliabilities for WEALTH and HOMEPOS were computed without this item.

Interest and enjoyment of science learning

Eight items are used to measure general interest in science learning in PISA 2006. While the interest items which are embedded in the test instrument provide data on interest in specific contexts, the items here will provide data on students' interest in more general terms. All items were inverted for scaling and positive scores indicate higher levels of interest in learning science. Item wording and model parameters are displayed in Table 16.8.

Table 16.8
Item parameters for interest in science learning (INTSCIE)

Item	How much interest do you have in learning about the following <broad science> topics?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST21Q01	a) Topics in physics	0.04	–1.52	–0.04	1.57
ST21Q02	b) Topics in chemistry	–0.05	–1.42	0.01	1.40
ST21Q03	c) The biology of plants	0.03	–1.62	0.06	1.55
ST21Q04	d) Human biology	–0.76	–1.35	–0.12	1.47
ST21Q05	e) Topics in astronomy	–0.2	–1.28	0.11	1.17
ST21Q06	f) Topics in geology	0.32	–1.7	0.08	1.62
ST21Q07	g) Ways scientists design experiments	0.11	–1.43	0.07	1.35
ST21Q08	h) What is required for scientific explanations	0.51	–1.60	0.05	1.55

Note: Item categories were "high interest", "medium interest", "low interest" and "no interest"; all items were inverted for scaling.



Four items are used to measure enjoyment of science learning in PISA 2006. All items were inverted for IRT scaling so that positive WLE scores on this new index for PISA 2006 indicate higher levels of enjoyment of science. Table 16.9 shows the item wording and the international item parameters for this scale.

Table 16.9
Item parameters for enjoyment of science (JOYSCIE)

Item	How much do you agree with the statements below?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST16Q01	a) I generally have fun when I am learning <broad science> topics	–0.43	–4.17	–0.4	4.57
ST16Q02	b) I like reading about <broad science>	0.51	–4.39	–0.12	4.51
ST16Q03	c) I am happy doing <broad science> problems	1.01	–4.6	0.02	4.57
ST16Q04	d) I enjoy acquiring new knowledge in <broad science>	–0.69	–3.91	–0.61	4.52
ST16Q05	e) I am interested in learning about <broad science>	–0.41	–3.84	–0.42	4.26

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

The fit for a two-factor model was not satisfactory for the pooled international sample and most of the country sub-samples (see Table 16.10). However, the lack of fit is mostly due to correlated error terms between interest items about similar topics (like biology of plants and human biology). The results also show high correlations (typically between 0.70 and 0.80) between the two constructs whose strength does not vary much across country sub-samples.

Table 16.10
Model fit and estimated latent correlations for interest in and enjoyment of science learning[1]

			Model fit			Latent correlations between:
		RMSEA	RMR	CFI	NNFI	INTSCIE/JOYSCIE
OECD	Australia	0.114	0.048	0.90	0.90	0.77
	Austria	0.110	0.061	0.88	0.88	0.76
	Belgium	0.113	0.056	0.88	0.88	0.81
	Canada	0.122	0.063	0.88	0.88	0.81
	Czech Republic	0.119	0.068	0.83	0.83	0.72
	Denmark	0.150	0.060	0.84	0.84	0.80
	Finland	0.138	0.055	0.85	0.85	0.72
	France	0.107	0.052	0.89	0.89	0.77
	Germany	0.106	0.056	0.89	0.89	0.79
	Greece	0.113	0.070	0.87	0.87	0.74
	Hungary	0.105	0.060	0.86	0.86	0.68
	Iceland	0.137	0.051	0.88	0.88	0.78
	Ireland	0.119	0.060	0.88	0.88	0.81
	Italy	0.102	0.043	0.88	0.88	0.72
	Japan	0.106	0.048	0.91	0.91	0.81
	Korea	0.115	0.057	0.86	0.86	0.81
	Luxembourg	0.100	0.053	0.90	0.90	0.71
	Mexico	0.121	0.048	0.81	0.81	0.59
	Netherlands	0.136	0.058	0.85	0.85	0.81
	New Zealand	0.106	0.050	0.90	0.90	0.83
	Norway	0.097	0.036	0.94	0.94	0.74
	Poland	0.126	0.062	0.85	0.85	0.73
	Portugal	0.114	0.047	0.86	0.86	0.67
	Slovak Republic	0.111	0.055	0.85	0.85	0.61
	Spain	0.139	0.068	0.84	0.84	0.77
	Sweden	0.105	0.039	0.93	0.93	0.81
	Switzerland	0.090	0.048	0.92	0.92	0.78
	Turkey	0.118	0.065	0.87	0.87	0.71
	United Kingdom	0.103	0.046	0.89	0.89	0.67
	United States	0.099	0.039	0.93	0.93	0.73
	OECD	**0.106**	**0.048**	**0.90**	**0.90**	**0.75**

1. Model estimates based on international student calibration sample (500 students per OECD country).



Table 16.11 shows the scale reliabilities for both indices in OECD and partner countries. The internal consistency for both scales is very high and typically above 0.80 for *INTSCIE* and 0.90 for *JOYSCIE*.

Table 16.11
Scale reliabilities for interest in and enjoyment of science learning

		INTSCIE	JOYSCIE
OECD	Australia	0.87	0.94
	Austria	0.79	0.92
	Belgium	0.85	0.89
	Canada	0.83	0.93
	Czech Republic	0.76	0.88
	Denmark	0.87	0.93
	Finland	0.85	0.93
	France	0.83	0.90
	Germany	0.80	0.92
	Greece	0.81	0.92
	Hungary	0.75	0.89
	Iceland	0.89	0.95
	Ireland	0.84	0.93
	Italy	0.80	0.88
	Japan	0.86	0.93
	Korea	0.81	0.91
	Luxembourg	0.82	0.92
	Mexico	0.81	0.82
	Netherlands	0.85	0.92
	New Zealand	0.85	0.93
	Norway	0.90	0.94
	Poland	0.79	0.92
	Portugal	0.83	0.87
	Slovak Republic	0.81	0.88
	Spain	0.83	0.90
	Sweden	0.88	0.95
	Switzerland	0.82	0.91
	Turkey	0.83	0.91
	United Kingdom	0.85	0.91
	United States	0.87	0.93
	Median	**0.83**	**0.92**

		INTSCIE	JOYSCIE
Partners	Argentina	0.83	0.88
	Azerbaijan	0.81	0.83
	Brazil	0.85	0.83
	Bulgaria	0.82	0.85
	Chile	0.82	0.87
	Colombia	0.78	0.85
	Croatia	0.78	0.88
	Estonia	0.75	0.87
	Hong Kong-China	0.83	0.91
	Indonesia	0.76	0.83
	Israel	0.88	0.94
	Jordan	0.81	0.82
	Kyrgyzstan	0.75	0.79
	Latvia	0.72	0.82
	Liechtenstein	0.87	0.94
	Lithuania	0.73	0.87
	Macao-China	0.79	0.89
	Montenegro	0.81	0.87
	Qatar	0.88	0.91
	Romania	0.81	0.83
	Russian Federation	0.76	0.84
	Serbia	0.77	0.85
	Slovenia	0.79	0.90
	Chinese Taipei	0.87	0.91
	Thailand	0.84	0.82
	Tunisia	0.71	0.77
	Uruguay	0.80	0.89
	Median	**0.81**	**0.87**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Motivation to learn science

Five items measuring the construct of instrumental motivation were included in the PISA 2006 main study. All items were inverted for IRT scaling: positive WLE scores on this new index for PISA 2006 indicate higher levels of instrumental motivation to learn science.

Table 16.12
Item parameters for instrumental motivation to learn science (INSTSCIE)

Item	How much do you agree with the statements below? (Strongly agree/Agree/Disagree/Strongly disagree)	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST35Q01	a) Making an effort in my <school science> subject(s) is worth it because this will help me in the work I want to do later on	–0.21	–3.46	–0.39	3.85
ST35Q02	b) What I learn in my <school science> subject(s) is important for me because I need this for what I want to study later on	0.24	–3.62	–0.17	3.79
ST35Q03	c) I study <school science> because I know it is useful for me	–0.37	–3.66	–0.67	4.33
ST35Q04	d) Studying my <school science> subject(s) is worthwhile for me because what I learn will improve my career prospects	0.00	–3.66	–0.45	4.11
ST35Q05	e) I will learn many things in my <school science> subject(s) that will help me get a job	0.34	–3.76	–0.29	4.05

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.



Expectations about tertiary science studies and working in science-related careers are another important aspect of student motivations to learning science. Four items measuring students' motivations to take up a science-related career were included in the student questionnaire. All items are reverse scored so that positive WLE scores on this index indicate higher levels of motivation to take up a science-related career.

Table 16.13
Item parameters for future-oriented science motivation (SCIEFUT)

Item	How much do you agree with the statements below?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST29Q01	a) I would like to work in a career involving <broad science>	–0.77	–3.58	0.34	3.24
ST29Q02	b) I would like to study <broad science> after <secondary school>	–0.27	–3.57	0.44	3.13
ST29Q03	c) I would like to spend my life doing advanced <broad science>	0.71	–3.81	0.57	3.23
ST29Q04	d) I would like to work on <broad science> projects as an adult	0.33	–3.78	0.30	3.48

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Table 16.14
Model fit and estimated latent correlations for motivation to learn science[1]

		Model fit			Latent correlations between:
		RMSEA	RMR	CFI	NNFI	INSTSCIE/SCIEFUT
OECD	Australia	0.130	0.028	0.95	0.95	0.79
	Austria	0.064	0.028	0.98	0.98	0.59
	Belgium	0.079	0.018	0.98	0.98	0.82
	Canada	0.092	0.023	0.98	0.98	0.78
	Czech Republic	0.089	0.019	0.97	0.97	0.65
	Denmark	0.065	0.019	0.99	0.99	0.71
	Finland	0.095	0.023	0.97	0.97	0.73
	France	0.132	0.038	0.94	0.94	0.80
	Germany	0.064	0.018	0.98	0.98	0.67
	Greece	0.089	0.022	0.97	0.97	0.72
	Hungary	0.071	0.018	0.98	0.98	0.67
	Iceland	0.070	0.016	0.99	0.99	0.77
	Ireland	0.112	0.027	0.96	0.96	0.79
	Italy	0.059	0.017	0.99	0.99	0.73
	Japan	0.106	0.019	0.97	0.97	0.74
	Korea	0.116	0.021	0.95	0.95	0.68
	Luxembourg	0.062	0.023	0.98	0.98	0.69
	Mexico	0.069	0.020	0.97	0.97	0.58
	Netherlands	0.080	0.018	0.98	0.98	0.60
	New Zealand	0.121	0.031	0.95	0.95	0.79
	Norway	0.061	0.019	0.99	0.99	0.66
	Poland	0.061	0.015	0.98	0.98	0.59
	Portugal	0.129	0.033	0.94	0.94	0.73
	Slovak Republic	0.105	0.018	0.96	0.96	0.71
	Spain	0.097	0.027	0.97	0.97	0.78
	Sweden	0.071	0.022	0.98	0.98	0.71
	Switzerland	0.078	0.031	0.97	0.97	0.70
	Turkey	0.100	0.022	0.96	0.96	0.63
	United Kingdom	0.117	0.026	0.95	0.95	0.73
	United States	0.078	0.021	0.98	0.98	0.67
	OECD	**0.086**	**0.020**	**0.97**	**0.96**	**0.72**

1. Model estimates based on international student calibration sample (500 students per OECD country).



The fit for the two-factor model was satisfactory for the pooled OECD sample (RMSEA = 0.086) and in most country sub-samples. The latent correlation between the two construct ranges is quite high and ranges between 0.59 and 0.82.

Table 16.15 shows that the reliabilities for both scales are highly satisfactory around 0.90 in most countries.

Table 16.15
Scale reliabilities for instrumental and future-oriented science motivation

		INTSCIE	SCIEFUT			INTSCIE	SCIEFUT
OECD	Australia	0.95	0.93	Partners	Argentina	0.88	0.90
	Austria	0.91	0.90		Azerbaijan	0.86	0.88
	Belgium	0.92	0.93		Brazil	0.86	0.89
	Canada	0.94	0.94		Bulgaria	0.87	0.89
	Czech Republic	0.89	0.92		Chile	0.91	0.91
	Denmark	0.92	0.95		Colombia	0.88	0.90
	Finland	0.92	0.91		Croatia	0.92	0.91
	France	0.91	0.92		Estonia	0.85	0.87
	Germany	0.90	0.91		Hong Kong-China	0.94	0.93
	Greece	0.89	0.93		Indonesia	0.86	0.83
	Hungary	0.88	0.91		Israel	0.92	0.93
	Iceland	0.95	0.94		Jordan	0.81	0.82
	Ireland	0.93	0.92		Kyrgyzstan	0.84	0.82
	Italy	0.88	0.90		Latvia	0.85	0.89
	Japan	0.94	0.94		Liechtenstein	0.91	0.90
	Korea	0.93	0.92		Lithuania	0.89	0.91
	Luxembourg	0.92	0.92		Macao-China	0.91	0.91
	Mexico	0.86	0.88		Montenegro	0.90	0.91
	Netherlands	0.93	0.93		Qatar	0.87	0.89
	New Zealand	0.94	0.93		Romania	0.86	0.87
	Norway	0.92	0.93		Russian Federation	0.88	0.86
	Poland	0.91	0.91		Serbia	0.89	0.90
	Portugal	0.94	0.91		Slovenia	0.91	0.90
	Slovak Republic	0.90	0.94		Chinese Taipei	0.92	0.94
	Spain	0.92	0.93		Thailand	0.84	0.87
	Sweden	0.93	0.93		Tunisia	0.82	0.80
	Switzerland	0.91	0.90		Uruguay	0.91	0.93
	Turkey	0.91	0.93				
	United Kingdom	0.92	0.92				
	United States	0.91	0.92				
	Median	**0.92**	**0.92**		**Median**	**0.88**	**0.90**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Table 16.16
Item parameters for science self-efficacy (SCIEEFF)

Item	How easy do you think it would be for you to perform the following tasks on your own?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST17Q01	a) Recognise the science question that underlies a newspaper report on a health issue	–0.28	–1.93	–0.49	2.41
ST17Q02	b) Explain why earthquakes occur more frequently in some areas than in others	–0.63	–1.49	–0.15	1.64
ST17Q03	c) Describe the role of antibiotics in the treatment of disease	0.17	–1.55	–0.12	1.68
ST17Q04	d) Identify the science question associated with the disposal of garbage	0.09	–1.80	–0.21	2.02
ST17Q05	e) Predict how changes to an environment will affect the survival of certain species	–0.05	–1.48	–0.19	1.67
ST17Q06	f) Interpret the scientific information provided on the labelling of food items	–0.05	–1.61	–0.17	1.78
ST17Q07	g) Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars	0.49	–1.43	–0.14	1.57
ST17Q08	h) Identify the better of two explanations for the formation of acid rain	0.25	–1.46	–0.16	1.62

Note: Item categories were "I could do this easily", "I could do this with a bit of effort", "I would struggle to do this on my own" and "I couldn't do this"; all items were inverted for scaling.

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



Self-related cognitions in science

Eight items measuring students' science self-efficacy (their confidence in performing science-related tasks) were included. These items cover important themes identified in the science literacy framework: identifying scientific questions, explaining phenomena scientifically and using scientific evidence. All items are reverse coded for IRT scaling so that positive WLE scores on this new index for PISA 2006 indicate higher levels of self-efficacy in science.

Six items on science self-concept were included in the student questionnaire. The items were inverted for scaling so that positive WLE scores on this new PISA 2006 index indicate a positive self-concept in science.

Table 16.17
Item parameters for science self-concept (SCSCIE)

Item	How much do you agree with the statements below? (Strongly agree/Agree/Disagree/Strongly disagree)	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST37Q01	a) Learning advanced <school science> topics would be easy for me	0.56	–4.23	–0.07	4.29
ST37Q02	b) I can usually give good answers to <test questions> on <school science> topics	–0.55	–4.35	–0.47	4.82
ST37Q03	c) I learn <school science> topics quickly	–0.19	–4.30	–0.09	4.38
ST37Q04	d) <School science> topics are easy for me	0.41	–4.35	0.13	4.23
ST37Q05	e) When I am being taught <school science>. I can understand the concepts very well	–0.22	–4.32	–0.3	4.62
ST37Q06	f) I can easily understand new ideas in <school science>	–0.01	–4.32	–0.16	4.49

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Table 16.18
Model fit and estimated latent correlations for science self-efficacy and science self-concept[1]

		RMSEA	Model fit		NNFI	Latent correlations between:
			RMR	CFI		SCIEEFF/SCSCIE
OECD	Australia	0.074	0.029	0.95	0.95	0.66
	Austria	0.068	0.034	0.94	0.94	0.58
	Belgium	0.060	0.032	0.96	0.96	0.58
	Canada	0.053	0.028	0.97	0.98	0.54
	Czech Republic	0.044	0.026	0.96	0.97	0.44
	Denmark	0.062	0.027	0.96	0.96	0.67
	Finland	0.039	0.020	0.98	0.98	0.61
	France	0.049	0.030	0.96	0.96	0.49
	Germany	0.035	0.024	0.98	0.98	0.65
	Greece	0.049	0.038	0.96	0.96	0.45
	Hungary	0.043	0.027	0.97	0.97	0.35
	Iceland	0.054	0.028	0.97	0.97	0.64
	Ireland	0.059	0.031	0.96	0.96	0.66
	Italy	0.046	0.028	0.96	0.97	0.45
	Japan	0.059	0.024	0.97	0.97	0.49
	Korea	0.066	0.027	0.95	0.95	0.46
	Luxembourg	0.054	0.029	0.96	0.96	0.59
	Mexico	0.057	0.026	0.95	0.95	0.39
	Netherlands	0.055	0.028	0.96	0.96	0.52
	New Zealand	0.049	0.021	0.97	0.97	0.63
	Norway	0.042	0.021	0.98	0.98	0.53
	Poland	0.035	0.020	0.98	0.98	0.43
	Portugal	0.054	0.024	0.96	0.96	0.34
	Slovak Republic	0.062	0.030	0.94	0.94	0.42
	Spain	0.055	0.032	0.97	0.97	0.46
	Sweden	0.056	0.027	0.97	0.97	0.57
	Switzerland	0.033	0.024	0.99	0.99	0.57
	Turkey	0.064	0.032	0.95	0.95	0.43
	United Kingdom	0.050	0.023	0.97	0.97	0.64
	United States	0.046	0.025	0.98	0.98	0.60
	OECD	**0.041**	**0.017**	**0.98**	**0.98**	**0.55**

1. Model estimates based on international student calibration sample (500 students per OECD country).



Table 16.18 shows the results of confirmatory factor analyses (CFA) for a two-dimensional model of self-efficacy and self-concept items. The model fit is very well for the pooled OECD sample and also for all country sub-samples. The estimated latent correlation between the two constructs is moderately high and ranges between 0.35 and 0.67.

Table 16.19 shows internal consistencies for the two scales. Both constructs have high reliabilities across participating countries, for *SCIEEFF* the reliabilities are typically around 0.80 and for *SCSCIE* even higher (around 0.90).

Table 16.19
Scale reliabilities for science self-efficacy and science self-concept

		SCIEEFF	SCSCIE			SCIEEFF	SCSCIE
OECD	Australia	0.88	0.93	*Partners*	Argentina	0.76	0.89
	Austria	0.80	0.90		Azerbaijan	0.78	0.88
	Belgium	0.82	0.91		Brazil	0.79	0.86
	Canada	0.85	0.94		Bulgaria	0.81	0.87
	Czech Republic	0.78	0.88		Chile	0.81	0.89
	Denmark	0.84	0.94		Colombia	0.77	0.87
	Finland	0.83	0.92		Croatia	0.79	0.89
	France	0.79	0.91		Estonia	0.76	0.86
	Germany	0.82	0.90		Hong Kong-China	0.83	0.93
	Greece	0.77	0.90		Indonesia	0.73	0.86
	Hungary	0.76	0.88		Israel	0.84	0.92
	Iceland	0.88	0.94		Jordan	0.75	0.83
	Ireland	0.82	0.93		Kyrgyzstan	0.76	0.82
	Italy	0.75	0.89		Latvia	0.74	0.82
	Japan	0.85	0.93		Liechtenstein	0.85	0.93
	Korea	0.83	0.92		Lithuania	0.77	0.86
	Luxembourg	0.83	0.91		Macao-China	0.80	0.92
	Mexico	0.77	0.86		Montenegro	0.77	0.87
	Netherlands	0.84	0.91		Qatar	0.85	0.88
	New Zealand	0.87	0.92		Romania	0.79	0.84
	Norway	0.87	0.92		Russian Federation	0.79	0.84
	Poland	0.82	0.88		Serbia	0.78	0.90
	Portugal	0.84	0.91		Slovenia	0.80	0.90
	Slovak Republic	0.77	0.88		Chinese Taipei	0.85	0.93
	Spain	0.83	0.92		Thailand	0.79	0.87
	Sweden	0.87	0.93		Tunisia	0.66	0.82
	Switzerland	0.82	0.92		Uruguay	0.78	0.90
	Turkey	0.81	0.92				
	United Kingdom	0.85	0.91				
	United States	0.87	0.93				
	Median	**0.83**	**0.92**		**Median**	**0.79**	**0.87**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Table 16.20
Item parameters for general value of science (GENSCIE)

Item	How much do you agree with the statements below? (Strongly agree/Agree/Disagree/Strongly disagree)	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST18Q01	a) Advances in <broad science and technology> usually improve people's living conditions	–0.42	–1.68	–0.96	2.64
ST18Q02	b) <Broad science> is important for helping us to understand the natural world	–0.52	–1.71	–0.92	2.63
ST18Q04	d) Advances in <broad science and technology> usually help improve the economy	0.31	–2.37	–0.45	2.82
ST18Q06	f) <Broad science> is valuable to society	0.04	–1.93	–0.80	2.73
ST18Q09	i) Advances in <broad science and technology> usually bring social benefits	0.60	–2.43	–0.36	2.79

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.



Value of science

Five items measuring perceptions of the general value of science were included in the student questionnaire. The items are reverse coded for scaling so that positive WLE scores on this new PISA 2006 index indicate positive students' perceptions of the general value of science. Table 16.20 shows the item wording and international IRT parameters that were used for scaling.

Five items measuring perceptions of the personal value of science were included in the student questionnaire. The items were inverted for scaling so that positive WLE scores on this new PISA 2006 index indicate positive students' perceptions of the general value of science. Table 16.21 shows the item wording and international IRT parameters used for scaling.

Table 16.21
Item parameters for personal value of science (PERSCIE)

Item	How much do you agree with the statements below?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST18Q03	c) Some concepts in <broad science> help me see how I relate to other people	0.05	–2.97	–0.05	3.02
ST18Q05	e) I will use <broad science> in many ways when I am an adult	–0.02	–2.52	–0.21	2.74
ST18Q07	g) <Broad science> is very relevant to me	0.26	–2.36	–0.08	2.44
ST18Q08	h) I find that <broad science> helps me to understand the things around me	–0.52	–2.36	–0.45	2.81
ST18Q10	j) When I leave school there will be many opportunities for me to use <broad science>	0.24	–2.35	–0.18	2.53

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Table 16.22
Model fit and estimated latent correlations for general and personal value of science[1]

		Model fit				Latent correlations between:
		RMSEA	RMR	CFI	NNFI	GENSCIE/PERSCIE
OECD	Australia	0.090	0.029	0.93	0.94	0.75
	Austria	0.085	0.033	0.92	0.92	0.83
	Belgium	0.066	0.025	0.95	0.95	0.77
	Canada	0.083	0.027	0.94	0.94	0.82
	Czech Republic	0.077	0.025	0.92	0.92	0.74
	Denmark	0.062	0.022	0.96	0.96	0.71
	Finland	0.101	0.026	0.90	0.90	0.72
	France	0.070	0.026	0.95	0.95	0.72
	Germany	0.086	0.033	0.93	0.93	0.77
	Greece	0.059	0.026	0.92	0.93	0.71
	Hungary	0.083	0.031	0.91	0.91	0.74
	Iceland	0.107	0.039	0.92	0.92	0.79
	Ireland	0.078	0.030	0.94	0.95	0.75
	Italy	0.072	0.023	0.93	0.93	0.77
	Japan	0.092	0.031	0.92	0.92	0.79
	Korea	0.076	0.027	0.93	0.93	0.63
	Luxembourg	0.091	0.037	0.91	0.91	0.74
	Mexico	0.068	0.021	0.93	0.93	0.86
	Netherlands	0.062	0.018	0.96	0.96	0.74
	New Zealand	0.099	0.032	0.93	0.93	0.81
	Norway	0.077	0.025	0.95	0.95	0.79
	Poland	0.095	0.026	0.90	0.90	0.76
	Portugal	0.074	0.017	0.94	0.94	0.83
	Slovak Republic	0.054	0.018	0.96	0.96	0.69
	Spain	0.080	0.027	0.93	0.93	0.77
	Sweden	0.102	0.032	0.93	0.93	0.84
	Switzerland	0.064	0.026	0.96	0.96	0.77
	Turkey	0.090	0.029	0.91	0.92	0.75
	United Kingdom	0.084	0.027	0.94	0.94	0.77
	United States	0.098	0.030	0.93	0.93	0.77
	OECD	**0.076**	**0.023**	**0.94**	**0.94**	**0.78**

1. Model estimates based on international student calibration sample (500 students per OECD country).

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



Table 16.23 shows the results of a CFA for general and personal value of science items. The model fit is satisfactory for the pooled sample and in all but three country sub-samples. Not unexpectedly, the estimated latent correlation between the two construct is quite high and ranges between 0.63 and 0.86.

Table 16.23 shows the scale reliabilities for general and personal value of science. For both constructs, the internal consistencies are high across participating countries. However, reliabilities for *GENSCIE* are somewhat lower in many partner countries.

Table 16.23
Scale reliabilities for general and personal value of science

		GENSCIE	PERSCIE			GENSCIE	PERSCIE
OECD	Australia	0.81	0.86	Partners	Argentina	0.69	0.77
	Austria	0.72	0.80		Azerbaijan	0.68	0.67
	Belgium	0.70	0.78		Brazil	0.67	0.75
	Canada	0.78	0.85		Bulgaria	0.73	0.76
	Czech Republic	0.71	0.79		Chile	0.72	0.78
	Denmark	0.70	0.85		Colombia	0.61	0.71
	Finland	0.76	0.83		Croatia	0.69	0.79
	France	0.68	0.80		Estonia	0.65	0.74
	Germany	0.75	0.81		Hong Kong-China	0.80	0.79
	Greece	0.66	0.74		Indonesia	0.62	0.66
	Hungary	0.67	0.77		Israel	0.79	0.83
	Iceland	0.80	0.87		Jordan	0.69	0.69
	Ireland	0.75	0.83		Kyrgyzstan	0.65	0.69
	Italy	0.68	0.73		Latvia	0.65	0.73
	Japan	0.80	0.76		Liechtenstein	0.79	0.84
	Korea	0.77	0.75		Lithuania	0.70	0.77
	Luxembourg	0.79	0.83		Macao-China	0.72	0.73
	Mexico	0.65	0.71		Montenegro	0.68	0.78
	Netherlands	0.78	0.78		Qatar	0.81	0.82
	New Zealand	0.79	0.85		Romania	0.69	0.71
	Norway	0.82	0.85		Russian Federation	0.64	0.77
	Poland	0.71	0.80		Serbia	0.68	0.76
	Portugal	0.74	0.79		Slovenia	0.74	0.81
	Slovak Republic	0.71	0.76		Chinese Taipei	0.82	0.79
	Spain	0.72	0.79		Thailand	0.72	0.72
	Sweden	0.82	0.85		Tunisia	0.64	0.62
	Switzerland	0.73	0.80		Uruguay	0.68	0.80
	Turkey	0.79	0.81				
	United Kingdom	0.78	0.83				
	United States	0.82	0.84				
	Median	**0.75**	**0.80**		**Median**	**0.69**	**0.76**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Science-related activities

Student participation in non-compulsory activities related to science or choice of course combinations with an emphasis on this subject are important indicators of engagement. Furthermore, out-of-school activities relating to science can contribute considerably to students' engagement and learning in science.

Six items measuring students' activities related to science were included in the student questionnaire. The items are reverse scored for scaling so that positive WLE scores on this new PISA 2006 index indicate higher frequencies of students' science activities. Table 16.24 shows the item wording and the international IRT parameters used for scaling.

Table 16.24
Item parameters for science activities (SCIEACT)

Item	How often do you do these things? (Very often/Regularly/Sometimes/Never or hardly ever)	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST19Q01	a) Watch TV programmes about <broad science>	–1.99	–2.50	0.82	1.68
ST19Q02	b) Borrow or buy books on <broad science> topics	0.31	–1.72	0.60	1.12
ST19Q03	c) Visit web sites about <broad science> topics	–0.17	–1.75	0.59	1.17
ST19Q04	d) Listen to radio programmes about advances in <broad science>	0.58	–1.45	0.43	1.03
ST19Q05	e) Read <broad science> magazines or science articles in newspapers	–0.68	–1.89	0.51	1.29
ST19Q06	f) Attend a <science club>	0.96	–0.21	–0.02	0.23

Note: Item categories were "very often", "regularly", "sometimes" and "never or hardly ever"; all items were inverted for scaling.



Table 16.25 shows the scale reliabilities across countries, which are satisfactory and range typically between 0.75 and 0.80 in a majority of countries.

Table 16.25
Scale reliabilities for the science activities index

		SCIEACT			SCIEACT
OECD	Australia	0.80	Partners	Argentina	0.77
	Austria	0.76		Azerbaijan	0.71
	Belgium	0.77		Brazil	0.80
	Canada	0.80		Bulgaria	0.75
	Czech Republic	0.77		Chile	0.81
	Denmark	0.79		Colombia	0.76
	Finland	0.76		Croatia	0.78
	France	0.75		Estonia	0.75
	Germany	0.77		Hong Kong-China	0.84
	Greece	0.82		Indonesia	0.71
	Hungary	0.77		Israel	0.88
	Iceland	0.81		Jordan	0.67
	Ireland	0.79		Kyrgyzstan	0.76
	Italy	0.76		Latvia	0.76
	Japan	0.80		Liechtenstein	0.78
	Korea	0.80		Lithuania	0.75
	Luxembourg	0.80		Macao-China	0.80
	Mexico	0.78		Montenegro	0.75
	Netherlands	0.78		Qatar	0.83
	New Zealand	0.78		Romania	0.76
	Norway	0.81		Russian Federation	0.76
	Poland	0.76		Serbia	0.73
	Portugal	0.80		Slovenia	0.81
	Slovak Republic	0.75		Chinese Taipei	0.84
	Spain	0.78		Thailand	0.77
	Sweden	0.79		Tunisia	0.60
	Switzerland	0.78		Uruguay	0.78
	Turkey	0.82			
	United Kingdom	0.78			
	United States	0.80			
	Median	**0.78**		**Median**	**0.76**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Scientific literacy and environment

Five items measuring students' awareness of environmental issues were included in the student questionnaire. Positive WLE scores on this index indicate higher levels of students' awareness of environmental issues. Table 16.26 shows the item wording and international IRT parameters for this scale.

Table 16.26
Item parameters for awareness of environmental issues (ENVAWARE)

		Parameter estimates			
Item	**How informed are you about the following environmental issues?**	**Delta**	**Tau(1)**	**Tau(2)**	**Tau(3)**
ST22Q01	a) The increase of greenhouse gases in the atmosphere	–0.05	–1.87	–0.01	1.88
ST22Q02	b) Use of genetically modified organisms (<GMO>)	0.88	–2.03	0.14	1.88
ST22Q03	c) Acid rain	–0.16	–2.13	0.07	2.07
ST22Q04	d) Nuclear waste	0.02	–2.47	0.23	2.25
ST22Q05	e) The consequences of clearing forests for other land use	–0.68	–1.56	–0.05	1.61

Six items measuring students' perception of environmental issues as a concern were included in the student questionnaire. The items were reverse scored for scaling so that positive WLE scores on this index indicate higher levels of students' concerns about environmental issues. Table 16.27 shows the item wording and the international IRT parameters for this scale.



Table 16.27
Item parameters for perception of environmental issues (ENVPERC)

Item	Do you see the environmental issues below as a serious concern for yourself and/or others?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST24Q01	a) Air pollution	–0.6	–0.81	–0.03	0.84
ST24Q02	b) Energy shortages	0.12	–1.53	0.13	1.40
ST24Q03	c) Extinction of plants and animals	0.11	–1.06	–0.02	1.08
ST24Q04	d) Clearing of forests for other land use	0.07	–1.5	0.21	1.29
ST24Q05	e) Water shortages	–0.06	–2.09	1.33	0.76
ST24Q06	f) Nuclear waste	0.35	–1.61	0.23	1.38

Students' optimism regarding environmental issues was measured by six items in the student questionnaire. The items were inverted for scaling so that positive WLE scores on this index indicate higher levels of students' optimism about environmental issues. Table 16.28 shows the item wording and the international IRT parameters for this scale.

Table 16.28
Item parameters for environmental optimism (ENVOPT)

Item	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years?	Parameter estimates		
		Delta	Tau(1)	Tau(2)
ST25Q01	a) Air pollution	0.20	0.05	–0.05
ST25Q02	b) Energy shortages	–0.45	–0.71	0.71
ST25Q03	c) Extinction of plants and animals	0.18	–0.57	0.57
ST25Q04	d) Clearing of forests for other land use	0.32	–0.37	0.37
ST25Q05	e) Water shortages	–0.25	–0.75	0.75
ST25Q06	f) Nuclear waste	0.00	–0.73	0.73

Seven items measuring students' responsibility for sustainable development were included in the student questionnaire. The items were reverse coded for scaling so that positive WLE scores on this new PISA 2006 index indicate higher levels of students' responsibility for sustainable development. Table 16.29 shows the item wording and the international IRT parameters for this scale.

Table 16.29
Item parameters for responsibility for sustainable development (RESPDEV)

Item	How much do you agree with the statements below?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST26Q01	a) It is important to carry out regular checks on the emissions from cars as a condition of their use	–0.42	–1.45	–0.86	2.32
ST26Q02	b) It disturbs me when energy is wasted through the unnecessary use of electrical appliances	0.61	–1.94	–0.05	1.99
ST26Q03	c) I am in favour of having laws that regulate factory emissions even if this would increase the price of products	0.65	–1.91	–0.06	1.97
ST26Q04	d) To reduce waste, the use of plastic packaging should be kept to a minimum	0.06	–1.79	–0.35	2.14
ST26Q05	e) Industries should be required to prove that they safely dispose of dangerous waste materials	–0.59	–1.21	–0.76	1.97
ST26Q06	f) I am in favour of having laws that protect the habitats of endangered species	–0.58	–1.12	–0.68	1.80
ST26Q07	g) Electricity should be produced from renewable sources as much as possible, even if this increases the cost	0.28	–1.70	–0.32	2.02

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.



Table 16.30 shows the model fit for a four-dimensional model for the environment-related items in PISA 2006. The model fit is satisfactory across participating countries and for the pooled OECD sample.

Table 16.30
Model fit environment-related constructs[1]

		Model fit			
		RMSEA	RMR	CFI	NNFI
OECD	Australia	0.056	0.027	0.90	0.90
OECD	Austria	0.043	0.030	0.90	0.90
OECD	Belgium	0.049	0.032	0.86	0.87
OECD	Canada	0.052	0.028	0.90	0.91
OECD	Czech Republic	0.051	0.030	0.86	0.86
OECD	Denmark	0.046	0.033	0.90	0.90
OECD	Finland	0.049	0.033	0.91	0.91
OECD	France	0.044	0.030	0.90	0.91
OECD	Germany	0.055	0.036	0.85	0.85
OECD	Greece	0.045	0.031	0.89	0.89
OECD	Hungary	0.041	0.025	0.90	0.90
OECD	Iceland	0.049	0.040	0.91	0.91
OECD	Ireland	0.046	0.032	0.91	0.92
OECD	Italy	0.055	0.032	0.83	0.83
OECD	Japan	0.043	0.023	0.94	0.94
OECD	Korea	0.041	0.020	0.93	0.93
OECD	Luxembourg	0.047	0.036	0.92	0.92
OECD	Mexico	0.049	0.025	0.89	0.90
OECD	Netherlands	0.050	0.030	0.87	0.87
OECD	New Zealand	0.057	0.032	0.89	0.89
OECD	Norway	0.057	0.041	0.89	0.89
OECD	Poland	0.051	0.028	0.89	0.89
OECD	Portugal	0.049	0.022	0.91	0.91
OECD	Slovak Republic	0.050	0.032	0.88	0.88
OECD	Spain	0.041	0.022	0.93	0.93
OECD	Sweden	0.047	0.036	0.92	0.92
OECD	Switzerland	0.051	0.035	0.87	0.87
OECD	Turkey	0.048	0.025	0.93	0.93
OECD	United Kingdom	0.049	0.029	0.92	0.92
OECD	United States	0.056	0.032	0.91	0.91
	OECD	**0.044**	**0.021**	**0.92**	**0.92**

1. Model estimates based on international student calibration sample (500 students per OECD country).

Table 16.31
Estimated latent correlations for environment-related constructs[1]

		Latent correlations between					
		RESPDEV/ ENVAWARE	RESPDEV/ENVPERC	RESPDEV/ENVOPT	ENVAWARE/ ENVPERC	ENVAWARE/ ENVOPT	ENVPERC/ENVOPT
OECD	Australia	0.42	0.44	-0.12	0.23	-0.12	-0.15
OECD	Austria	0.29	0.26	-0.15	0.09	-0.15	-0.16
OECD	Belgium	0.30	0.42	-0.13	0.17	-0.14	-0.11
OECD	Canada	0.39	0.39	-0.15	0.18	-0.09	-0.15
OECD	Czech Republic	0.46	0.22	-0.08	-0.06	-0.06	-0.16
OECD	Denmark	0.37	0.31	-0.10	0.03	-0.04	-0.09
OECD	Finland	0.42	0.48	-0.32	0.18	-0.20	-0.13
OECD	France	0.39	0.43	-0.25	0.26	-0.25	-0.19
OECD	Germany	0.31	0.52	-0.06	0.19	-0.08	-0.17
OECD	Greece	0.34	0.66	-0.24	0.18	-0.28	-0.23
OECD	Hungary	0.37	0.52	-0.22	0.08	-0.22	-0.12
OECD	Iceland	0.38	0.20	-0.13	-0.13	-0.07	0.05
OECD	Ireland	0.55	0.37	-0.15	0.09	-0.07	-0.04
OECD	Italy	0.32	0.49	-0.24	0.16	-0.17	-0.26
OECD	Japan	0.45	0.60	-0.02	0.32	0.08	-0.07
OECD	Korea	0.31	0.43	-0.10	0.24	-0.06	-0.03
OECD	Luxembourg	0.27	0.47	-0.23	0.00	-0.16	-0.30
OECD	Mexico	0.28	0.43	-0.10	0.06	-0.06	-0.15
OECD	Netherlands	0.31	0.33	-0.09	0.04	-0.18	-0.17
OECD	New Zealand	0.48	0.43	-0.12	0.14	-0.15	-0.07
OECD	Norway	0.48	0.41	-0.05	0.19	-0.08	-0.01
OECD	Poland	0.35	0.33	-0.05	0.01	-0.13	-0.01
OECD	Portugal	0.44	0.23	-0.29	0.15	-0.34	-0.17
OECD	Slovak Republic	0.46	0.20	-0.17	-0.05	-0.26	-0.15
OECD	Spain	0.43	0.45	-0.18	0.21	-0.30	-0.17
OECD	Sweden	0.34	0.32	-0.18	0.10	-0.18	-0.17
OECD	Switzerland	0.50	0.35	-0.22	0.19	-0.15	-0.09
OECD	Turkey	0.21	0.34	-0.05	0.09	-0.27	-0.04
OECD	United Kingdom	0.40	0.33	-0.19	0.15	-0.10	-0.03
OECD	United States	0.31	0.34	-0.08	0.02	-0.05	0.02
	OECD	**0.34**	**0.44**	**-0.13**	**0.12**	**-0.14**	**-0.11**

1. Model estimates based on international student calibration sample (500 students per OECD country).



Table 16.33 shows the estimated latent correlations for the four environment-related constructs. The highest correlations (0.44 for the pooled sample) are found for *RESPDEV* and *ENVPERC*. Environmental optimism has (weak) negative correlations with all other constructs.

Table 16.32
Scale reliabilities for environment-related scales in OECD countries

	ENVAWARE	ENVPERC	ENVOPT	RESPDEV
OECD Australia	0.79	0.85	0.79	0.80
Austria	0.76	0.77	0.68	0.75
Belgium	0.75	0.74	0.74	0.77
Canada	0.77	0.84	0.79	0.82
Czech Republic	0.73	0.74	0.73	0.72
Denmark	0.76	0.81	0.72	0.79
Finland	0.75	0.78	0.75	0.83
France	0.73	0.71	0.76	0.76
Germany	0.77	0.78	0.69	0.76
Greece	0.66	0.71	0.77	0.71
Hungary	0.68	0.72	0.76	0.74
Iceland	0.79	0.82	0.72	0.82
Ireland	0.76	0.82	0.73	0.76
Italy	0.73	0.68	0.74	0.70
Japan	0.79	0.84	0.79	0.81
Korea	0.75	0.81	0.78	0.78
Luxembourg	0.78	0.82	0.78	0.80
Mexico	0.74	0.76	0.85	0.70
Netherlands	0.74	0.75	0.74	0.76
New Zealand	0.79	0.84	0.79	0.79
Norway	0.78	0.85	0.79	0.84
Poland	0.77	0.81	0.79	0.79
Portugal	0.79	0.77	0.84	0.77
Slovakia	0.74	0.74	0.77	0.71
Spain	0.77	0.79	0.78	0.75
Sweden	0.78	0.85	0.76	0.82
Switzerland	0.75	0.76	0.74	0.79
Turkey	0.72	0.85	0.87	0.84
United Kingdom	0.79	0.84	0.80	0.81
United States	0.79	0.88	0.83	0.80
Median	**0.76**	**0.80**	**0.77**	**0.78**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Table 16.33
Scale reliabilities for environment-related scales in non-OECD countries

	ENVAWARE	ENVPERC	ENVOPT	RESPDEV
OECD Argentina	0.72	0.75	0.84	0.69
Azerbaijan	0.74	0.77	0.85	0.72
Brazil	0.77	0.80	0.88	0.68
Bulgaria	0.75	0.81	0.85	0.72
Chile	0.74	0.73	0.82	0.71
Colombia	0.74	0.79	0.87	0.64
Croatia	0.75	0.77	0.80	0.69
Estonia	0.70	0.72	0.76	0.72
Hong Kong-China	0.72	0.80	0.78	0.75
Indonesia	0.64	0.81	0.80	0.59
Israel	0.78	0.83	0.83	0.85
Jordan	0.66	0.78	0.83	0.73
Kyrgyzstan	0.71	0.83	0.82	0.68
Latvia	0.67	0.64	0.72	0.64
Liechtenstein	0.72	0.81	0.76	0.81
Lithuania	0.71	0.73	0.78	0.71
Macao-China	0.70	0.84	0.80	0.70
Montenegro	0.76	0.76	0.83	0.71
Qatar	0.77	0.83	0.82	0.81
Romania	0.71	0.81	0.80	0.69
Russian Federation	0.73	0.75	0.78	0.68
Serbia	0.75	0.77	0.84	0.73
Slovenia	0.73	0.79	0.77	0.76
Chinese Taipei	0.81	0.93	0.84	0.80
Thailand	0.73	0.80	0.86	0.72
Tunisia	0.55	0.64	0.73	0.64
Uruguay	0.73	0.75	0.84	0.72
Median	**0.73**	**0.79**	**0.82**	**0.71**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

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



Table 16.32 shows the scale reliabilities for environment-related scale in OECD countries, Table 16.3 those for partner countries. For all four constructs the internal consistencies are generally satisfactory across participating countries. Only in few countries scale reliabilities are below 0.70.

Science career preparation

Four items measuring students' perceptions of the usefulness of schooling as preparation for science-related careers were included in the student questionnaire. All items were inverted so that positive WLE scores on this index indicate higher levels of agreement with usefulness of schooling for this purpose. Item wording and international IRT parameter are shown in Table 16.34.

Table 16.34
Item parameters for school preparation for science career (CARPREP)

Item	How much do you agree with the statements below?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST27Q01	a) The subjects available at my school provide students with the basic skills and knowledge for a <science-related career>	–0.38	–2.81	–0.76	3.57
ST27Q02	b) The <school science> subjects at my school provide students with the basic skills and knowledge for many different careers	–0.26	–2.96	–0.61	3.57
ST27Q03	c) The subjects I study provide me with the basic skills and knowledge for a <science-related career>	0.28	–2.86	–0.37	3.23
ST27Q04	d) My teachers equip me with the basic skills and knowledge I need for a <science-related career>	0.35	–2.65	–0.59	3.24

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Four items measuring students' perceptions of being informed about science-related careers are included in the student questionnaire. Items were reverse coded so that positive WLE scores on this index indicate higher levels of information about science-related careers. Table 16.35 shows the wording of items and the international IRT parameters used for scaling.

Table 16.35
Item parameters for student information on science careers (CARINFO)

Item	How informed are you about these topics?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST28Q01	a) <Science-related careers> that are available in the job market	–0.02	–3.34	0.06	3.28
ST28Q02	b) Where to find information about <science-related careers>	–0.35	–3.04	0.06	2.98
ST28Q03	c) The steps a student needs to take if they want a <science-related career>	–0.19	–2.82	0.07	2.75
ST28Q04	d) Employers or companies that hire people to work in <science-related careers>	0.57	–3.03	0.16	2.87

Note: Item categories were "Very well informed", "Fairly informed", "Not well informed" and "Not informed at all"; all items were inverted for scaling.

Table 16.36 shows the results of a CFA for the items related to science career preparation. The model fit is satisfactory for the pooled sample and in most country sub-samples. The estimated latent correlation between the two constructs is moderate to high: between 0.26 and 0.57.

Table 16.37 shows the scale reliabilities for *CARINFO* and *CARPREP* across participating countries. For both scales, the internal consistencies are high around 0.80.



Table 16.36
Model fit and estimated latent correlations for science career preparation indices[1]

		Model fit			Latent correlations between:
		RMSEA	RMR	CFI	NNFI	CARPREP/CARINFO
OECD	Australia	0.098	0.027	0.95	0.95	0.54
	Austria	0.060	0.023	0.98	0.98	0.53
	Belgium	0.088	0.025	0.95	0.95	0.44
	Canada	0.082	0.024	0.96	0.96	0.47
	Czech Republic	0.057	0.021	0.98	0.98	0.43
	Denmark	0.047	0.015	0.99	0.99	0.56
	Finland	0.025	0.010	1.00	1.00	0.43
	France	0.066	0.026	0.96	0.96	0.44
	Germany	0.047	0.018	0.98	0.98	0.35
	Greece	0.037	0.020	0.99	0.99	0.48
	Hungary	0.051	0.018	0.96	0.96	0.36
	Iceland	0.060	0.018	0.98	0.98	0.52
	Ireland	0.101	0.033	0.94	0.94	0.51
	Italy	0.045	0.019	0.98	0.98	0.33
	Japan	0.078	0.017	0.97	0.97	0.47
	Korea	0.050	0.015	0.98	0.98	0.26
	Luxembourg	0.060	0.022	0.98	0.98	0.47
	Mexico	0.050	0.024	0.98	0.98	0.43
	Netherlands	0.092	0.022	0.94	0.94	0.40
	New Zealand	0.114	0.028	0.92	0.92	0.44
	Norway	0.057	0.018	0.98	0.98	0.57
	Poland	0.053	0.014	0.98	0.98	0.39
	Portugal	0.108	0.023	0.93	0.93	0.40
	Slovak Republic	0.057	0.018	0.98	0.98	0.41
	Spain	0.078	0.021	0.95	0.95	0.45
	Sweden	0.047	0.019	0.99	0.99	0.47
	Switzerland	0.031	0.014	0.99	0.99	0.47
	Turkey	0.086	0.023	0.96	0.96	0.32
	United Kingdom	0.053	0.020	0.98	0.98	0.48
	United States	0.078	0.021	0.97	0.97	0.45
	OECD	**0.054**	**0.012**	**0.98**	**0.98**	**0.45**

1. Model estimates based on international student calibration sample (500 students per OECD country).

Table 16.37
Scale reliabilities for science career preparation indices

		CARPREP	CARINFO			CARPREP	CARINFO
OECD	Australia	0.81	0.86	Partners	Argentina	0.79	0.80
	Austria	0.83	0.79		Azerbaijan	0.74	0.76
	Belgium	0.81	0.78		Brazil	0.78	0.79
	Canada	0.83	0.84		Bulgaria	0.79	0.80
	Czech Republic	0.81	0.78		Chile	0.80	0.81
	Denmark	0.81	0.84		Colombia	0.79	0.74
	Finland	0.86	0.80		Croatia	0.83	0.78
	France	0.81	0.76		Estonia	0.78	0.76
	Germany	0.83	0.78		Hong Kong-China	0.79	0.77
	Greece	0.74	0.77		Indonesia	0.72	0.76
	Hungary	0.75	0.69		Israel	0.84	0.82
	Iceland	0.84	0.84		Jordan	0.71	0.68
	Ireland	0.79	0.83		Kyrgyzstan	0.70	0.71
	Italy	0.79	0.72		Latvia	0.76	0.75
	Japan	0.83	0.87		Liechtenstein	0.87	0.82
	Korea	0.80	0.78		Lithuania	0.80	0.72
	Luxembourg	0.82	0.81		Macao-China	0.80	0.77
	Mexico	0.75	0.81		Montenegro	0.76	0.80
	Netherlands	0.72	0.82		Qatar	0.81	0.79
	New Zealand	0.81	0.84		Romania	0.75	0.74
	Norway	0.82	0.87		Russian Federation	0.76	0.73
	Poland	0.80	0.82		Serbia	0.79	0.79
	Portugal	0.77	0.82		Slovenia	0.79	0.79
	Slovak Republic	0.81	0.80		Chinese Taipei	0.84	0.77
	Spain	0.78	0.80		Thailand	0.78	0.76
	Sweden	0.85	0.85		Tunisia	0.68	0.67
	Switzerland	0.82	0.78		Uruguay	0.75	0.79
	Turkey	0.88	0.83				
	United Kingdom	0.83	0.85				
	United States	0.82	0.85				
	Median	**0.81**	**0.82**		**Median**	**0.79**	**0.77**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

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



Science learning and teaching

Four items measuring students' reports on the frequency of interactive teaching in science lessons were included in the student questionnaire. Items were inverted such that positive WLE scores on this index indicate higher frequencies of interactive science teaching. Table 16.38 shows the item wording and international IRT parameters used for scaling.

Table 16.38
Item parameters for science teaching: interaction (SCINTACT)

Item	When learning <school science> topics at school, how often do the following activities occur?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST34Q01	a) Students are given opportunities to explain their ideas	–0.64	–1.72	0.11	1.61
ST34Q05	e) The lessons involve students' opinions about the topics	–0.07	–1.72	0.09	1.63
ST34Q09	i) There is a class debate or discussion	0.48	–1.70	0.26	1.44
ST34Q13	m) The students have discussions about the topics	0.23	–1.71	0.08	1.63

Four items measuring students' reports on the frequency of hands-on activities in science lessons are included in the main study. These were reverse scored so that positive WLE scores on this index indicate higher frequencies of this type of science teaching. Table 16.39 shows the item wording and the international item parameters used for scaling.

Table 16.39
Item parameters for science teaching: hands-on activities (SCHANDS)

Item	When learning <school science> topics at school, how often do the following activities occur?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST34Q02	b) Students spend time in the laboratory doing practical experiments	0.57	–2.17	0.38	1.8
ST34Q03	c) Students are required to design how a <school science> question could be investigated in the laboratory	0.64	–1.77	0.15	1.62
ST34Q06	f) Students are asked to draw conclusions from an experiment they have conducted	–0.74	–1.82	–0.02	1.84
ST34Q14	n) Students do experiments by following the instructions of the teacher	–0.47	–1.71	0.05	1.67

Three items measuring students' reports on the frequency of student investigations in science lessons were included in the student questionnaire. Responses were inverted so that positive WLE scores on this index indicate perceived higher frequencies of this type of science teaching. Table 16.40 shows the item wording and the international IRT parameters for this scale.

Table 16.40
Item parameters for science teaching: student investigations (SCINVEST)

Item	When learning <school science> topics at school, how often do the following activities occur?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST34Q08	h) Students are allowed to design their own experiments	0.16	–1.36	0.08	1.28
ST34Q11	k) Students are given the chance to choose their own investigations	0.12	–1.78	0.24	1.53
ST34Q16	p) Students are asked to do an investigation to test out their own ideas	–0.28	–1.88	0.16	1.72

Note: Item categories were "In all lessons", "In most lessons", "In some lessons" and "Never or hardly ever"; all items were inverted for scaling.



Five items measuring students' reports on the frequency of teaching in science lessons with a focus on applications are included in the student questionnaire. All items were reverse scored so that positive WLE scores on this index indicate higher frequencies of this type of science teaching. Table 16.41 shows the item wording and the international IRT parameters for this scale.

Table 16.41
Item parameters for science teaching: focus on models or applications (SCAPPLY)

Item	When learning <school science> topics at school, how often do the following activities occur?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
ST34Q07	g) The teacher explains how a <school science> idea can be applied to a number of different phenomena (*e.g.* the movement of objects, substances with similar properties)	–0.64	–1.85	–0.02	1.87
ST34Q12	l) The teacher uses science to help students understand the world outside school	0.3	–1.99	0.13	1.87
ST34Q15	o) The teacher clearly explains the relevance of <broad science> concepts to our lives	–0.11	–2.05	0.18	1.87
ST34Q17	q) The teacher uses examples of technological application to show how <school science> is relevant to society	0.45	–1.95	0.15	1.8

Note: Item categories were "In all lessons", "In most lessons", "In some lessons" and "Never or hardly ever"; all items were inverted for scaling.

Table 16.42
Model fit for CFA with science teaching and learning[1]

		Model fit			
		RMSEA	RMR	CFI	NNFI
OECD	Australia	0.079	0.039	0.91	0.91
	Austria	0.067	0.039	0.94	0.94
	Belgium	0.077	0.045	0.91	0.91
	Canada	0.078	0.046	0.92	0.92
	Czech Republic	0.089	0.040	0.88	0.88
	Denmark	0.070	0.036	0.91	0.92
	Finland	0.094	0.037	0.83	0.84
	France	0.079	0.054	0.86	0.87
	Germany	0.071	0.042	0.91	0.91
	Greece	0.076	0.050	0.91	0.91
	Hungary	0.069	0.048	0.92	0.92
	Iceland	0.077	0.046	0.90	0.90
	Ireland	0.079	0.042	0.90	0.90
	Italy	0.069	0.046	0.93	0.93
	Japan	0.100	0.048	0.87	0.87
	Korea	0.067	0.031	0.93	0.93
	Luxembourg	0.079	0.047	0.91	0.91
	Mexico	0.094	0.057	0.86	0.87
	Netherlands	0.067	0.041	0.92	0.92
	New Zealand	0.072	0.039	0.92	0.92
	Norway	0.081	0.038	0.90	0.90
	Poland	0.084	0.040	0.90	0.90
	Portugal	0.075	0.040	0.91	0.91
	Slovak Republic	0.074	0.037	0.91	0.91
	Spain	0.090	0.051	0.88	0.88
	Sweden	0.100	0.049	0.85	0.85
	Switzerland	0.088	0.049	0.88	0.88
	Turkey	0.092	0.044	0.90	0.90
	United Kingdom	0.083	0.037	0.89	0.89
	United States	0.094	0.049	0.89	0.89
	OECD	**0.071**	**0.035**	**0.93**	**0.93**

1. Model estimates based on international student calibration sample (500 students per OECD country).

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



Table 16.43 shows the model fit for a four-dimensional model for the science teaching and learning items in PISA 2006. The model fit is satisfactory for the pooled OECD sample and in all but two OECD countries.

Table 16.43
Estimated latent correlations for constructs related to science teaching and learning[1]

		Latent correlations between					
		SCINTACT/ SCHANDS	SCINTACT/ SCINVEST	SCINTACT/ SCAPPLY	SCHANDS/ SCINVEST	SCHANDS/ SCAPPLY	SCINVEST/ SCAPPLY
OECD	Australia	0.57	0.55	0.76	0.67	0.58	0.68
	Austria	0.57	0.60	0.58	0.83	0.63	0.59
	Belgium	0.49	0.64	0.56	0.70	0.81	0.60
	Canada	0.59	0.62	0.78	0.67	0.57	0.54
	Czech Republic	0.65	0.64	0.77	0.74	0.77	0.70
	Denmark	0.52	0.57	0.76	0.42	0.71	0.45
	Finland	0.64	0.62	0.68	0.74	0.73	0.60
	France	0.45	0.78	0.64	0.44	0.70	0.67
	Germany	0.47	0.63	0.59	0.74	0.52	0.61
	Greece	0.67	0.60	0.80	0.94	0.73	0.68
	Hungary	0.48	0.48	0.75	1.00	0.55	0.59
	Iceland	0.47	0.35	0.69	0.89	0.49	0.41
	Ireland	0.70	0.67	0.71	0.76	0.73	0.62
	Italy	0.43	0.64	0.68	0.73	0.59	0.80
	Japan	0.63	0.92	0.77	0.89	0.61	0.90
	Korea	0.71	0.92	0.55	0.80	0.49	0.66
	Luxembourg	0.65	0.68	0.83	0.76	0.70	0.67
	Mexico	0.75	0.75	0.80	0.78	0.86	0.82
	Netherlands	0.45	0.62	0.57	0.51	0.67	0.57
	New Zealand	0.70	0.51	0.69	0.65	0.65	0.51
	Norway	0.58	0.54	0.81	0.66	0.66	0.63
	Poland	0.82	0.72	0.86	0.94	0.85	0.74
	Portugal	0.64	0.76	0.77	0.82	0.80	0.71
	Slovak Republic	0.57	0.70	0.69	0.81	0.82	0.85
	Spain	0.61	0.70	0.66	0.78	0.69	0.68
	Sweden	0.70	0.80	0.73	0.67	0.64	0.71
	Switzerland	0.49	0.69	0.71	0.67	0.70	0.61
	Turkey	0.72	0.85	0.75	0.91	0.81	0.84
	United Kingdom	0.69	0.68	0.74	0.72	0.74	0.70
	United States	0.77	0.76	0.81	0.66	0.67	0.73
	OECD	**0.55**	**0.66**	**0.74**	**0.71**	**0.66**	**0.67**

1. Model estimates based on international student calibration sample (500 students per OECD country).

Table 16.44 shows the estimated latent correlations for the four environment-related constructs. All four constructs are positively correlated with each other, the highest correlations are found between *SCINTACT* and *SCINVEST* and between *SCHANDS* and *SCINVEST*.

Table 16.45 shows the scale reliabilities for the indices related to science teaching and learning. The internal consistency of all four scales is satisfactory across countries and is typically between 0.70 and 0.80. Similar reliabilities are found in partner countries (see Table 16.44).



Table 16.44
Scale reliabilities for scales to science teaching and learning in OECD countries

	SCINTACT	SCHANDS	SCINVEST	SCAPPLY
Australia	0.80	0.71	0.75	0.81
Austria	0.82	0.78	0.79	0.73
Belgium	0.74	0.72	0.77	0.76
Canada	0.79	0.72	0.77	0.81
Czech Republic	0.79	0.72	0.76	0.76
Denmark	0.79	0.73	0.76	0.74
Finland	0.69	0.70	0.64	0.74
France	0.71	0.69	0.68	0.72
Germany	0.75	0.72	0.77	0.73
Greece	0.77	0.76	0.74	0.74
Hungary	0.75	0.74	0.74	0.74
Iceland	0.67	0.75	0.72	0.77
Ireland	0.78	0.69	0.72	0.78
Italy	0.74	0.79	0.75	0.71
Japan	0.70	0.78	0.65	0.83
Korea	0.76	0.74	0.73	0.78
Luxembourg	0.76	0.76	0.78	0.75
Mexico	0.69	0.75	0.69	0.76
Netherlands	0.73	0.75	0.75	0.75
New Zealand	0.79	0.70	0.74	0.81
Norway	0.80	0.76	0.78	0.76
Poland	0.73	0.65	0.78	0.76
Portugal	0.79	0.72	0.68	0.79
Slovakia	0.77	0.71	0.74	0.72
Spain	0.77	0.74	0.75	0.77
Sweden	0.76	0.75	0.71	0.78
Switzerland	0.72	0.76	0.75	0.75
Turkey	0.76	0.81	0.76	0.81
United Kingdom	0.77	0.69	0.75	0.77
United States	0.77	0.75	0.79	0.80
Median	**0.76**	**0.74**	**0.75**	**0.76**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

Table 16.45
Scale reliabilities for scales to science teaching and learning in partner countries/economies

	SCINTACT	SCHANDS	SCINVEST	SCAPPLY
Argentina	0.70	0.75	0.73	0.74
Azerbaijan	0.64	0.76	0.71	0.71
Brazil	0.75	0.74	0.73	0.75
Bulgaria	0.74	0.79	0.74	0.78
Chile	0.74	0.76	0.70	0.78
Colombia	0.64	0.68	0.68	0.71
Croatia	0.82	0.77	0.75	0.80
Estonia	0.72	0.70	0.67	0.72
Hong Kong-China	0.78	0.80	0.72	0.82
Indonesia	0.72	0.72	0.73	0.77
Israel	0.77	0.81	0.78	0.80
Jordan	0.70	0.72	0.67	0.71
Kyrgyzstan	0.63	0.69	0.71	0.69
Latvia	0.72	0.64	0.68	0.67
Liechtenstein	0.77	0.73	0.79	0.79
Lithuania	0.71	0.65	0.69	0.73
Macao-China	0.77	0.73	0.73	0.75
Montenegro	0.71	0.79	0.80	0.80
Qatar	0.78	0.81	0.77	0.79
Romania	0.66	0.73	0.74	0.74
Russian Federation	0.70	0.68	0.74	0.73
Serbia	0.72	0.79	0.76	0.78
Slovenia	0.71	0.73	0.71	0.78
Chinese Taipei	0.78	0.78	0.79	0.83
Thailand	0.70	0.72	0.68	0.76
Tunisia	0.64	0.58	0.64	0.62
Uruguay	0.74	0.73	0.71	0.74
Median	**0.72**	**0.73**	**0.73**	**0.75**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.



ICT familiarity

The ICT familiarity questionnaire was an optional instrument administered which was administered in 40 of the participating countries in PISA 2006, for which four scaled indices were computed.

As in PISA 2003, six items measuring the frequency of ICT use related to Internet and entertainment were included in the PISA 2006 student questionnaire. All items are reverse scored so that positive WLE scores on this index indicate high frequencies of ICT use. Table 16.46 shows the item wording and the international parameters used for scaling.

Table 16.46
Item parameters for ICT Internet/entertainment use (INTUSE)

		Parameter estimates				
Item	How often do you use computers for the following reasons?	Delta	Tau(1)	Tau(2)	Tau(3)	Tau(4)
IC04Q01	a) Browse the Internet for information about people, things, or ideas	–0.29	–0.52	–0.54	–0.13	1.18
IC04Q02	b) Play games	–0.05	–0.24	–0.10	–0.19	0.52
IC04Q04	d) Use the Internet to collaborate with a group or team	0.42	0.03	–0.44	–0.07	0.48
IC04Q06	f) Download software from the Internet to (including games)	0.31	0.09	–0.3	–0.22	0.43
IC04Q09	i) Download music from the Internet	–0.05	0.57	–0.45	–0.35	0.24
IC04Q11	k) For communication (*e.g.* e-mail or "chat rooms")	–0.34	0.65	–0.08	–0.43	–0.14

Note: Item categories were "Almost every day", "Once or twice a week", "A few times a month", "Once a month or less" and "Never"; all items were inverted for scaling.

As in PISA 2003, six items measuring the frequency of ICT use related to programming and software packages are included in the PISA 2006 student questionnaire. All items are reverse coded so that positive WLE scores on this index indicate high frequencies of ICT use. Table 16.47 shows the item wording and the international parameters used for scaling.

Table 16.47
Item parameters for ICT program/software use (PRGUSE)

		Parameter estimates				
Item	How often do you use computers for the following reasons?	Delta	Tau(1)	Tau(2)	Tau(3)	Tau(4)
IC04Q03	c) Write documents (*e.g.* with <Word® or WordPerfect®>)	–0.79	–1.04	–0.86	0.16	1.75
IC04Q05	e) Use spreadsheets (*e.g.* <Lotus 1 2 3® or Microsoft Excel®>)	0.21	–0.77	–0.53	0.02	1.27
IC04Q07	g) Drawing, painting or using graphics programs	–0.19	–0.71	–0.27	0.04	0.94
IC04Q08	h) Use educational software such as Mathematics programs	0.46	–0.47	–0.45	0.00	0.92
IC04Q10	j) Writing computer programs	0.31	0.15	–0.39	–0.16	0.40

Note: Item categories were "Almost every day", "Once or twice a week", "A few times a month", "Once a month or less" and "Never"; all items were inverted for scaling.

As in PISA 2003, items measuring students' confidence in doing ICT Internet tasks were included. However, a modified set of six items was used in the PISA 2006 student questionnaire where three items were already included in the previous cycle. All items were inverted for IRT scaling and positive WLE scores on this index indicate high self-confidence. Table 16.48 shows the item wording and the international parameters used for scaling.

Table 16.48
Item parameters for ICT self-confidence in Internet tasks (INTCONF)

		Parameter estimates			
Item	How often do you use computers for the following reasons?	Delta	Tau(1)	Tau(2)	Tau(3)
IC05Q01	a) Chat online	0.01	–1.24	0.73	0.50
IC05Q07	g) Search the Internet for information	–0.71	–0.55	0.45	0.10
IC05Q08	h) Download files or programs from the Internet	0.13	–1.39	0.21	1.18
IC05Q09	i) Attach a file to an e-mail message	0.55	–1.26	0.19	1.07
IC05Q13	m) Download music from the Internet	0.19	–1.54	0.42	1.13
IC05Q15	o) Write and send e-mails	–0.18	–1.13	0.48	0.65

Note: Item categories were " I can do this very well by myself", "I can do this with help from someone", "I know what this means but I cannot do it" and "I don't know what this means"; all items were inverted for scaling.

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



As in PISA 2003, items measuring student's confidence in doing ICT high-level tasks were included in the PISA 2006 student questionnaire. The set of eight items used in the PISA 2006 main study is modified somewhat from the 2003 item set. Items are inverted for IRT scaling and positive WLE scores on this index indicate high self-confidence. Item wording and international IRT parameters for scaling are shown in Table 16.49.

Table 16.49
Item parameters for ICT self-confidence in high-level ICT tasks (HIGHCONF)

Item	How often do you use computers for the following reasons?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
IC05Q02	b) Use software to find and get rid of computer viruses	0.09	–1.59	0.47	1.11
IC05Q03	c) Edit digital photographs or other graphic images	–0.40	–1.31	0.30	1.01
IC05Q04	d) Create a database (*e.g.* using <Microsoft Access®>)	1.10	–1.05	–0.10	1.15
IC05Q10	j) Use a word processor (*e.g.* to write an essay for school)	–0.96	–0.30	0.26	0.04
IC05Q11	k) Use a spreadsheet to plot a graph	0.08	–0.84	–0.17	1.01
IC05Q12	l) Create a presentation (*e.g.* using <Microsoft PowerPoint®>)	–0.16	–0.73	0.01	0.72
IC05Q14	n) Create a multi-media presentation (with sound, pictures, video)	–0.07	–1.55	0.10	1.46
IC05Q16	p) Construct a web page	0.33	–1.9	0.19	1.71

Note: Item categories were " I can do this very well by myself", "I can do this with help from someone", "I know what this means but I cannot do it" and "I don't know what this means"; all items were inverted for scaling.

Table 16.50 shows the model fit for a four-dimensional model for the ICT familiarity items in PISA 2006. The model fit is satisfactory for the pooled OECD sample and in all but two OECD countries.

Table 16.50
Model fit for CFA with ICT familiarity items[1]

		RMSEA	Model fit		NNFI
			RMR	CFI	
OECD	Australia	0.088	0.073	0.72	0.73
	Austria	0.081	0.079	0.79	0.79
	Belgium	0.080	0.080	0.78	0.78
	Canada	0.097	0.083	0.75	0.75
	Czech Republic	0.084	0.076	0.84	0.84
	Denmark	0.099	0.084	0.69	0.70
	Finland	0.108	0.088	0.69	0.70
	Germany	0.089	0.084	0.76	0.76
	Greece	0.084	0.097	0.84	0.84
	Hungary	0.087	0.083	0.81	0.81
	Iceland	0.089	0.078	0.71	0.72
	Ireland	0.090	0.093	0.79	0.79
	Italy	0.082	0.106	0.84	0.84
	Japan	0.086	0.071	0.83	0.83
	Korea	0.077	0.060	0.79	0.80
	Netherlands	0.079	0.061	0.72	0.72
	New Zealand	0.081	0.086	0.81	0.81
	Norway	0.100	0.082	0.76	0.76
	Poland	0.091	0.099	0.84	0.84
	Portugal	0.096	0.082	0.77	0.77
	Slovak Republic	0.084	0.090	0.83	0.83
	Spain	0.091	0.117	0.78	0.78
	Sweden	0.095	0.091	0.72	0.72
	Switzerland	0.084	0.080	0.76	0.76
	Turkey	0.084	0.079	0.84	0.84
	OECD	**0.084**	**0.082**	**0.81**	**0.81**

1. Model estimates based on international student calibration sample (500 students per OECD country).

Table 16.51 shows the estimated latent correlations for the four environment-related constructs. All four constructs are positively correlated with each other, the highest correlations are found between the two constructs reflecting self-confidence in ICT tasks.



Table 16.51
Estimated latent correlations for constructs related to ICT familiarity[1]

		Latent correlations between					
		INTUSE/PRGUSE	INTUSE/INTCONF	INTUSE/ HIGHCONF	PRGUSE/INTCONF	PRGUSE/ HIGHCONF	INTCONF/ HIGHCONF
OECD	Australia	0.53	0.58	0.52	0.16	0.49	0.71
	Austria	0.62	0.53	0.40	0.29	0.53	0.78
	Belgium	0.52	0.64	0.61	0.16	0.58	0.61
	Canada	0.61	0.61	0.56	0.12	0.46	0.67
	Czech Republic	0.54	0.62	0.56	0.37	0.62	0.79
	Denmark	0.70	0.42	0.57	0.17	0.59	0.73
	Finland	0.57	0.73	0.55	0.37	0.69	0.78
	Germany	0.59	0.64	0.53	0.32	0.57	0.78
	Greece	0.78	0.72	0.64	0.44	0.59	0.89
	Hungary	0.66	0.59	0.60	0.21	0.49	0.81
	Iceland	0.65	0.59	0.61	0.17	0.56	0.69
	Ireland	0.61	0.76	0.60	0.31	0.57	0.79
	Italy	0.55	0.73	0.51	0.37	0.66	0.80
	Japan	0.67	0.65	0.55	0.26	0.47	0.84
	Korea	0.76	0.15	0.29	0.01	0.50	0.40
	Netherlands	0.62	0.43	0.60	0.17	0.53	0.47
	New Zealand	0.47	0.54	0.42	0.18	0.44	0.71
	Norway	0.67	0.51	0.66	0.14	0.52	0.68
	Poland	0.53	0.64	0.53	0.27	0.50	0.84
	Portugal	0.57	0.67	0.48	0.24	0.39	0.80
	Slovak Republic	0.60	0.70	0.62	0.39	0.59	0.82
	Spain	0.55	0.76	0.55	0.32	0.66	0.68
	Sweden	0.56	0.58	0.58	0.16	0.57	0.64
	Switzerland	0.59	0.57	0.64	0.10	0.45	0.72
	Turkey	0.77	0.68	0.64	0.37	0.60	0.87
	OECD	**0.61**	**0.65**	**0.60**	**0.21**	**0.54**	**0.76**

1. Model estimates based on international student calibration sample (500 students per OECD country).

Table 16.52
Scale reliabilities for ICT familiarity scales

		INTUSE	PRGUSE	INTCONF	HIGHCONF
OECD	Australia	0.75	0.71	0.83	0.80
	Austria	0.71	0.68	0.80	0.80
	Belgium	0.71	0.72	0.82	0.78
	Canada	0.71	0.74	0.83	0.79
	Czech Republic	0.83	0.78	0.85	0.85
	Denmark	0.66	0.73	0.76	0.80
	Finland	0.72	0.73	0.76	0.83
	Germany	0.71	0.72	0.80	0.82
	Greece	0.82	0.79	0.82	0.84
	Hungary	0.79	0.71	0.86	0.83
	Iceland	0.69	0.75	0.74	0.77
	Ireland	0.78	0.75	0.84	0.84
	Italy	0.82	0.73	0.86	0.83
	Japan	0.80	0.75	0.88	0.89
	Korea	0.66	0.71	0.81	0.82
	Netherlands	0.63	0.73	0.80	0.76
	New Zealand	0.76	0.75	0.83	0.82
	Norway	0.73	0.75	0.84	0.79
	Poland	0.86	0.79	0.90	0.86
	Portugal	0.79	0.75	0.84	0.84
	Slovak Republic	0.78	0.79	0.87	0.86
	Spain	0.77	0.72	0.84	0.82
	Sweden	0.71	0.75	0.77	0.83
	Switzerland	0.72	0.74	0.83	0.82
	Turkey	0.85	0.83	0.89	0.89
	Median	**0.75**	**0.74**	**0.83**	**0.82**
Partners	Bulgaria	0.81	0.79	0.90	0.85
	Chile	0.81	0.80	0.87	0.83
	Colombia	0.83	0.77	0.88	0.85
	Croatia	0.80	0.79	0.86	0.85
	Jordan	0.84	0.80	0.87	0.86
	Latvia	0.78	0.74	0.82	0.77
	Liechtenstein	0.67	0.71	0.81	0.83
	Lithuania	0.82	0.78	0.86	0.83
	Macao-China	0.73	0.73	0.83	0.80
	Qatar	0.84	0.83	0.87	0.87
	Russian Federation	0.89	0.81	0.93	0.88
	Serbia	0.85	0.75	0.92	0.87
	Slovenia	0.78	0.78	0.86	0.82
	Thailand	0.85	0.78	0.89	0.88
	Uruguay	0.82	0.84	0.89	0.85
	Median	**0.82**	**0.78**	**0.87**	**0.85**



Table 16.52 shows the scale reliabilities for the countries that administered the ICT familiarity questionnaire. The internal consistencies are high most countries; in only very few countries there are reliabilities 0.7 for *INTUSE* and *PRGUSE.*

School questionnaire scale indices

The Index on Teacher Shortage (*TCSHORT*) was derived from four items measuring the school principal's perceptions of potential factors hindering instruction at school. Similar items were used in PISA 2000 and 2003. The items were not inverted for scaling such that higher WLE scores indicate higher rates of teacher shortage at a school. Table 16.53 shows the item wording and the international parameters used for IRT scaling.

Table 16.53
Item parameters for teacher shortage (TCSHORT)

Item	Is your school's capacity to provide instruction hindered by any of the following?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
SC14Q01	a) A lack of qualified science teachers	0.10	–1.24	–0.53	1.76
SC14Q02	b) A lack of qualified mathematics teachers	–0.05	–0.92	–0.21	1.12
SC14Q03	c) A lack of qualified <test language> teachers	0.25	–0.82	–0.18	1.00
SC14Q04	d) A lack of qualified teachers of other subjects	–0.30	–1.79	–0.31	2.10

Note: Categories were "not at all", "very little", "to some extent" and "a lot".

The index on the school's educational resources (*SCMATEDU*) was computed on the basis of seven items measuring the school principal's 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. All items were inverted for IRT scaling and positive WLE scores indicate better quality of educational resources. Table 16.54 shows the item wording and the international parameters used for IRT scaling.

Table 16.54
Item parameters for quality of educational resources (SCMATEDU)

Item	Is your school's capacity to provide instruction hindered by any of the following?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
SC14Q07	g) Shortage or inadequacy of science laboratory equipment	0.40	–1.47	0.25	1.22
SC14Q08	h) Shortage or inadequacy of instructional materials (*e.g.* textbooks)	–0.43	–1.85	0.28	1.57
SC14Q09	i) Shortage or inadequacy of computers for instruction	0.05	–1.49	0.18	1.31
SC14Q10	j) Lack or inadequacy of internet connectivity	–0.50	–0.81	0.04	0.78
SC14Q11	k) Shortage or inadequacy of computer software for instruction	0.12	–1.64	0.13	1.50
SC14Q12	l) Shortage or inadequacy of library materials	0.06	–1.92	0.04	1.88
SC14Q13	m) Shortage or inadequacy of audio-visual resources	0.31	–1.64	–0.02	1.66

Note: Categories were "not at all", "very little", "to some extent" and "a lot"; all items were inverted for scaling.

School principals are asked to report what activities to promote students' learning of science occur at their school. Items were coded (Yes=1, No=0) so that positive WLE scores indicate higher levels of school activities in this area. Table 16.55 shows the item wording and the international parameters used for IRT scaling.

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



Table 16.55
Item parameters for school activities to promote the learning of science (SCIPROM)

Item	Is your school involved in any of the following activities to promote engagement with science among students in <national modal grade for 15-year-olds>? (Yes/No)	Parameter estimates
		Delta
SC20Q01	a) Science clubs	0.90
SC20Q02	b) Science fairs	0.76
SC20Q03	c) Science competitions	0.23
SC20Q04	d) Extracurricular science projects (including research)	0.24
SC20Q05	e) Excursions and field trips	–2.13

Note: Categories were "Yes" and "No"; all items were inverted for scaling.

School principals are asked to report what activities to promote students' learning of environmental topics occur at their school. Items will be coded (Yes=1, No=0) so that positive WLE scores indicate higher levels of school activities in this area. Table 1656 shows the item wording and the international parameters used for IRT scaling.

Table 16.56
Item parameters for school activities for learning environmental topics (ENVLEARN)

Item	Does your school organise any of the following activities to provide opportunities to students in <national modal grade for 15-year-olds> to learn about environmental topics?	Parameter estimates
		Delta
SC22Q01	a) <Outdoor education>	–0.37
SC22Q02	b) Trips to museums	–0.77
SC22Q03	c) Trips to science and/or technology centres	–0.09
SC22Q04	d) Extracurricular environmental projects (including research)	0.76
SC22Q05	e) Lectures and/or seminars (*e.g.* guest speakers)	0.46

Note: Categories were "Yes" and "No"; all items were inverted for scaling.

Table 16.57
Scale reliabilities for school-level scales in OECD countries

	TCSHORT	SCMATEDU	SCIPROM	ENVLEARN
Australia	0.87	0.90	0.35	0.60
Austria	0.71	0.87	0.65	0.58
Belgium	0.87	0.84	0.43	0.48
Canada	0.85	0.87	0.59	0.63
Czech Republic	0.72	0.79	0.63	0.46
Denmark	0.71	0.84	0.45	0.57
Finland	0.64	0.86	0.26	0.51
Germany	0.78	0.86	0.63	0.49
Greece	0.92	0.81	0.49	0.34
Hungary	0.67	0.81	0.49	0.53
Iceland	0.82	0.76	0.49	0.50
Ireland	0.75	0.84	0.62	0.74
Italy	0.86	0.86	0.53	0.60
Japan	0.79	0.86	0.62	0.69
Korea	0.87	0.85	0.59	0.60
Luxembourg	0.87	0.86	0.66	0.64
Mexico	0.89	0.90	0.62	0.67
Netherlands	0.75	0.82	0.64	0.62
New Zealand	0.71	0.88	0.59	0.73
Norway	0.75	0.78	0.42	0.49
Poland	0.55	0.85	0.37	0.40
Portugal	0.52	0.83	0.41	0.44
Slovakia	0.67	0.79	0.59	0.32
Spain	0.84	0.85	0.50	0.51
Sweden	0.77	0.82	0.49	0.42
Switzerland	0.76	0.83	0.38	0.42
Turkey	0.93	0.85	0.67	0.62
United Kingdom	0.81	0.89	0.58	0.71
United States	0.84	0.86	0.61	0.67
Median	**0.78**	**0.85**	**0.58**	**0.57**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.

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



Table 16.57 shows the scale reliabilities for school-level indices in OECD countries. Both *TCSHORT* and *SCMATEDU* have high reliabilities across countries. The internal consistencies for the scales on school activities to learn science and environmental issues are rather low (in some countries even below 0.5).

Table 16.58 shows the scale reliabilities for partner countries. Again, high reliabilities can be observed for the two indices related to school resources but the internal consistencies for the two indices on school activities are low and even very low in some of the countries.

Table 16.58
Scale reliabilities for environment-related scales in partner countries/economies

	TCSHORT	SCMATEDU	SCIPROM	ENVLEARN
Argentina	0.85	0.87	0.66	0.62
Azerbaijan	0.85	0.75	0.57	0.49
Brazil	0.86	0.90	0.41	0.57
Bulgaria1	0.42	0.69	0.45	0.41
Chile	0.83	0.89	0.71	0.66
Colombia	0.87	0.89	0.49	0.49
Croatia	0.71	0.82	0.72	0.59
Estonia	0.71	0.76	0.08	0.49
Hong Kong-China	0.88	0.89	0.30	0.59
Indonesia	0.84	0.89	0.71	0.57
Israel	0.78	0.90	0.69	0.71
Jordan	0.93	0.84	0.60	0.52
Kyrgyzstan	0.85	0.85	0.48	0.57
Latvia	0.72	0.81	0.33	0.55
Liechtenstein	0.88	0.86	0.53	0.28
Lithuania	0.84	0.81	0.43	0.44
Macao-China	0.93	0.84	0.40	0.71
Montenegro	0.72	0.86	0.65	0.65
Qatar	0.95	0.86	0.64	0.39
Romania	0.64	0.83	0.58	0.44
Russian Federation	0.88	0.81	0.41	0.45
Serbia	0.67	0.77	0.66	0.50
Slovenia	0.62	0.80	0.62	0.57
Chinese Taipei	0.93	0.93	0.72	0.68
Thailand	0.74	0.92	0.62	0.61
Tunisia	0.70	0.74	0.59	0.52
Uruguay	0.82	0.90	0.65	0.62
Median	**0.84**	**0.85**	**0.59**	**0.57**

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.
1. Reliability for SCIPROM in Bulgaria was calculated without the item SC20Q01 ("science clubs").

Parent questionnaire scale indices

Parent questionnaire indices are only available for the 16 countries which chose to administer the optional parent questionnaire.

Six items measuring students' activities related to science at age 10 were included in the parent questionnaire. The items were inverted for scaling so that positive WLE scores on this index indicate higher frequencies of students' science activities. The item wording and international parameters for IRT scaling are shown in Table 16.59.

Seven items measuring parents' perceptions of the quality of school learning were included in the parent questionnaire. The items were reverse scored prior to scaling so that positive WLE scores on this index indicate positive evaluations of the school's quality. Table 16.60 shows then item wording and the international parameters used for IRT scaling.



Table 16.59
Item parameters for science activities at age 10 (PQSCIACT)

Item	Thinking back to when your child was about 10 years old, how often would your child have done these things?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
PA02Q01	a) Watched TV programmes about science	–0.89	–2.40	1.08	1.31
PA02Q02	b) Read books on scientific discoveries	–0.05	–1.85	0.89	0.97
PA02Q03	c) Watched, read or listened to science fiction	–0.79	–1.81	0.69	1.12
PA02Q04	d) Visited web sites about science topics	0.63	–0.97	0.60	0.38
PA02Q05	e) Attended a science club	1.09	–0.36	0.11	0.24

Note: Categories were "very often", "regularly", "sometimes" and "never"; all items were inverted for scaling.

Table 16.60
Item parameters for parent's perception of school quality (PQSCHOOL)

Item	How much do you agree with the following statements?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
PA03Q01	a) Most of my child's school teachers seem competent and dedicated	–0.4	–2.8	–1.11	3.91
PA03Q02	b) Standards of achievement are high in my child's school	0.11	–3.35	–0.3	3.64
PA03Q03	c) I am happy with the content taught and the instructional methods used in my child's school	0.01	–3.02	–0.75	3.77
PA03Q04	d) I am satisfied with the disciplinary atmosphere in my child's school	0.13	–2.38	–0.71	3.09
PA03Q05	e) My child's progress is carefully monitored by the school	0.19	–2.87	–0.56	3.43
PA03Q06	f) My child's school provides regular and useful information on my child's progress	0.35	–2.49	–0.46	2.95
PA03Q07	g) My child's school does a good job in educating students	–0.37	–2.69	–0.83	3.52

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Four items measuring parents' views on the importance of science were included in the PISA 2006 parent questionnaire. The items were inverted for scaling so that positive WLE scores on this index will indicate positive evaluations of the school's quality. Table 16.61 shows then item wording and the international parameters used for IRT scaling.

Table 16.61
Item parameters for parent's views on importance of science (PQSCIMP)

Item	We are interested in what you think about the need for science skills in the job market today. How much do you agree with the following statements?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
PA04Q01	a) It is important to have good scientific knowledge and skills in order to get any good job in today's world	–0.40	–3.73	0.02	3.71
PA04Q02	b) Employers generally appreciate strong scientific knowledge and skills among their employees	0.48	–4.2	0.12	4.08
PA04Q03	c) Most jobs today require some scientific knowledge and skills	0.33	–4.27	0.04	4.22
PA04Q04	d) It is an advantage in the job market to have good scientific knowledge and skills	–0.41	–3.55	–0.32	3.87

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Four items measuring parents' reports on science career motivation for their child were included in the PISA 2006 parent questionnaire. The items were e inverted for scaling so that positive WLE scores on this index indicate higher levels of science career motivation. One item in this set (*PA05Q01* "Does anybody in your family (including you) work in a <science-related career>?') was not included in the scale since it is unrelated to the construct of career motivation of parents for their child. Item wording and international IRT parameters are shown in Table 16.62.



Table 16.62
Item parameters for parent's reports on science career motivation (PQSCCAR)

Item	Please answer the questions below (Yes/No)	Parameter estimates
PA05Q02	b) Does your child show an interest to work in a <science-related career>?	–0.42
PA05Q03	c) Do you expect your child will go into a <science-related career>?	–0.44
PA05Q04	d) Has your child shown interest in studying science after completing <secondary school>?	0.03
PA05Q05	e) Do you expect your child will study science after completing <secondary school>?	–0.29

Note: Categories were "Yes" and "No"; all items were inverted for scaling.
Five items measuring parents' perceptions of the general value of science were included in the PISA 2006 parent questionnaire; similar items were also included in the student questionnaire. As with the student scale, the items are reverse scored for scaling so that positive WLE scores on this new PISA 2006 index indicate positive parents' perceptions of the general value of science. Table 55 shows the item wording and international parameters used for scaling.

Five items measuring parents' perceptions of the general value of science were included in the PISA 2006 parent questionnaire; similar items were also included in the student questionnaire. As with the student scale, the items are reverse scored for scaling so that positive WLE scores on this new PISA 2006 index indicate positive parents' perceptions of the general value of science. Table 16.63 shows the item wording and international parameters used for scaling.

Table 16.63
Item parameters for parents' view on general value of science (PQGENSCI)

Item	The following question asks about your views towards science. How much do you agree with the following statements?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
PA06Q01	a) Advances in <broad science and technology> usually improve people's living conditions	–0.29	–2.45	–1.10	3.56
PA06Q02	b) <Broad science> is important for helping us to understand the natural world	–0.49	–2.34	–1.32	3.66
PA06Q04	d) Advances in <broad science and technology> usually help improve the economy	0.30	–2.86	–0.67	3.54
PA06Q06	f) <Broad science> is valuable to society	–0.09	–2.42	–1.14	3.56
PA06Q09	i) Advances in <broad science and technology> usually bring social benefits	0.56	–2.82	–0.68	3.50

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Four items measuring parents' perceptions of the personal value of science are included in the PISA 2006 parent questionnaire; similar items are included in the student questionnaire. The items were inverted for scaling so that positive WLE scores indicate positive students' perceptions of the general value of science. Table 16.64 shows the item wording and international parameters used for scaling.

Table 16.64
Item parameters for parent's view on personal value of science (PQPERSCI)

Item	The following question asks about your views towards science. How much do you agree with the following statements?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
PA06Q03	c) Some concepts in <broad science> help me to see how I relate to other people	0.10	–3.51	–0.2	3.71
PA06Q05	e) There are many opportunities for me to use <broad science> in my everyday life	0.51	–3.18	–0.05	3.23
PA06Q07	g) <Broad science> is very relevant to me	–0.03	–2.74	–0.22	2.96
PA06Q08	h) I find that <broad science> helps me to understand the things around me	–0.57	–2.87	–0.63	3.49

Note: Item categories were "strongly agree", "agree", "disagree" and "strongly disagree"; all items were inverted for scaling.

Six items measuring perception of environmental issues as a concern were included in the PISA 2006 parent questionnaire; similar items were also included in the student questionnaire. The items were reverse scored for scaling so that positive WLE scores on this index indicate higher levels of parents' concerns about environmental issues. Table 16.65 shows the item wording and international parameters used for scaling.



Table 16.65
Item parameters for parent's perception of environmental issues (PQENPERC)

Item	Do you see the environmental issues below as a serious concern for yourself and/or others?	Parameter estimates			
		Delta	Tau(1)	Tau(2)	Tau(3)
PA07Q01	a) Air pollution	–0.79	–0.74	0.20	0.54
PA07Q02	b) Energy shortages	–0.06	–1.24	0.44	0.80
PA07Q03	c) Extinction of plants and animals	0.30	–1.3	0.35	0.95
PA07Q04	d) Clearing of forests for other land use	0.27	–1.73	0.8	0.92
PA07Q05	e) Water shortages	–0.1	–1.87	1.37	0.51
PA07Q06	f) Nuclear waste	0.37	–1.79	1.04	0.75

Note: Item categories were "This is a serious concern for me personally as well as others" , "This is a serious concern for other people in my country but not me personally", "This is a serious concern for people in other countries" and "This is not a serious concern to anyone"; all items were inverted for scaling.

Six items measuring parents' optimism regarding environmental issues were included in the PISA 2006 parent questionnaire similar to items on the student questionnaire. These were inverted for scaling so that positive WLE scores on the index indicate higher levels of parents' optimism about environmental issues. Table 16.66 shows the item wording and international parameters used for scaling.

Table 16.66
Item parameters for parent's environmental optimism (PQENVOPT)

Item	Do you think problems associated with the environmental issues below will improve or get worse over the next 20 years?	Parameter estimates		
		Delta	Tau(1)	Tau(2)
PA08Q01	a) Air pollution	–0.04	–0.14	0.14
PA08Q02	b) Energy shortages	–0.33	–0.64	0.64
PA08Q03	c) Extinction of plants and animals	0.14	–0.64	0.64
PA08Q04	d) Clearing of forests for other land use	0.17	–0.44	0.44
PA08Q05	e) Water shortages	0.04	–0.64	0.64
PA08Q06	f) Nuclear waste	0.01	–0.64	0.64

Note: Item categories were "Improve", "Stay about the same" and "Get worse"; all items were inverted for scaling.

Table 16.67 shows the reliabilities for the scale indices derived from the parent questionnaire. Most indices have high reliabilities across countries, only the index PQSCIEACT has somewhat lower internal consistency but it is still satisfactory in most country sub-samples.

Table 16.67
Scale reliabilities for parent questionnaire scales

		PQSCIEACT	PQSCHOOL	PQSCIMP	PQSCCAR	PQGENSCI	PQPERSCI	PQENPERC	PQENVOPT
OECD	Denmark	0.63	0.90	0.88	0.93	0.81	0.83	0.82	0.75
	Germany	0.50	0.84	0.86	0.80	0.77	0.78	0.80	0.76
	Iceland	0.72	0.87	0.84	0.96	0.83	0.82	0.85	0.78
	Italy	0.65	0.82	0.83	0.93	0.77	0.72	0.75	0.82
	Korea	0.78	0.84	0.76	0.82	0.83	0.77	0.84	0.87
	Luxembourg	0.60	0.84	0.86	0.86	0.80	0.79	0.81	0.85
	New Zealand	0.67	0.88	0.86	0.94	0.83	0.83	0.82	0.84
	Poland[1]		0.84						
	Portugal	0.73	0.83	0.83	0.95	0.80	0.76	0.81	0.86
	Turkey	0.67	0.80	0.72	0.85	0.70	0.69	0.77	0.83
Partners	Bulgaria[2]	0.78	0.84	0.73	0.88	0.78	0.72	0.81	0.88
	Colombia	0.67	0.84	0.71	0.81	0.77	0.73	0.71	0.91
	Croatia	0.67	0.78	0.82	0.88	0.81	0.80	0.76	0.85
	Hong Kong-China	0.76	0.80	0.85	0.79	0.81	0.76	0.82	0.80
	Macao-China	0.74	0.82	0.82	0.79	0.80	0.77	0.85	0.86
	Qatar	0.72	0.87	0.75	0.87	0.81	0.78	0.81	0.86

Note: Reliabilities (Cronbach's alpha) computed with weighted national samples.
1. Poland did not submit results for any items in PQSCIEACT, PQSCIMP, PQSCCAR, PQGENSCI, PQENPERC, PQENVOPT.
2. Reliability for the index of PQSCIEACT in Bulgaria was calculated with the omission of PA02Q05.

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



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 WLE estimates based on student reports on home possessions).

The *ESCS* for PISA 2003 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.[6] The rationale for using these three components is that socio-economic status is usually seen as based on education, occupational status and income. As no direct income measure is available from the PISA data, the existence of household items is used as proxy for family wealth.

The *ESCS* has been slightly modified because: (i) there were more indicators available in the recent 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, the components comprising *ESCS* for 2006 are home possessions, *HOMEPOS* (which comprises all items on the *WEALTH*, *CULTPOS* and *HEDRES* scales (except ST14Q04), as well as books in the home (ST15Q01) recoded into a three-level categorical variable (less than 25 books, 25-100 books, more than 100 books), the higher parental occupation (*HISEI*) and the higher parental education expressed as years of schooling (*PARED*).

Missing values for students with missing data for only one component were imputed with predicted values plus a random component based on a regression on the other two variables. 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 the standard deviation across equally weighted OECD countries. For partner countries, *ESCS* scores were obtained as

16.6

$$ESCS = \frac{\beta_1 HISEI' + \beta_2 PARED' + \beta_3 HOMEPOS'}{\varepsilon_f}$$

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.[7]

Consistency across cycles

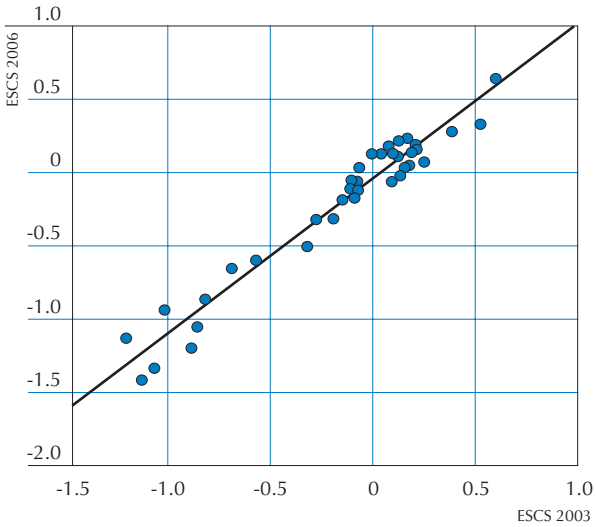
Results for similar *ESCS* indices in 2003 and 2000 showed quite a high degree of consistency (see Schulz, 2006b). Comparing *ESCS* mean scores per country shows that in spite of these differences there is a very high correlation of 0.98 between ESCS 2003 and ESCS 2006 country means (*see* Figure 16.3).

Consistency across countries

Using principal component analysis (PCA) to derive factor loading for each participating country provides insight into the extent to which there are similar relationships between the three components. Table 16.68 shows the PCA results for the OECD countries and Table 16.69 those for partner countries. The tables also include the scale reliabilities for the *z*-standardised variables (Cronbach's Alpha).



Figure 16.3
Scatterplot of country means for ESCS 2003 and ESCS 2006



Note: Weighted averages for OECD and partner countries and economies participating in both cycles.

Table 16.68
Factor loadings and internal consistency of ESCS 2006 in OECD countries

	Factor loadings			
	HISEI	PARED	HOMEPOS	Reliability[1]
Australia	0.80	0.78	0.67	0.59
Austria	0.81	0.78	0.71	0.64
Belgium	0.83	0.80	0.71	0.68
Canada	0.79	0.78	0.67	0.60
Czech Republic	0.84	0.78	0.70	0.65
Denmark	0.79	0.78	0.70	0.63
Finland	0.77	0.75	0.63	0.52
France	0.82	0.79	0.73	0.67
Germany	0.81	0.76	0.72	0.64
Greece	0.84	0.82	0.72	0.71
Hungary	0.83	0.85	0.77	0.74
Iceland	0.80	0.80	0.59	0.57
Ireland	0.81	0.79	0.74	0.67
Italy	0.84	0.81	0.73	0.71
Japan	0.72	0.77	0.68	0.53
Korea	0.76	0.81	0.75	0.66
Luxembourg	0.83	0.81	0.73	0.69
Mexico	0.85	0.86	0.82	0.80
Netherlands	0.82	0.78	0.75	0.68
New Zealand	0.79	0.76	0.69	0.59
Norway	0.78	0.77	0.66	0.55
Poland	0.87	0.86	0.74	0.73
Portugal	0.86	0.85	0.80	0.77
Slovakia	0.85	0.82	0.74	0.72
Spain	0.84	0.82	0.70	0.69
Sweden	0.77	0.73	0.70	0.57
Switzerland	0.80	0.78	0.68	0.62
Turkey	0.80	0.83	0.79	0.72
United Kingdom	0.78	0.75	0.71	0.60
United States	0.80	0.81	0.74	0.67
Median	**0.81**	**0.79**	**0.72**	**0.67**

1. Reliabilities (Standardised Cronbach's alpha) computed with weighted national samples.



Comparing results from within-country PCA reveals that patterns of factor loadings are generally similar across countries. Only in a few countries somehow distinct patterns emerge, however, all three components contribute more or less equally to this index with factor loadings ranging from 0.55 to 0.87. Internal consistency ranges between 0.52 and 0.80, the median scale reliability for the pooled OECD countries is 0.67.

Table 16.69
Factor loadings and internal consistency of ESCS 2006 in partner countries/economies

	Factor loadings			
	HISEI	PARED	HOMEPOS	Reliability[1]
Argentina	0.81	0.78	0.79	0.69
Azerbaijan	0.83	0.83	0.73	0.70
Brazil	0.82	0.83	0.80	0.73
Bulgaria	0.84	0.83	0.77	0.74
Chile	0.86	0.85	0.83	0.80
Colombia	0.82	0.82	0.79	0.73
Croatia	0.83	0.81	0.73	0.69
Estonia	0.81	0.77	0.72	0.63
Hong Kong-China	0.83	0.82	0.77	0.72
Indonesia	0.81	0.83	0.78	0.73
Israel	0.78	0.75	0.73	0.60
Jordan	0.83	0.83	0.75	0.73
Kyrgyzstan	0.76	0.76	0.71	0.57
Latvia	0.81	0.78	0.74	0.66
Liechtenstein	0.83	0.81	0.62	0.63
Lithuania	0.81	0.79	0.76	0.68
Macao-China	0.79	0.77	0.75	0.65
Montenegro	0.80	0.80	0.73	0.66
Qatar	0.82	0.86	0.55	0.60
Romania	0.82	0.75	0.80	0.69
Russian Federation	0.81	0.79	0.69	0.59
Serbia	0.84	0.84	0.72	0.71
Slovenia	0.84	0.84	0.71	0.71
Chinese Taipei	0.77	0.79	0.70	0.61
Thailand	0.85	0.84	0.82	0.78
Tunisia	0.86	0.85	0.83	0.79
Uruguay	0.83	0.81	0.81	0.74
Median	**0.82**	**0.81**	**0.75**	**0.69**

1. Reliabilities (Cronbach's alpha) computed with weighted national samples.



Notes

1. Data on public/private school ownership in Australia are not included in the PISA 2003 database. In Austria, the question on funding was omitted and only for private schools information on government funding was provided to construct this index.

2. The raw index was transformed as (*RESPRES*_raw-2.57)/2.2.

3. The raw index was transformed as (*RESPCURR*_raw-2.72)/1.8.

4. A similar approach was used in the IEA Civic Education Study (*see* Schulz, 2004).

5. This analysis did not include the country-specific items.

6. Here, home possessions only included items from ST17, as well as books in the home (*ST19Q01*) which was recoded into a dichotomous item (0 = "Less than 100 books', 1 = "100 books or more") (see OECD, 2004, p. 283).

7. Only one principal component with an eigenvalue greater than 1 was identified in each of the participating countries.



17

Validation of the Embedded Attitudinal Scales

Introduction 352
International scalability 353
- Analysis of item dimensionality with exploratory and confirmatory factor analysis 353
- Fit to item response model 353
- Reliability 355
- Differential item functioning 355
- Summary of scalability 357

Relationship and comparisons with other variables 357
- Within-country student level correlations with achievement and selected background variables 358
- Relationships between embedded scales and questionnaire 360
- Country level correlations with achievement and selected background variables 361
- Variance decomposition 363
- Observations from other cross-national data collections 363
- Summary of relations with other variables 364

Conclusion 364

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



INTRODUCTION

The development processes that are employed by PISA to ensure the cross-national validity of its scales consist of four steps. First, the construct should have well-established theoretical underpinnings. That is the construct should be underpinned by a body of academic literature and it should be supported by leading theorists and academics working in an area. Within PISA this is ensured through the articulation of the constructs in widely discussed and reviewed assessment frameworks. For the *embedded interest* and *embedded support* scales the articulation can be found in the *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006) (also see Chapter 2).

Second, the approach to operationalising the construct must be widely supported – that is, there must be wide agreement that the items that are used in PISA are reflective of the underlying conceptual definition of the domain. For the *embedded interest* and *embedded support* scales the procedures used in PISA to ensure this parallel those used for the cognitive assessment items. The procedures that PISA puts in place to achieve this include:

- The use of skilled professional test development teams from a variety of PISA participating countries;
- Review of the items as they are prepared by experts who have been directly involved in and were often responsible for the conceptualisation of the underpinning construct definitions;
- Opportunities for review and evaluation of the drafted items by PISA participating countries on multiple occasions;
- A detailed set of translation and translation verification protocols that are aimed at ensuring the conceptual and psychometric equivalence of the items across languages and cultures;
- A range of small, medium and large trial testing activities where students are asked to respond to the item and to reflect upon the meaning of the items to them.

Third, psychometric analyses are undertaken to ensure that the sets of items that are deemed to be reflective of the underlying construct can indeed be brought together in a coherent fashion to provide indicators of the underlying construct. These analyses pay particular attention to the scalability, reliability and cross-country consistency of the behaviour of the items.

Finally, the constructed scales are reviewed for their nomothetic span. That is, the extent to which relations with other variables make conceptual sense.

This chapter is concerned with the range of analyses that were undertaken as part of the last two steps in the above-described process for validating the scales that were constructed from the attitudinal items. The purpose of these analyses was to confirm the empirical validity of the scales for the purposes of cross-national comparisons.

For the main study, attitudinal items were embedded within units of the science test in order to obtain measures of two attitudinal dimensions (or constructs): *interest in science* and *support for scientific inquiry*.[1] In short, these domains will be referred to as *embedded interest* and *embedded support.*

As the analyses reported here were undertaken for validation purposes, prior to the finalisation of the international database, they were undertaken with data that had not been fully cleaned and weighted. The majority of the analyses reported use data from 51 different data sets – this was made up of all 30 OECD countries and 21 partner countries. Where this is not the case it is noted.



The thirty OECD datasets were used for the analyses of scale reliability, gender DIF, general confirmation of the expected dimensional structure for embedded science attitude items and the correlation between scales.[2] Random sub-samples of 5000 cases were taken for countries that used over-sampling in the main study. In particular, reduced samples were used for Australia, Belgium, Canada, Italy, Mexico, Spain and Switzerland. The UH booklet responses were excluded from these analyses.[3]

For the item response theory analyses, a calibration sample of 500 cases from each of the OECD datasets was used to estimate item parameters. These item parameter estimates were then used to estimate weighted likelihood estimates for each case for each of the five test scales (*mathematics, reading, science, embedded interest* and *embedded support*).

Preliminary weights were available for all countries except Australia and USA at the time of analysing the data for this report.

INTERNATIONAL SCALABILITY

Analysis of item dimensionality with exploratory and confirmatory factor analysis

Software packages *Mplus* (Muthén and Muthén, 2004) and ACER *ConQuest*® (Wu, Adams and Wilson, 1997) were used to confirm the two dimensional structure of the embedded attitudinal measures (*embedded interest* and *embedded support*).

When the items have Likert-type response categories it is recommended that factor analyses should be conducted on the matrix of polychoric inter-item correlations rather than on the matrix of product-moment correlations. Unfortunately, exploratory factor analyses (EFA) based on polychoric correlations has only been implemented in the software package *Mplus* for complete datasets. Because PISA uses a rotated booklet design, EFA was undertaken with *Mplus* with the variables defined as continuous and with product-moment correlations.[4]

Appendix 7 gives the *Mplus* results for both an EFA (with a *promax* rotation) and a two dimensional confirmatory factor analysis (CFA). The results can be summarised as follows:

- *Interest in learning science (embedded interest)*: Solutions with two factors confirmed that interest items generally loaded on one dimension. However, some items (S456N-THE CHEETAH, S519N-AIRBAGS and S527N-EXTINCTION OF THE DINOSAURS) were loading on the second factor – that is the *support* factor;
- *Support for scientific inquiry (embedded support)*: the items selected for main study for this domain items loaded on one factor;
- For the CFA the estimated latent correlation between *embedded interest* and *embedded support* was 0.594;
- The RMSEA measure of model fit produced by *Mplus* was 0.025, which was considered quite acceptable.

Fit to item response model

An alternative approach to assessing item dimensionality is to assess the fit of the data to a multi-dimensional IRT model. Here, a five-dimensional model (reading, mathematics, science, embedded interest and embedded support) was fit to the data using the *ConQuest*® software (Wu, Adams and Wilson, 1997).



The item-level fit statistics for each of the attitudinal items is given in Appendix 7, a normal probability plot of the fit mean squares is given in Figure 17.1 and the estimated latent correlations for the five-dimensional IRT model are given in Table 17.1. The normal probability plot provides a comparison of the distribution of the fit statistics with normal distribution that would be expected if the data did fit the model.

The range and distributions of the fit statistics show an acceptable fit to the multi-dimensional item response model. The fit mean squares are close to normally distributed and the worst fit mean square is 1.18.

Figure 17.1
Distribution of item fit mean square statistics for embedded attitude items

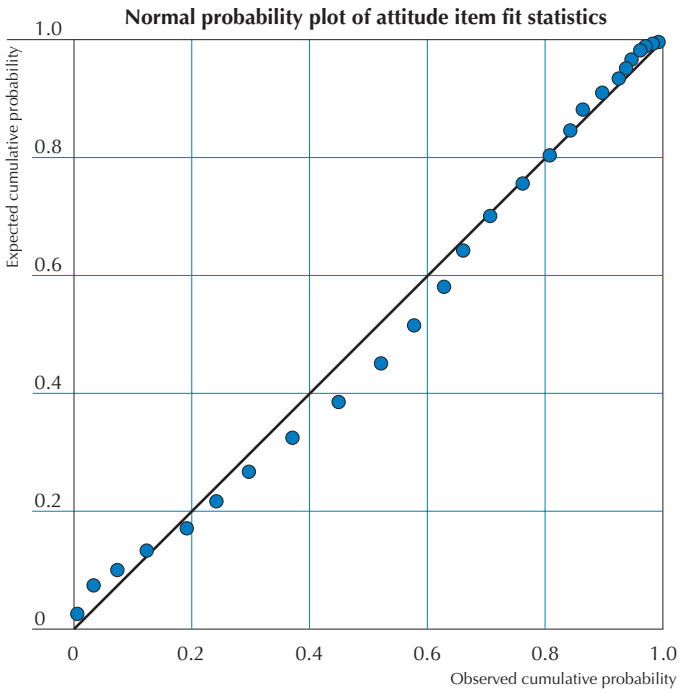


Table 17.1 shows that the estimated latent correlation between *embedded interest* and *embedded support* is 0.623. This is very similar to the corresponding value, 0.594, that was estimated using the CFA.

The correlation of the *embedded support* scale with the three achievement scales is about 0.20, while the correlation between the *embedded interest* scale with the three achievement scales is about –0.15.[5] We return to explore this negative correlation, at the student level, later in this chapter.

Table 17.1
Student-level latent correlations between mathematics, reading, science, embedded interest and embedded support

	Mathematics	Reading	Science	Embedded Interest
Reading	0.780			
Science	0.871	0.831		
Embedded Interest	–0.194	–0.151	–0.133	
Embedded Support	0.136	0.215	0.223	0.623



Reliability

Further scaling properties of the embedded items are reported in Chapter 13, where the overall reliability of the *embedded interest* scale is estimated as 0.892, and of the *embedded support* scale 0.818 (using WLEs). The reliabilities by country are also reported in Chapter 12.

At the country level the reliabilities are greater than 0.80 for *embedded interest* and greater than 0.70 for the *embedded support* scale. As discussed in Chapter 12 the lower reliability for the *embedded support* scale is likely due to the fact that the majority of student responded positively to the support items – *i.e.* students overwhelming expressed positive support for science.

Differential item functioning

Country DIF

IRT models were also estimated for each country data set separately. Comparing the outcomes with the results for the pooled international sample (51 countries) provides information about potential item-by-country interactions, which is a case of differential item function (DIF) associated with the country of test. In addition, it is informative to review item discrimination and item fit statistics in order to assess whether the scaling model holds across countries.

Table 17.2
Summary of the IRT scaling results across countries

	Item-by-country Interaction			Number of items with a discrimination < 0.20	Weighted MNSQ Fit	
	Number of items easier than expected	Number of items harder than expected	Number of items with country DIF		Number of items with fit < 0.8	Number of items with fit > 1.2
For interest in...						
No countries	26	32	13	52	52	42
1 or 2 countries	22	8	22	0	0	4
3 countries or more	4	12	17	0	0	6
N (items)	52	52	52	52	52	52
For support in...						
No countries	25	22	12	37	37	33
1 or 2 countries	11	12	20	0	0	3
3 countries or more	1	3	5	0	0	1
N (items)	37	37	37	37	37	37

Table 17.2 summarises the results of the national scaling analyses. For each attitude scale it shows the number of items that were significantly easier, harder, or different (easier or harder) compared to the pooled international sample in the following categories: (i) in no country, (ii) in only one or two countries or (iii) in three countries or more. The fourth column gives the number of items with low discrimination (item-score correlations below 0.20), the fifth column the number of items with a weighted MNSQ item fit lower than 0.8 or higher than 1.2 in each of the categories described above.

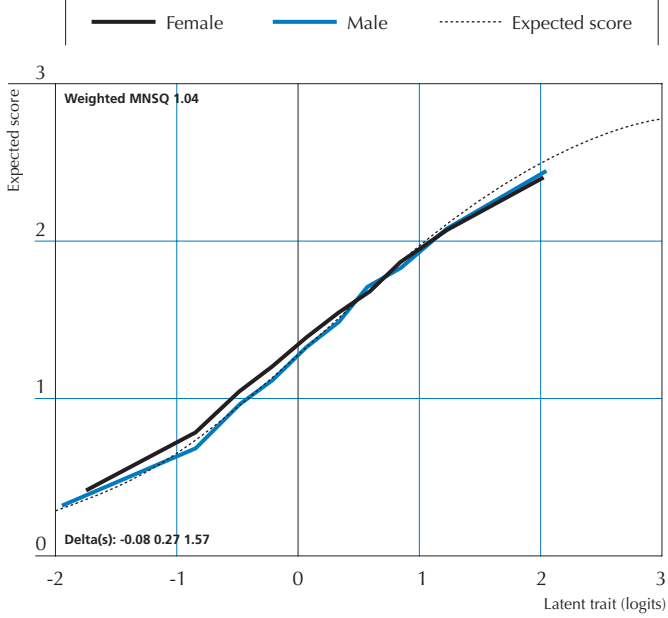


Item-by-country interactions indicate the degree of parameter invariance across countries. The results show that parameters for items measuring both *embedded interest* and *embedded support* tend to be fairly stable across countries. Furthermore, there were no items with a discrimination value less than 0.20. A full set of item-by-country interaction plots for the embedded items and cognitive items was constructed. An analysis of the item-by-country interactions shows that the embedded item parameter estimates are more stable across countries than the parameter estimates for the cognitive items.

Gender DIF

To investigate any effect of gender DIF on item performance, Expected Score Curves (ESC) were constructed and reported. A full set of plots showing the gender DIF for each embedded item was constructed. Figure 17.2 shows an example of ESC plots for item S408RNA (S408QNA recoded so that strongly agree = 3, agree = 2, disagree = 1, strongly disagree = 0). The solid line represents a predicted score and dots are observed scores for females and males separately.

Figure 17.2
An example of the ESC plot for item S408RNA



The gender DIF analysis was performed by one run of a multi-facet Rasch model, where item difficulty was modelled as a function of item, gender and item-by-gender interaction terms. Table 17.3 shows a tabular report of the gender DIF for embedded attitude items. For each item:

- The columns headed 'DIF' contain the difference between the estimates of item difficulty for girls and boys;
- The columns headed '|DIF|>0.3' provide an indicator of the magnitude of the difference. The value +1 indicate that the item is easier for males than for females and the value –1 indicate that it is easier for females than for males.

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



Table 17.3
Gender DIF table for embedded attitude items[1]

Item	DIF	|DIF|>0.3	ItemId	DIF	|DIF|>0.3	ItemId	DIF	|DIF|>0.3
S408QNA	0.01	0	S485QNB	0.2	0	S425QSA	–0.18	0
S408QNB	–0.17	0	S485QNC	0.06	0	S425QSB	–0.08	0
S408QNC	–0.14	0	S498QNA	0.03	0	S425QSC	0.02	0
S413QNA	**0.39**	**1**	S498QNB	**–0.31**	**–1**	S426QSA	0.08	0
S413QNB	**0.37**	**1**	S498QNC	–0.19	0	S426QSB	0.01	0
S413QNC	0.21	0	S508QNA	0	0	S426QSC	–0.05	0
S416QNA	–0.18	0	S508QNB	–0.15	0	S438QSA	–0.04	0
S416QNB	**–0.37**	**–1**	S508QNC	–0.13	0	S438QSB	0.16	0
S428QNA	–0.19	0	S514QNA	**0.43**	**1**	S438QSC	–0.09	0
S428QNB	**–0.33**	**–1**	S514QNB	0.5	**1**	S456QSA	–0.12	0
S428QNC	–0.15	0	S514QNC	0.32	**1**	S456QSB	–0.09	0
S437QNA	0.28	0	S519QNA	–0.04	0	S456QSC	–0.02	0
S437QNB	0.26	0	S519QNB	0.22	0	S465QSA	0.20	0
S437QNC	**0.32**	**1**	S519QNC	**0.44**	**1**	S465QSB	–0.01	0
S438QNA	0.02	0	S521QNA	**0.36**	**1**	S476QSA	0.09	0
S438QNB	0.02	0	S521QNB	0.17	0	S476QSB	–0.22	0
S438QNC	–0.13	0	S524QNA	0.17	0	S476QSC	–0.27	0
S456QNA	–0.18	0	S524QNB	–0.07	0	S477QSA	0.05	0
S456QNB	–0.09	0	S524QNC	0.09	0	S477QSB	0.05	0
S456QNC	–0.04	0	S527QNA	0.13	0	S477QSC	–0.01	0
S466QNA	–0.01	0	S527QNB	0.15	0	S485QSB	0.14	0
S466QNB	0.10	0	S527QNC	–0.12	0	S485QSC	–0.05	0
S466QNC	**–0.36**	**–1**	S408QSA	0.04	0	S498QSA	0.00	0
S476QNA	**–0.34**	**–1**	S408QSB	–0.01	0	S498QSB	–0.13	0
S476QNB	–0.25	0	S408QSC	–0.06	0	S519QSA	0.02	0
S476QNC	**–0.41**	**–1**	S416QSA	0.09	0	S519QSB	0.13	0
S478QNA	–0.20	0	S416QSB	0.08	0	S519QSC	–0.04	0
S478QNB	**–0.39**	**–1**	S416QSC	–0.02	0	S527QSB	0.17	0
S478QNC	–0.23	0	S421QSA	0.27	0	S527QSC	–0.11	0
S485QNA	–0.06	0	S421QSC	0	0			

1. Absolute values greater than 0.3 are displayed in bold in this table.

While this analysis has shown the existence of DIF for some *embedded interest* items, no substantial DIF were detected for *embedded support* items.

Summary of scalability

In summary the basic psychometric characteristics of the embedded scales appear to be sound. The existence of two factors is confirmed and both the fit to the scaling model and the reliabilities of scales appear to be adequate.

The review of differential item functioning with respect to country (item-by-country interactions) and gender shows that the embedded attitude items have fewer instances of DIF (by country and gender) than do the PISA cognitive items.

RELATIONSHIP AND COMPARISONS WITH OTHER VARIABLES

Having confirmed the adequacy of the psychometric properties of the embedded attitude scales we now consider the so-called nomothetic span of these scales. Loosely speaking, nomothetic span considers the extent to which a construct relates with other constructs in expected ways. We do this by examining the relationships of the embedded attitude scales with proficiencies, student background variables and the other PISA affective scales.



Within-country student level correlations with achievement and selected background variables

Table 0.4 shows the estimated within-country student-level correlations between *embedded interest* and *embedded support* scales and *reading, mathematics* and *science* performance and highest occupational status of parents (*HISEI*). The estimates reported in Table 17.4 were computed from weighted likelihood estimates of proficiency and then disattenuated by dividing the uncorrected correlation by the square root of the product of the reliabilities for each scale.

The correlations of *embedded support* with *reading, mathematics* and *science* have medians of 0.30, 0.24 and 0.28, respectively. For reading and science, approximately 50% of the values lie between 0.25 and 0.35, whereas for mathematics the values are typically a little lower, with 50% ranging between 0.18 and 0.28.

Table 17.4
Correlation amongst attitudinal scales, performance scales and HISEI[1]

		Correlation Embedded Support (WLE) with					Correlation Embedded Interest (WLE) with			
		Science (WLE)	Maths (WLE)	Read (WLE)	Emb. Int. (WLE)	HISEI	Science (WLE)	Maths (WLE)	Read (WLE)	HISEI
OECD	Australia	**0.38**	**0.30**	**0.33**	**0.54**	0.14	0.12	0.10	0.07	0.03
	Austria	**0.28**	**0.24**	**0.29**	**0.55**	0.13	0.06	0.01	0.02	–0.02
	Belgium	**0.29**	**0.16**	**0.22**	**0.55**	0.14	0.04	–0.04	–0.05	0.01
	Canada	**0.33**	**0.27**	**0.30**	**0.57**	0.12	0.13	0.12	0.10	0.03
	Czech Republic	**0.22**	0.13	**0.20**	**0.52**	0.03	0.04	–0.01	0.04	–0.03
	Denmark	**0.34**	**0.27**	**0.30**	**0.51**	0.15	**0.16**	0.13	0.14	0.07
	Finland	**0.31**	**0.21**	**0.35**	**0.56**	0.12	0.20	0.12	**0.21**	0.06
	France	**0.33**	**0.27**	**0.28**	**0.62**	0.14	0.13	0.11	0.10	0.04
	Germany	**0.32**	**0.30**	**0.37**	**0.61**	0.11	0.09	0.07	0.03	0.01
	Greece	**0.36**	**0.27**	**0.33**	**0.59**	**0.16**	0.07	0.03	0.09	0.00
	Hungary	**0.25**	**0.18**	**0.23**	**0.58**	0.08	–0.01	–0.02	–0.01	–0.01
	Iceland	**0.42**	**0.32**	**0.37**	**0.63**	0.13	**0.28**	0.20	**0.19**	0.09
	Ireland	**0.38**	**0.30**	**0.31**	**0.57**	0.13	0.14	0.10	0.06	0.05
	Italy	**0.27**	**0.16**	**0.27**	**0.65**	0.12	–0.04	–0.06	–0.03	–0.01
	Japan	**0.26**	**0.21**	**0.21**	**0.69**	0.07	**0.19**	0.16	0.15	0.05
	Luxembourg	**0.34**	**0.26**	**0.30**	**0.57**	0.17	0.08	0.07	0.05	0.03
	Mexico	**0.24**	**0.22**	**0.26**	**0.57**	0.10	–0.05	–0.03	–0.02	–0.11
	Netherlands	**0.26**	**0.18**	**0.22**	**0.50**	0.13	0.06	0.02	–0.04	0.03
	New Zealand	**0.35**	**0.28**	**0.34**	**0.57**	0.15	0.08	0.05	0.04	0.03
	Norway	**0.42**	**0.34**	**0.37**	**0.63**	0.15	**0.27**	**0.21**	**0.21**	0.05
	Poland	**0.30**	**0.23**	**0.29**	**0.47**	0.10	–0.06	–0.08	–0.11	–0.06
	Portugal	**0.28**	**0.24**	**0.24**	**0.60**	0.11	–0.08	–0.07	–0.11	–0.11
	Korea	**0.33**	**0.30**	**0.28**	**0.55**	0.06	0.20	0.19	**0.16**	0.07
	Scotland	**0.43**	**0.34**	**0.34**	**0.53**	**0.17**	0.20	0.16	0.15	0.11
	Slovak Republic	**0.28**	**0.24**	**0.23**	**0.55**	0.11	0.01	0.01	0.01	–0.01
	Spain	**0.26**	**0.20**	**0.20**	**0.53**	0.06	0.01	–0.01	–0.07	–0.05
	Sweden	**0.38**	**0.31**	**0.43**	**0.56**	**0.19**	0.20	0.16	0.22	0.09
	Switzerland	**0.26**	**0.18**	**0.24**	**0.55**	0.11	0.12	0.07	0.08	0.04
	Turkey	**0.38**	**0.22**	**0.37**	**0.55**	0.07	0.01	–0.06	–0.02	–0.07
	United States	**0.31**	**0.28**	**0.31**	**0.53**	0.09	–0.03	–0.07	–0.06	–0.07
Partners	Azerbaijan	0.13	0.14	0.12	**0.78**	–0.02	0.01	0.08	–0.03	–0.08
	Brazil	**0.24**	0.15	**0.25**	**0.63**	0.08	–0.16	–0.18	–0.11	–0.14
	Colombia	**0.24**	**0.16**	0.15	**0.55**	0.11	0.04	0.08	0.05	–0.01
	Croatia	**0.22**	**0.18**	**0.23**	**0.51**	0.03	–0.05	–0.08	–0.06	–0.10
	Estonia	**0.26**	**0.21**	**0.24**	**0.53**	0.09	0.01	0.00	0.00	–0.02
	Hong Kong-China	**0.35**	**0.24**	**0.26**	**0.62**	0.03	**0.23**	0.18	0.14	0.02
	Israel	**0.27**	0.15	**0.22**	**0.63**	0.00	–0.01	–0.07	–0.06	–0.12
	Jordan	**0.37**	**0.31**	**0.30**	**0.76**	0.12	0.11	0.06	0.10	–0.03
	Kyrgyzstan	0.15	0.12	**0.26**	**0.82**	0.03	–0.04	0.02	0.03	0.01
	Latvia	**0.27**	**0.20**	**0.22**	**0.51**	0.09	–0.07	–0.10	–0.10	–0.05
	Lithuania	**0.32**	**0.25**	**0.29**	**0.53**	0.09	–0.04	–0.04	–0.03	–0.05
	Macao-China	**0.33**	**0.24**	**0.31**	**0.56**	0.04	0.18	0.15	0.13	0.02
	Montenegro	**0.28**	**0.17**	**0.28**	**0.51**	0.09	0.16	0.09	0.15	0.00
	Qatar	**0.41**	**0.27**	**0.40**	**0.89**	–0.01	0.05	0.00	0.08	–0.09
	Romania	**0.33**	**0.31**	**0.35**	**0.70**	0.16	–0.04	0.02	0.07	–0.08
	Russian Federation	**0.30**	**0.24**	**0.28**	**0.53**	0.08	–0.06	–0.05	–0.08	–0.04
	Serbia	**0.22**	0.14	**0.19**	**0.51**	0.07	–0.06	–0.10	–0.06	–0.09
	Slovenia	**0.26**	0.15	**0.29**	**0.59**	0.08	0.10	0.06	0.06	–0.03
	Chinese Taipei	**0.24**	**0.17**	**0.20**	**0.55**	0.07	**0.20**	**0.17**	0.13	0.03
	Thailand	**0.25**	**0.25**	**0.31**	**0.74**	0.10	**0.18**	**0.18**	**0.20**	0.00
	Tunisia	**0.44**	**0.31**	**0.41**	**0.78**	0.13	0.13	**0.18**	0.17	–0.01

1. Absolute values greater than 0.15 are displayed in bold in this table.

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



The correlations of *embedded interest* with *reading*, *mathematics* and *science* are lower and have medians of 0.06, 0.05 and 0.05, respectively. Approximately 50% of the values lie between –0.03 and 0.15 for each of the proficiencies.

Correlations of both *embedded support* and *embedded interest* with *HISEI* are lower than the correlations with achievement variables. The median for *embedded support* is 0.10, while the median for *embedded interest* is –0.01.

To provide a frame of reference for assessing whether these results are reasonable a set of parallel correlation between relevant questionnaire variables and achievement was undertaken. The questionnaire variables that were chosen where:

- Interest in science learning: *INTSCIE*;
- Enjoyment of science: *JOYSCIE*;
- General value of science: *GENSCIE*; and
- Personal value of science: *PERSCIE*.

Table 17.5
Correlations for science scale[1]

		Correlation science (WLE) with			
		INTSCIE	JOYSCIE	GENSCIE	PERSCIE
OECD	Austria	**0.23**	**0.29**	**0.24**	0.11
	Belgium	**0.37**	**0.35**	**0.21**	**0.22**
	Canada	**0.24**	**0.34**	**0.27**	**0.27**
	Czech Republic	0.19	**0.23**	**0.25**	0.11
	Denmark	**0.22**	**0.27**	**0.24**	0.18
	Finland	0.30	0.30	**0.29**	**0.26**
	France	0.30	**0.29**	**0.23**	**0.23**
	Germany	**0.23**	**0.31**	**0.26**	0.17
	Greece	**0.23**	**0.24**	**0.23**	0.19
	Hungary	0.19	**0.24**	**0.22**	0.06
	Iceland	**0.31**	**0.42**	**0.32**	**0.32**
	Ireland	**0.31**	**0.37**	0.30	**0.31**
	Italy	0.20	**0.22**	**0.25**	0.12
	Luxembourg	0.20	**0.25**	**0.26**	0.13
	Mexico	0.05	0.06	0.14	0.03
	Netherlands	**0.23**	**0.25**	**0.29**	0.18
	Norway	0.30	**0.36**	**0.33**	**0.25**
	Poland	0.15	0.15	**0.25**	0.05
	Scotland	**0.34**	**0.37**	**0.34**	**0.31**
	Slovak Republic	0.19	0.15	**0.27**	0.04
	Spain	**0.25**	**0.33**	**0.24**	**0.24**
	Sweden	**0.29**	**0.33**	0.30	0.26
Partners	Colombia	–0.08	–0.05	0.09	–0.06
	Croatia	0.18	0.12	0.18	0.03
	Estonia	0.14	0.20	**0.28**	0.17
	Hong Kong-China	**0.28**	**0.31**	0.20	0.19
	Israel	0.15	0.23	**0.22**	0.15
	Jordan	0.15	0.16	**0.25**	0.14
	Kyrgyzstan	–0.07	–0.13	0.09	–0.1
	Latvia	0.05	0.09	**0.21**	0.07
	Lithuania	0.17	0.17	**0.24**	0.14
	Montenegro	0.13	–0.04	0.15	–0.06
	Netherlands	**0.23**	**0.25**	**0.29**	0.18
	Qatar	0.07	0.13	0.20	0.13
	Romania	0.12	0.09	**0.24**	0.05
	Russian Federation	0.08	0.11	0.15	0.03
	Serbia	0.07	–0.08	0.11	–0.08
	Slovenia	0.19	0.11	**0.25**	0.12
	Tunisia	0.14	0.12	**0.21**	0.16

1. Correlations in this table are not disattenuated for unreliability of the scales. Values greater than 0.20 are displayed in bold in this table.



The first two of the above listed variables are parallels to *embedded interest* and the second two are parallels to *embedded support*.

The estimated correlations between science proficiency and each of these four questionnaire scales are given in Table 17.4. The results reported in the table are based upon the 39 countries for which the context questionnaire data had been cleaned at the time of analysis.

The correlations of *INTSCIE*, *JOYSCIE, GENSCIE*, and *PERSCIE* with science have medians of 0.17, 0.15, 0.24 and 0.14 respectively. After accounting for the fact that these correlations have not been disattenuated for measurement error it appears that the support correlations are a little lower than the corresponding values for embedded support and the interest values are a little higher.

Relationships between embedded scales and questionnaire

Of particular interest were the relationships between the variables that quantify achievement in reading, mathematics and science, and the embedded affective variables, which were gathered using the same instruments. Similarly, of interest also were the relationships between the context questionnaire interest and support variables and the embedded affective variables, which were gathered using the different instruments but were intended to tap related constructs.

An overview of these relationships is shown in Table 17.6 which reports the results of a principal components analysis that was undertaken using the final PISA database and included all 30 OECD countries. The analysis confirms that the first component is an achievement component, the second is an interest component and the third a support component.

Table 17.6
Loadings of the achievement, interest and support variables on three varimax rotated components

	Component One	Component Two	Component Three
Science	0.956	0.054	0.081
Mathematics	0.943	0.014	0.043
Reading	0.922	0.001	0.095
Interest in science learning:	0.084	0.872	0.157
Enjoyment of science:	0.107	0.814	0.253
Embedded Interest	-0.163	0.732	0.343
General value of science	0.113	0.159	0.899
Embedded support	0.133	0.390	0.698
Personal value of science	-0.005	0.525	0.639

Table 17.7 shows the correlations, for each country, of *embedded interest* and *embedded support* with the questionnaire interest variables (*INTSCIE* and *JOYSCIE*) and questionnaire support variables (*GENSCIE*, and *PERSCIE)*. The correlations show that the embedded scales are clearly related to their parallel questionnaire scales, but they do not seem to measure exactly the same constructs.

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



Table 17.7
Correlation between embedded attitude scales and questionnaire attitude scales

		Correlation Interest (WLE) with		Correlation Support (WLE) with	
		INTSCIE	JOYSCIE	GENSCIE	PERSCIE
OECD	Australia	0.52	0.46	0.51	0.44
	Austria	0.49	0.46	0.45	0.34
	Belgium	0.47	0.43	0.40	0.36
	Canada	0.54	0.49	0.47	0.41
	Czech Republic	0.48	0.42	0.41	0.31
	Denmark	0.53	0.50	0.41	0.33
	Finland	0.58	0.50	0.48	0.40
	France	0.51	0.48	0.42	0.38
	Germany	0.51	0.49	0.46	0.38
	Greece	0.47	0.39	0.37	0.31
	Hungary	0.47	0.39	0.40	0.34
	Iceland	0.55	0.53	0.47	0.39
	Ireland	0.51	0.47	0.47	0.40
	Italy	0.45	0.40	0.39	0.34
	Japan	0.51	0.47	0.44	0.41
	Korea	0.45	0.41	0.43	0.36
	Luxembourg	0.49	0.46	0.45	0.35
	Mexico	0.43	0.37	0.36	0.32
	Netherlands	0.57	0.50	0.42	0.35
	New Zealand	0.55	0.51	0.49	0.44
	Norway	0.53	0.50	0.49	0.42
	Poland	0.42	0.32	0.40	0.30
	Portugal	0.41	0.38	0.46	0.42
	Slovak Republic	0.42	0.44	0.43	0.32
	Spain	0.43	0.41	0.41	0.35
	Sweden	0.56	0.50	0.51	0.42
	Switzerland	0.47	0.47	0.44	0.34
	Turkey	0.49	0.46	0.53	0.43
	United Kingdom	0.51	0.45	0.51	0.40
	United States	0.51	0.43	0.49	0.44
Partners	Azerbaijan	0.25	0.26	0.25	0.22
	Brazil	0.43	0.37	0.41	0.34
	Colombia	0.36	0.27	0.36	0.31
	Croatia	0.49	0.46	0.43	0.33
	Estonia	0.52	0.45	0.42	0.34
	Hong Kong-China	0.58	0.53	0.45	0.39
	Israel	0.52	0.47	0.38	0.39
	Jordan	0.40	0.32	0.39	0.35
	Kyrgyzstan	0.38	0.32	0.34	0.27
	Latvia	0.46	0.41	0.38	0.32
	Liechtenstein	0.54	0.41	0.49	0.41
	Lithuania	0.44	0.36	0.43	0.33
	Macao-China	0.50	0.46	0.41	0.34
	Montenegro	0.48	0.39	0.40	0.30
	Qatar	0.36	0.36	0.37	0.32
	Romania	0.32	0.34	0.40	0.30
	Russian Federation	0.47	0.36	0.33	0.24
	Serbia	0.45	0.44	0.39	0.30
	Slovenia	0.48	0.42	0.44	0.34
	Chinese Taipei	0.53	0.49	0.45	0.36
	Thailand	0.43	0.37	0.43	0.43
	Tunisia	0.40	0.31	0.39	0.36

Country level correlations with achievement and selected background variables

The results reported above have all been concerned with the overall student-level or the student-level within country. In this section we consider country-level relationships.

Table 17.8 shows the rank order correlations between the country means for the five cognitive domains, the four questionnaire attitude indices and for HISEI. Negative rank order correlations are shaded.



Table 17.8
Rank order correlation five test domains, questionnaire attitude scales and HISEI

	MATH	READ	SCIE	INT	SUP	INTSCIE	GENSCIE	JOYSCIE	PERSCIE
READ	0.94								
SCIE	0.95	0.95							
INT	–0.75	–0.80	–0.74						
SUP	–0.53	–0.58	–0.54	0.85					
INTSCIE	–0.69	–0.71	–0.68	0.86	0.73				
GENSCIE	–0.48	–0.47	–0.46	0.71	0.72	0.61			
JOYSCIE	–0.59	–0.62	–0.61	0.77	0.65	0.81	0.71		
PERSCIE	–0.62	–0.58	–0.59	0.73	0.64	0.71	0.90	0.80	
HISEI	0.40	0.39	0.38	–0.60	–0.53	–0.50	–0.46	–0.45	–0.40

Table 17.9
Intra-class correlation (rho)[1]

		Cognitive scales			Embedded scales		Questionnaire scales			
		SCIE	READ	MATH	Interest	Support	INTSCIE	JOYSCIE	GENSCIE	PERSCIE
OECD	Australia	0.16	0.17	0.15	0.02	0.03	0.06	0.09	0.07	0.06
	Austria	**0.44**	**0.43**	**0.43**	0.03	0.03	0.15	0.20	0.18	0.13
	Belgium	**0.33**	**0.33**	**0.35**	0.04	0.04	0.16	0.14	0.09	0.08
	Canada	0.16	0.14	0.15	0.03	0.04	0.09	0.10	0.09	0.09
	Czech Republic	**0.45**	**0.47**	**0.46**	0.04	0.02	0.11	0.12	0.09	0.11
	Denmark	0.12	0.09	0.09	0.03	0.04	0.13	0.12	0.08	0.07
	Finland	0.04	0.04	0.04	0.02	0.01	0.07	0.05	0.06	0.04
	France	**0.43**	**0.42**	**0.41**	0.04	0.04	0.12	0.12	0.11	0.10
	Germany	**0.47**	**0.47**	**0.52**	0.02	0.04	0.14	0.13	0.09	0.07
	Greece	**0.34**	**0.26**	**0.31**	0.01	0.05	0.06	0.08	0.07	0.06
	Hungary	**0.48**	**0.49**	**0.49**	0.03	0.03	0.10	0.11	0.09	0.11
	Iceland	0.06	0.05	0.06	0.02	0.02	0.07	0.08	0.07	0.07
	Ireland	0.14	0.12	0.14	0.02	0.02	0.16	**0.21**	0.17	0.14
	Italy	**0.42**	**0.38**	**0.39**	0.05	0.06	0.17	0.14	0.11	0.14
	Japan	**0.38**	**0.40**	**0.37**	0.03	0.04	0.06	0.08	0.08	0.06
	Luxembourg	**0.24**	**0.25**	**0.23**	0.01	0.02	0.09	0.11	0.09	0.10
	Mexico	**0.25**	**0.24**	**0.22**	0.06	0.03	0.06	0.07	0.04	0.02
	Netherlands	**0.49**	**0.47**	**0.47**	0.05	0.03	0.10	0.11	0.09	0.07
	New Zealand	0.13	0.11	0.13	0.03	0.04	0.07	0.07	0.08	0.08
	Norway	0.06	0.06	0.04	0.03	0.03	0.10	0.12	0.10	0.06
	Poland	0.16	0.16	0.10	0.03	0.01	0.09	0.09	0.08	0.08
	Portugal	**0.25**	**0.24**	**0.23**	0.05	0.03	0.20	0.18	0.15	0.18
	Korea	**0.29**	**0.30**	**0.27**	0.03	0.05	0.12	0.19	0.08	0.10
	Scotland	0.14	0.13	0.09	0.02	0.03	0.14	0.12	0.07	0.10
	Slovak Republic	**0.33**	**0.36**	**0.35**	0.04	0.03	0.17	0.12	0.12	0.14
	Spain	0.12	0.12	0.12	0.04	0.03	0.09	0.10	0.07	0.07
	Sweden	0.08	0.08	0.07	0.04	0.04	0.10	0.11	0.12	0.10
	Switzerland	**0.26**	**0.24**	**0.21**	0.03	0.04	0.12	0.14	0.11	0.10
	Turkey	**0.42**	**0.40**	**0.33**	0.04	0.05	0.09	0.11	0.12	0.10
	United States	0.20	0.19	0.18	0.05	0.02	0.05	0.06	0.06	0.05
Partners	Azerbaijan	**0.27**	0.19	**0.24**	0.09	0.05	0.11	0.13	0.13	0.14
	Brazil	**0.34**	**0.32**	**0.25**	0.04	0.02	0.11	0.14	0.10	0.12
	Colombia	0.19	**0.21**	0.13	0.04	0.01	0.11	0.11	0.06	0.10
	Croatia	**0.32**	**0.29**	**0.31**	0.03	0.02	0.06	0.05	0.05	0.04
	Estonia	0.15	0.15	0.20	0.05	0.02	0.08	0.09	0.08	0.07
	Hong Kong-China	**0.29**	**0.29**	**0.26**	0.01	0.02	0.16	0.12	0.10	0.10
	Israel	**0.24**	**0.28**	**0.27**	0.16	0.05	0.14	0.17	0.18	0.18
	Jordan	0.18	0.15	0.16	0.02	0.04	0.15	0.15	0.13	0.13
	Kyrgyzstan	**0.22**	**0.22**	**0.21**	0.05	0.02	0.09	0.10	0.09	0.07
	Latvia	0.14	0.15	0.14	0.06	0.03	0.04	0.04	0.04	0.02
	Liechtenstein	**0.37**	**0.33**	**0.42**	0.03	0.02	0.10	0.11	0.07	0.09
	Lithuania	**0.21**	**0.23**	0.20	0.06	0.03	0.11	0.12	0.12	0.05
	Macao – China	**0.21**	0.17	0.17	0.02	0.02	0.20	0.17	0.08	0.12
	Montenegro	0.19	0.16	0.20	0.03	0.01	0.09	0.09	0.10	0.09
	Qatar	**0.27**	**0.30**	**0.34**	0.06	0.10	0.07	0.06	0.06	0.07
	Romania	**0.36**	**0.36**	**0.33**	0.04	0.05	0.10	0.07	0.11	0.08
	Russian Federation	0.19	0.19	0.18	0.07	0.03	0.14	0.13	0.08	0.12
	Serbia	**0.30**	**0.29**	**0.28**	0.07	0.01	0.08	0.11	0.11	0.13
	Slovenia	**0.50**	**0.45**	**0.54**	0.04	0.04	0.19	0.18	0.17	0.17
	Chinese Taipei	**0.43**	**0.42**	**0.35**	0.04	0.03	0.07	0.08	0.05	0.04
	Thailand	**0.31**	**0.29**	**0.28**	0.04	0.03	0.12	0.09	0.08	0.08
	Tunisia	**0.28**	**0.29**	**0.23**	0.03	0.05	0.08	0.08	0.08	0.07

1. Values greater than 0.20 are displayed in bold in this table.



Rank order correlation coefficients between cognitive scales and attitude scales (both embedded and questionnaire) are negative at country level. The strongest negative relationship is between country ranks in *embedded interest* and performance. All attitude scales have strong positive rank correlations with each other. *HISEI* has a positive correlation with cognitive scales and a negative correlation with attitude scales.

Variance decomposition

Table 17.9 provides the intra-class correlation for each country and each cognitive domain; each embedded attitudinal domain and each questionnaire index. The intra-class correlation can be interpreted as the percentage of the total variance that is accounted for by differences among schools

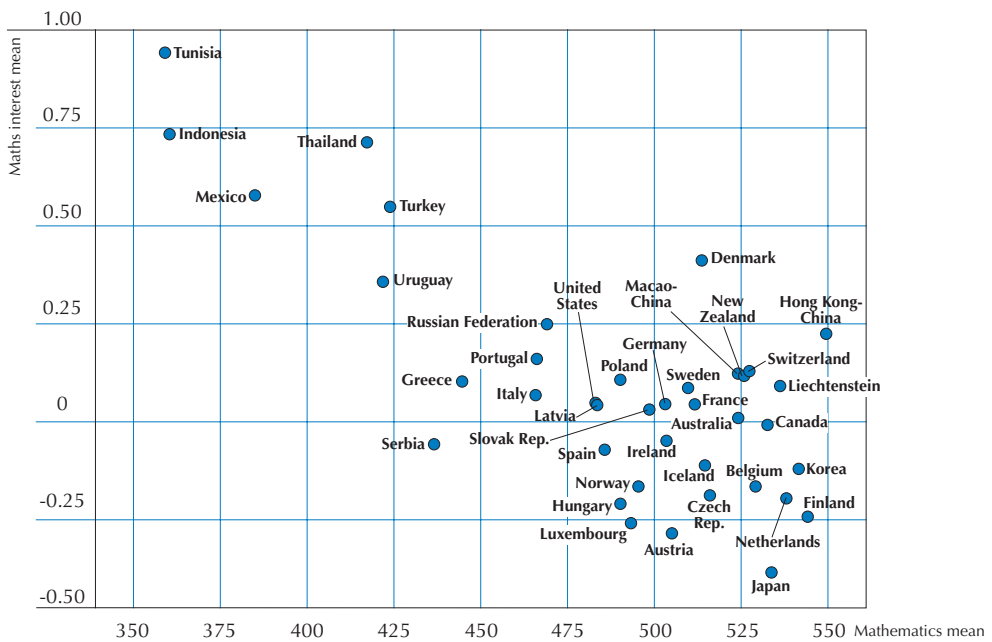
For *mathematics*, *reading* and *science* the intra-class correlation coefficient is greater than 0.20 for a number of countries, but for both the *embedded interest* and *embedded support* scale it is small for all countries. The questionnaire scales also have small intra-class correlations, although slightly larger than the embedded attitude scales. This observation is consistent with questionnaire results from previous cycles.

Observations from other cross-national data collections

We conclude the chapter by noting the relationships between similar attitudinal variables and achievement variables in PISA 2000 and 2003.

In PISA 2000 the variable closest to interest in reading (the major domain) was *JOYREAD*. For the 43 participating countries in PISA 2000 and PISA Plus the median within-country between-student correlation between reading achievement and *JOYREAD* was 0.30, with 50% of the values lying between 0.27 and 0.40. At the country level the correlation between mean reading achievement and mean *JOYREAD* was –0.63.

Figure 17.3
Scatterplot of mean mathematics interest against mean mathematics for PISA 2003



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



In PISA 2003 the interest variable *INTMATH* – mathematics being the major domain – was a close match to the INTSCIE variable included in PISA 2006. For the 40 participating countries in PISA 2003 the median within-country between-student correlation between reading achievement and *INTMATH* was 0.14, with 50% of the values lying between 0.10 and 0.24. At the country level the correlation between mean reading achievement and mean *INTMATH* was –0.76.

The country-level correlation for interest in mathematics and mathematics is shown in Figure 17.3. The correlation and the scatterplot is quite consistent with results that are observed for the attitude scales in PISA 2006 – both for the embedded attitude scales and the attitude scales that are included in the context questionnaires. Furthermore, the results are consistent with those found in other international studies such as the Trends in Mathematics and Science Study (*TIMSS*).

Summary of relations with other variables

The embedded items behave in expected and predictable ways with the other PISA variables. Principal component analysis supports that they are distinct dimensions that correlate appropriately with parallel scales that were included in the context questionnaires. Further their correlations, both at the student-level within-country and at the country level, with various other variables are consistent with observations that are made in other PISA data collections and in other studies.

CONCLUSION

The purpose of this chapter was to present analyses that support the use of the embedded scales as constructs that have the potential to provide useful and valid across-country comparisons. The purpose was not to present a comprehensive set of analyses that fully explore the relations between the embedded attitude scales and other PISA variables – such analyses will be reported elsewhere in research that draws upon the PISA databases.

The main conclusions are that embedded scales have been well constructed and are strongly supported by theory that is articulated in the PISA 2006 assessment frameworks (OECD, 2006). Statistical analysis indicates that from a psychometric perspective the embedded scales are equivalent, in terms of robustness and cross-participant validity, to the PISA cognitive scales.

In terms of their basic relationships with other variables, the embedded items generally behave in ways that are consistent with other affective variables. Our discussion of this, however, does suggest a number of important research issues that need to be explored with PISA and other data sources. Some issues that would seem worthy of pursuing are:

Why do affective variables (both embedded and otherwise) typically show a much lower intra-class correlation than do achievement variables, and to a lesser extent than do other student contextual variables?

Why do so many affective variables (both embedded and otherwise) have a negative correlation at the country level with performance measures? To what extent are these negative correlations simply examples of ecological fallacies, interpretable and important findings or cultural and misleading artifacts in the response behaviours of students?

Is there anything to be learned from the fact that lower correlations are observed between the embedded interest scales and student proficiency than between the questionnaire interest scales and student proficiency?



Notes

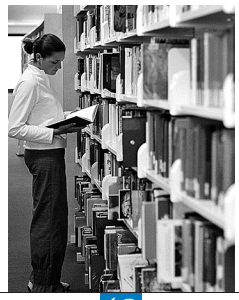
1. For an elaboration of these scales, see the *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD, 2006).

2. The reliability results presented in Chapter 12 and the IRT item parameter estimates reported in Appendix 7 are based upon a properly weighted and cleaned calibration sample and may differ a little from those reported here.

3. UH booklet is an optional one hour in length booklet, which some countries implemented in special educational needs settings (see Chapter 3).

4. Magnitudes and directions of booklet one factor loadings are very similar for the continuous and categorical approaches.

5. Note that these figures differ from those reported in Chapter 13 because the values reported in Chapter 13 were estimated using the final database.



18

International Database

Files in the database 368
- Student files 368
- School files 370
- Parent files 370

Records in the database 371
- Records included in the database 371
- Records excluded from the database 371

Representing missing data 371
How are students and schools identified? 372
Further information 373



FILES IN THE DATABASE

The PISA international database consists of six data files[1]: four with student responses, one with school responses and one with parent responses. All are provided in text (or ASCII format) with the corresponding SAS® and SPSS® control files.

Student files

Student performance and questionnaire data file (this file can be found on the PISA website *www.pisa.oecd.org*).

For each student who participated in the assessment, the following information is available:

- Identification variables for the country, school and student;
- The student responses to the two questionnaires, *i.e.*, the student questionnaire and the international option information communication technology *(ICT)* questionnaire;
- The indices derived from each student's responses to the original questions in the questionnaires;
- The students' performance scores in mathematics, reading, science, the three scales of science and embedded attitude scores in interest and support (five plausible values for each domain);
- The student weight variable and 80 Fay's replicates for the computation of the sampling variance estimates;
- Two weight factors to compute normalised (replicate) weights for multi-level analysis, one for countries and one for subnational entities;
- Three sampling related variables: the randomised final variance stratum, the final variance unit and the original explicit strata, mostly labeled by country;
- Some variables that come from the cognitive test: test language, effort variables and three science items that were internationally deleted because of students' misconceptions;
- Database version with the date of the release.

Two types of indices are provided in the student questionnaire files. The first set is based on a transformation of one variable or on a combination of the information included in two or more variables. Twenty-five indices are included in the database from this first type. The second set is the result of a Rasch scaling and consists of weighted likelihood estimate indices. Twenty-three indices from the student questionnaire and 4 indices from the international option on information communication technology questionnaire are included in the database from this second type. The index for socio-economic 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 and how to interpret them see Chapter 16.

For each domain, *i.e.* mathematics, reading and science, and for each scale in science, *i.e. identifying scientific issues, explaining phenomena scientifically* and *using scientific evidence,* a set of five plausible values (transformed to the PISA scale) are provided.

The metrics of the various scales are established so that in the year that the scale is first established the OECD student 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 is weighted to ensure that each OECD country is given equal weight.

In the case of science, the scale that was established in 2006, the average of the five plausible values means for the 30 equally weighted participating OECD countries has been set at 500 and the average of the five plausible values standard deviations has been set at 100. Note that it follows that the means and



variances of each of the five plausible values are not exactly 500 and 100. The same transformation was applied to the three science sub-scales.

Reading plausible values were mapped to the PISA 2000 scale and mathematics plausible values were mapped to the PISA 2003 scale. See chapter 12 for details of these mappings.

The variable *W_FSTUWT* is the final student weight. The sum of the weights constitutes an estimate of the size of the target population, *i.e.* the number of 15-year-old students in grade 7 or above in that country attending school. 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 2006 (OECD, 2007). 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.[2]

For a full description of the weighting methodology and the calculation of the weights, see Chapter 8). How to use weights in analysis of the database is described in detail in the PISA 2003 Data Analysis Manual for SPSS® or SAS® users (OECD, 2005), which is available through *www.pisa.oecd.org*. The data analysis manual also explains the theory behind sampling, plausible values and replication methodology and how to compute standard errors in case of two-stage, stratified sampling designs.

All student cognitive files can be found on the PISA website: *www.pisa.oecd.org*.

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 and attitude items. When original reponses 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 (without a T at the end of the variable name and the Q replaced by an R, see further below). The scored data file only has one single digit variable per item with scores instead of response categories.
- Test language;
- Effort self report;
- 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 four characters and form the first part of the variable names in the data files. The first character is R, M or S for reading, mathematics or science, respectively. The next three characters indicate the unit within the domain. For example, M155 is a mathematics unit. The item names (usually seven- or eight-digits) represent questions within a unit and are used as variable names (in the current example the item names within the unit are M155Q01, M155Q02T, M155Q03T and M155Q04T). Thus items within a unit have the same initial four characters plus a question number. Responses that needed to be recoded into single digit variables have a "T" at the end of the variable name. The original multiple digit responses have been added to the end of the unscored, single digit file without a "T" in the name and with the "Q" replaced by a "R" (for example, the variable M155Q02T is a recoded item with the corresponding original responses in M155R02 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 M155Q01 is "MATH - P2000 POPULATION PYRAMIDS (Q01)".

In all 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. The embedded attitude items have been placed after the cognitive items, first the embedded interest items followed by the embedded support 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 (this file can be found on the PISA website *www.pisa.oecd.org*).

For each school that participated in the assessment, the following information is available:

- 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;
- Explicit strata with national labels; and
- 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 14 indices from this first type. The second set is the result of a Rasch scaling and consists of weighted likelihood estimate indices. Four 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 analysis of school data see the PISA 2003 Data Analysis Manual for SPSS® or SAS® users (OECD, 2005), also available through *www.pisa.oecd.org*.

Parent file

The parent questionnaire file (this file can be found on the PISA website: *wwwpisa.oecd.org*). The following information is available:

- The identification variables for the country, school and student;
- The parents' responses on the parent questionnaire;

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



- 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 six indices from this first type. The second set is the result of a Rasch scaling and consists of weighted likelihood estimate indices. Eleven indices are included in the database from this second type. For a detailed description of the indices and how to interpret them see Chapter 9.

Due to the high parent non-response in most countries, caution is needed when analysing this data. Non-response is not random. When using the final student weights from the student file, the weights of valid students in the analysis do not sum up 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

Student and parent files

- All PISA students who attended test (assessment) 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.

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.

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



- 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 alternative answer.
- 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 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.

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 labelled *COUNTRY*. The country codes used in PISA are the ISO numerical three-digit country codes.
- A school identification variable labelled *SCHOOLID*.
- A student identification variable labelled *STIDSTD*.

A fourth variable has been included to differentiate adjudicated sub-national entities within countries. This variable (*SUBNATIO*) is used for four countries as follows:

- *Belgium*. The value "05601" is assigned to the Flemish region, "05602" to the French region and "05603" to the German region of Belgium
- *Italy*. The value "38001" is assigned to Provincia Autonoma of Bolzano, "38002" to Provincia Basilicata, "38003" to Provincia Campania, "38004" to Provincia Emilia Romagna, "38005" to Provincia Friuli Venezia Giulia, "38006" to Provincia Liguria, "38007" to Provincia Lombardia, "38008" to Provincia Piemonte, "38009" to Provincia Puglia, "38010" to Provincia Sardegna, "38011" to Provincia Sicilia, "38012" to Provincia Trento, "38013" to Provincia Veneto, "38014" to the rest of Italy.
- *Spain*. The value "72401" is assigned to Andalusia, "72402" to Aragon, "72403" to Asturias, "72406" to Cantabria, "72407" to Castile and Leon, "72409" to Catalonia, "72411" to Galicia, "72412" to La Rioja, "72415" to Navarre, "72416" to Basque Country, and
- *United Kingdom*. The value "82610" is assigned to England, Northern Ireland and Wales and the value "82620" is assigned to Scotland.

A fifth variable is added to make the identification of countries more convenient. The variable *CNT* uses the ISO 3166-1 ALPHA-3 classification, which is based on alphabetical characters rather than numeric characters (for example, for Sweden has COUNTRY=752 and CNT=SWE).

A sixth variable (*STRATUM*) is also included to differentiate sampling strata. Value labels are provided in the control files to indicate the population defined by each stratum.[3]

The school identification consists of two variables, which together form a unique identifier for each school:

- The country identification variable labelled *COUNTRY*. The country codes used in PISA are the ISO numerical three-digit country codes.
- The school identification variable labelled *SCHOOLID*.



FURTHER INFORMATION

A full description of the PISA 2006 database and guidelines on how to analyse it in accordance with the complex methodologies used to collect and process the data is provided in the *PISA 2006 Data Analysis Manual* (OECD, forthcoming) available through *www.pisa.oecd.org*.

Notes

1. Two additional data files were created and sent to countries on request. One file contains the student abilities in WLEs on the 5 domains. The other file contains plausible values for students abilities on an alternative set of science scales, the content subscales.

2. The definition of the OECD average has changed between PISA 2003 and PISA 2006. In previous cycles, 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.

3. Note that not all participants permit the identification of all sampling strata in the database.

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



References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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



Appendices

Appendix 1 PISA 2006 main study item pool characteristics 380
Appendix 2 Contrast coding used in conditioning 389
Appendix 3 Design effect tables 399
Appendix 4 Changes to core questionnaire items from 2003 to 2006 405
Appendix 5 Mapping of ISCED to years 411
Appendix 6 National household possession items 412
Appendix 7 Exploratory and confirmatory factor analyses for the embedded items 414
Appendix 8 PISA consortium, staff and consultants 416

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



APPENDIX 1 PISA 2006 main study item pool characteristics

[Part 1/1]

Table A1.1 2006 Main study reading item classification

									Item parameters (RP=.50)			Thresholds (RP=.62, PISA scale)		
Item	Name	Source	Language	Scale	Cluster	International % correct	SE % correct	Difficulty	Tau(1)	Tau(2)	Tau(3)	1	2	3
R055Q01	Drugged Spiders	CITO	English	Interpreting	R2.UHR	80.9	(0.20)	-1.53				375.9		
R055Q02	Drugged Spiders	CITO	English	Reflecting	R2.UHR	46.9	(0.25)	0.51				538.9		
R055Q03	Drugged Spiders	CITO	English	Interpreting	R2.UHR	57.5	(0.26)	-0.01				496.9		
R055Q05	Drugged Spiders	CITO	English	Interpreting	R2.UHR	71.1	(0.24)	-0.83				431.7		
R067Q01	Aesop	Greece	English/ Greek	Interpreting	R1	88.1	(0.16)	-1.86				349.0		
R067Q04	Aesop	Greece	English/ Greek	Reflecting	R1	55.6	(0.20)	0.33	-0.52	0.52		463.0	585.3	
R067Q05	Aesop	Greece	English/ Greek	Reflecting	R1	66.0	(0.23)	-0.11	0.57	-0.57		467.1	511.7	
R102Q04A	Shirts	CITO	English	Interpreting	R1	31.9	(0.22)	1.61				627.3		
R102Q05	Shirts	CITO	English	Interpreting	R1	43.3	(0.24)	0.98				576.6		
R102Q07	Shirts	CITO	English	Interpreting	R1	83.0	(0.19)	-1.40				386.0		
R104Q01	Telephone	New Zealand	English	Retrieving information	R2	80.4	(0.21)	-1.41				385.3		
R104Q02	Telephone	New Zealand	English	Retrieving information	R2	33.0	(0.23)	1.33				604.8		
R104Q05	Telephone	New Zealand	English	Retrieving information	R2	22.7	(0.15)	2.16	-1.17	1.17		571.5	772.2	
R111Q01	Exchange	Finland	Finnish	Interpreting	R2	63.4	(0.24)	-0.37				468.7		
R111Q02B	Exchange	Finland	Finnish	Reflecting	R2	33.8	(0.18)	1.23	-0.76	0.76		522.0	671.2	
R111Q06B	Exchange	Finland	Finnish	Reflecting	R2	40.7	(0.23)	0.81	0.81	-0.81		545.2	580.9	
R219Q01E	Employment	IALS	English	Interpreting	R1.UHR	57.5	(0.24)	0.30				522.6		
R219Q01T	Employment	IALS	English	Retrieving information	R1.UHR	68.8	(0.24)	-0.36				469.4		
R219Q02	Employment	IALS	English	Reflecting	R1.UHR	79.1	(0.21)	-1.14				406.6		
R220Q01	South Pole	France	French	Retrieving information	R1	42.2	(0.25)	1.03				580.9		
R220Q02B	South Pole	France	French	Interpreting	R1	61.1	(0.25)	0.03				500.7		
R220Q04	South Pole	France	French	Interpreting	R1	58.9	(0.25)	0.20				514.5		
R220Q05	South Pole	France	French	Interpreting	R1	80.9	(0.21)	-1.25				397.8		
R220Q06	South Pole	France	French	Interpreting	R1	65.9	(0.23)	-0.21				481.2		
R227Q01	Optician	Switzerland	German	Interpreting	R2	52.1	(0.24)	0.23				517.0		
R227Q02T	Optician	Switzerland	German	Retrieving information	R2	54.9	(0.17)	0.14	-1.08	1.08		414.8	603.5	
R227Q03	Optician	Switzerland	German	Reflecting	R2	53.2	(0.25)	0.26				518.9		
R227Q06	Optician	Switzerland	German	Retrieving information	R2	69.3	(0.24)	-0.67				444.9		

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



[Part 1/2]

Table A1.2 2006 Main study mathematics item classification

Item	Name	Source	Language	Scale	Cluster
M033Q01	P2000 A View Room	Consortium	Dutch	Space and Shape	M2
M034Q01T	P2000 Bricks	Consortium	Dutch	Space and Shape	M2
M155Q01	P2000 Population Pyramids	Consortium	Dutch	Change and Relationships	M2
M155Q02T	P2000 Population Pyramids	Consortium	Dutch	Change and Relationships	M2
M155Q03T	P2000 Population Pyramids	Consortium	Dutch	Space and Shape	M2
M155Q04T	P2000 Population Pyramids	Consortium	Dutch	Space and Shape	M2
M192Q01T	P2000 Containers	Germany	German	Space and Shape	M4
M273Q01T	P2000 Pipelines	Czech Republic	Czech	Space and Shape	M3
M302Q01T	Car Drive	TIMSS	English	Space and Shape	M1.UHM
M302Q02	Car Drive	TIMSS	English	Change and Relationships	M1.UHM
M302Q03	Car Drive	TIMSS	English	Change and Relationships	M1.UHM
M305Q01	Map	Consortium	English	Change and Relationships	M4
M406Q01	Running Tracks	Consortium	English	Change and Relationships	M4
M406Q02	Running Tracks	Consortium	English	Change and Relationships	M4
M408Q01T	Lotteries	Consortium	English	Change and Relationships	M3
M411Q01	Diving	Consortium	English	Change and Relationships	M2
M411Q02	Diving	Consortium	English	Uncertainty	M2
M420Q01T	Transport	Consortium	English	Change and Relationships	M3
M421Q01	Height	Consortium	English	Space and Shape	M1
M421Q02T	Height	Consortium	English	Space and Shape	M1
M421Q03	Height	Consortium	English	Change and Relationships	M1
M423Q01	Tossing Coins	Consortium	English	Change and Relationships	M4
M442Q02	Braille	Consortium	English	Change and Relationships	M2
M446Q01	Thermometer Cricket	Consortium	English	Space and Shape	M3
M446Q02	Thermometer Cricket	Consortium	English	Change and Relationships	M3
M447Q01	Tile Arrangement	Consortium	English	Change and Relationships	M3
M462Q01T	Third Side	Sweden	English	Space and Shape	M2.UHM
M464Q01T	The Fence	Sweden	English	Space and Shape	M3
M474Q01	Running Time	Canada	English	Space and Shape	M2
M496Q01T	Cash Withdrawal	Consortium	English	Uncertainty	M4
M496Q02	Cash Withdrawal	Consortium	English	Quantity	M4
M559Q01	Telephone Rates	Italy	English	Uncertainty	M3
M564Q01	Chair Lift	Italy	English	Quantity	M4.UHM
M564Q02	Chair Lift	Italy	English	Quantity	M4.UHM
M571Q01	Stop The Car	Germany	German	Quantity	M4
M598Q01	Making A Booklet	Switzerland	German	Uncertainty	M1
M603Q01T	Number Check	Austria	German	Uncertainty	M4
M603Q02T	Number Check	Austria	German	Uncertainty	M4
M710Q01	Forecast of Rain	Consortium	Japanese	Uncertainty	M1
M800Q01	Computer Game	Canada	English	Uncertainty	M3.UHM
M803Q01T	Labels	Canada	English	Quantity	M2
M810Q01T	Bicycles	Canada	English	Uncertainty	M1
M810Q02T	Bicycles	Canada	English	Uncertainty	M1
M810Q03T	Bicycles	Canada	English	Quantity	M1
M828Q01	Carbon Dioxide	The Netherlands	English	Change and Relationships	M3
M828Q02	Carbon Dioxide	The Netherlands	English	Change and Relationships	M3
M828Q03	Carbon Dioxide	The Netherlands	English	Space and Shape	M3
M833Q01T	Seeing the tower	The Netherlands	English	Space and Shape	M1

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



[Part 2/2]

Table A1.2 2006 Main study mathematics item classification

				Item parameters (RP=.50)			Thresholds (RP=.62, PISA scale)		
Item	International % correct	SE % correct	Difficulty	Tau(1)	Tau(2)	Tau(3)	1	2	3
M033Q01	76.7	(0.21)	-1.54				429.6		
M034Q01T	43.3	(0.27)	0.27				571.4		
M155Q01	64.7	(0.25)	-0.82				486.2		
M155Q02T	60.9	(0.24)	-0.51	0.76	-0.76		492.6	528.9	
M155Q03T	19.1	(0.18)	1.46	0.10	-0.10		629.8	698.3	
M155Q04T	55.7	(0.25)	-0.37				521.5		
M192Q01T	40.3	(0.24)	0.36				578.1		
M273Q01T	53.5	(0.25)	-0.34				523.3		
M302Q01T	95.5	(0.10)	-3.83				251.9		
M302Q02	80.6	(0.19)	-1.97				396.7		
M302Q03	29.1	(0.23)	1.02				629.3		
M305Q01	61.5	(0.24)	-0.63				500.8		
M406Q01	27.4	(0.24)	1.14				639.0		
M406Q02	17.0	(0.20)	1.93				700.5		
M408Q01T	44.0	(0.25)	0.15				561.6		
M411Q01	50.4	(0.27)	-0.10				542.2		
M411Q02	44.7	(0.25)	0.22				567.8		
M420Q01T	48.7	(0.25)	-0.08				543.4		
M421Q01	62.6	(0.26)	-0.78				489.3		
M421Q02T	16.4	(0.18)	1.93				700.5		
M421Q03	34.2	(0.23)	0.77				610.3		
M423Q01	79.9	(0.20)	-1.84				406.5		
M442Q02	39.1	(0.26)	0.55				592.7		
M446Q01	67.3	(0.26)	-1.00				472.2		
M446Q02	7.0	(0.13)	3.04				786.8		
M447Q01	68.5	(0.23)	-1.10				464.3		
M462Q01T	11.9	(0.15)	1.97	0.40	-0.40		677.6	728.7	
M464Q01T	24.7	(0.23)	1.28				649.3		
M474Q01	73.6	(0.22)	-1.36				444.3		
M496Q01T	50.1	(0.25)	-0.06				545.2		
M496Q02	64.0	(0.24)	-0.85				483.8		
M559Q01	63.5	(0.24)	-0.76				491.1		
M564Q01	46.9	(0.25)	0.11				558.6		
M564Q02	46.2	(0.25)	0.13				560.5		
M571Q01	47.4	(0.26)	0.06				554.4		
M598Q01	59.8	(0.25)	-0.67				498.4		
M603Q01T	45.0	(0.25)	0.15				562.3		
M603Q02T	34.8	(0.25)	0.79				611.0		
M710Q01	32.3	(0.23)	0.84				615.2		
M800Q01	89.4	(0.15)	-2.75				335.9		
M803Q01T	29.6	(0.24)	1.10				636.0		
M810Q01T	61.8	(0.25)	-0.74				492.3		
M810Q02T	69.0	(0.24)	-1.18				458.2		
M810Q03T	19.1	(0.18)	1.54	-0.03	0.03		631.4	708.6	
M828Q01	36.4	(0.25)	0.59				596.4		
M828Q02	54.6	(0.25)	-0.30				526.4		
M828Q03	28.9	(0.23)	1.12				637.8		
M833Q01T	30.2	(0.23)	1.07				633.5		



[Part 1/4]

Table A1.3 2006 Main study science item classification (cognitive)

Item	Name	Source	Language	Scale	Cluster
S114Q03T	Greenhouse	CITO	Dutch	Using scientific evidence	S2
S114Q04T	Greenhouse	CITO	Dutch	Using scientific evidence	S2
S114Q05T	Greenhouse	CITO	Dutch	Explaining phenomena scientifically	S2
S131Q02T	Good Vibrations	ACER	English	Using scientific evidence	S5
S131Q04T	Good Vibrations	ACER	English	Identifying scientific issues	S5
S213Q01T	Clothes	Australia	English	Identifying scientific issues	S7
S213Q02	Clothes	Australia	English	Explaining phenomena scientifically	S7
S256Q01	Spoons	TIMSS	English	Explaining phenomena scientifically	S5.UHS
S268Q01	Algae	Australia	English	Identifying scientific issues	S3
S268Q02T	Algae	Australia	English	Explaining phenomena scientifically	S3
S268Q06	Algae	Australia	English	Explaining phenomena scientifically	S3
S269Q01	Earth's Temperature	CITO	Dutch	Explaining phenomena scientifically	S1
S269Q03T	Earth's Temperature	CITO	Dutch	Explaining phenomena scientifically	S1
S269Q04T	Earth's Temperature	CITO	Dutch	Explaining phenomena scientifically	S1
S304Q01	Water	CITO	Dutch	Using scientific evidence	S6
S304Q02	Water	CITO	Dutch	Explaining phenomena scientifically	S6
S304Q03A	Water	CITO	Dutch	Using scientific evidence	S6
S304Q03B	Water	CITO	Dutch	Explaining phenomena scientifically	S6
S326Q01	Milk	CITO	Dutch	Using scientific evidence	S4
S326Q02	Milk	CITO	Dutch	Using scientific evidence	S4
S326Q03	Milk	CITO	Dutch	Using scientific evidence	S4
S326Q04T	Milk	CITO	Dutch	Explaining phenomena scientifically	S4
S408Q01	Wild Oat Grass	ILS	Norwegian	Explaining phenomena scientifically	S4
S408Q03	Wild Oat Grass	ILS	Norwegian	Explaining phenomena scientifically	S4
S408Q04T	Wild Oat Grass	ILS	Norwegian	Explaining phenomena scientifically	S4
S408Q05	Wild Oat Grass	ILS	Norwegian	Identifying scientific issues	S4
S413Q04T	Plastic Age	IPN	German	Using scientific evidence	S5
S413Q05	Plastic Age	IPN	German	Using scientific evidence	S5
S413Q06	Plastic Age	IPN	German	Explaining phenomena scientifically	S5
S415Q02	Solar Panels	NIER	Japanese	Explaining phenomena scientifically	S4
S415Q07T	Solar Panels	ACER	English	Identifying scientific issues	S4
S415Q08T	Solar Panels	ACER	English	Identifying scientific issues	S4
S416Q01	The Moon	ILS	Norwegian	Using scientific evidence	S7
S421Q01	Big and Small	ILS	Norwegian	Explaining phenomena scientifically	S7.UHS
S421Q03	Big and Small	ILS	Norwegian	Explaining phenomena scientifically	S7.UHS
S425Q02	Penguin Island	ACER	English	Using scientific evidence	S7
S425Q03	Penguin Island	ACER	English	Explaining phenomena scientifically	S7
S425Q04	Penguin Island	ACER	English	Using scientific evidence	S7
S425Q05	Penguin Island	ACER	English	Identifying scientific issues	S7
S426Q03	The Grand Canyon	ACER	English	Explaining phenomena scientifically	S1
S426Q05	The Grand Canyon	ACER	English	Explaining phenomena scientifically	S1
S426Q07T	The Grand Canyon	ACER	English	Identifying scientific issues	S1
S428Q01	Bacteria in Milk	IPN	German	Using scientific evidence	S6.UHS
S428Q03	Bacteria in Milk	IPN	German	Using scientific evidence	S6.UHS
S428Q05	Bacteria in Milk	IPN	German	Explaining phenomena scientifically	S6.UHS
S437Q01	Extinguishing Fires	ACER	German	Explaining phenomena scientifically	S4
S437Q03	Extinguishing Fires	ACER	English	Explaining phenomena scientifically	S4
S437Q04	Extinguishing Fires	ACER	English	Explaining phenomena scientifically	S4
S437Q06	Extinguishing Fires	ACER	English	Explaining phenomena scientifically	S4
S438Q01T	Green Parks	ACER	English	Identifying scientific issues	S6
S438Q02	Green Parks	ACER	English	Identifying scientific issues	S6
S438Q03T	Green Parks	ACER	English	Identifying scientific issues	S6

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



[Part 2/4]

Table A1.3 2006 Main study science item classification (cognitive)

				Item parameters (RP=.50)			Thresholds (RP=.62, PISA scale)		
Item	International % correct	SE % correct	Difficulty	Tau(1)	Tau(2)	Tau(3)	1	2	3
S114Q03T	53.9	(0.26)	0.00				529.6		
S114Q04T	34.5	(0.21)	0.91	-0.01	0.01		568.2	659.3	
S114Q05T	18.9	(0.19)	1.93				709.6		
S131Q02T	46.2	(0.26)	0.29				556.6		
S131Q04T	31.1	(0.23)	1.18				639.7		
S213Q01T	47.9	(0.26)	0.41				566.8		
S213Q02	79.4	(0.20)	-1.39				399.2		
S256Q01	87.8	(0.16)	-2.20				324.2		
S268Q01	72.5	(0.22)	-0.83				451.7		
S268Q02T	36.2	(0.24)	1.01				622.9		
S268Q06	55.2	(0.25)	0.10				538.4		
S269Q01	57.8	(0.25)	-0.28				502.7		
S269Q03T	41.2	(0.25)	0.59				583.5		
S269Q04T	34.1	(0.23)	0.94				617.0		
S304Q01	43.6	(0.25)	0.42				568.2		
S304Q02	62.1	(0.25)	-0.47				485.2		
S304Q03A	39.0	(0.24)	0.76				599.6		
S304Q03B	50.7	(0.26)	0.11				539.1		
S326Q01	59.0	(0.24)	-0.16				513.6		
S326Q02	63.7	(0.25)	-0.44				487.4		
S326Q03	58.3	(0.25)	-0.18				512.1		
S326Q04T	23.3	(0.22)	1.73				689.9		
S408Q01	62.9	(0.23)	-0.35				496.1		
S408Q03	30.5	(0.23)	1.28				647.6		
S408Q04T	50.7	(0.24)	0.25				552.2		
S408Q05	42.0	(0.24)	0.71				594.4		
S413Q04T	41.4	(0.25)	0.59				583.5		
S413Q05	65.6	(0.24)	-0.71				462.6		
S413Q06	37.8	(0.26)	0.81				604.0		
S415Q02	78.3	(0.21)	-1.28				410.2		
S415Q07T	72.1	(0.23)	-0.80				454.6		
S415Q08T	57.7	(0.25)	-0.04				525.3		
S416Q01	45.4	(0.25)	0.54				579.2		
S421Q01	39.8	(0.26)	0.83				606.8		
S421Q03	63.0	(0.25)	-0.42				489.5		
S425Q02	45.8	(0.25)	0.51				576.3		
S425Q03	41.4	(0.25)	0.68				592.3		
S425Q04	30.1	(0.23)	1.33				652.7		
S425Q05	69.0	(0.23)	-0.73				461.1		
S426Q03	67.6	(0.24)	-0.83				451.7		
S426Q05	75.8	(0.22)	-1.26				411.6		
S426Q07T	61.3	(0.23)	-0.47				485.2		
S428Q01	61.7	(0.24)	-0.46				485.9		
S428Q03	71.3	(0.24)	-1.08				428.4		
S428Q05	43.9	(0.26)	0.44				569.7		
S437Q01	72.2	(0.23)	-0.93				442.2		
S437Q03	49.4	(0.26)	0.33				559.5		
S437Q04	58.0	(0.24)	-0.15				514.3		
S437Q06	76.0	(0.22)	-1.15				421.8		
S438Q01T	83.2	(0.19)	-1.91				351.2		
S438Q02	65.6	(0.24)	-0.64				469.1		
S438Q03T	38.9	(0.25)	0.68				592.3		



[Part 3/4]

Table A1.3 2006 Main study science item classification (cognitive)

Item	Name	Source	Language	Scale	Cluster
S447Q02	Sunscreens	ACER	English	Identifying scientific issues	S5
S447Q03	Sunscreens	ACER	English	Identifying scientific issues	S5
S447Q04	Sunscreens	ACER	English	Identifying scientific issues	S5
S447Q05	Sunscreens	ACER	English	Using scientific evidence	S5
S458Q01	The Ice Mummy	ILS	Norwegian	Explaining phenomena scientifically	S6
S458Q02T	The Ice Mummy	ILS	Norwegian	Using scientific evidence	S6
S465Q01	Different Climates	ILS	Norwegian	Using scientific evidence	S5
S465Q02	Different Climates	ILS	Norwegian	Explaining phenomena scientifically	S5
S465Q04	Different Climates	ILS	Norwegian	Explaining phenomena scientifically	S5
S466Q01T	Forest Fires	ILS	Norwegian	Identifying scientific issues	S6.UHS
S466Q05	Forest Fires	ILS	Norwegian	Using scientific evidence	S6.UHS
S466Q07T	Forest Fires	ILS	Norwegian	Identifying scientific issues	S6.UHS
S476Q01	Heart Surgery	New Zealand	English	Explaining phenomena scientifically	S2.UHS
S476Q02	Heart Surgery	New Zealand	English	Explaining phenomena scientifically	S2.UHS
S476Q03	Heart Surgery	New Zealand	English	Explaining phenomena scientifically	S2.UHS
S477Q02	Mary Montagu	Norway	Norwegian	Explaining phenomena scientifically	S3
S477Q03	Mary Montagu	Norway	Norwegian	Explaining phenomena scientifically	S3
S477Q04	Mary Montagu	Norway	Norwegian	Explaining phenomena scientifically	S3
S478Q01	Antibiotics	France	French	Explaining phenomena scientifically	S5
S478Q02T	Antibiotics	France	French	Using scientific evidence	S5
S478Q03T	Antibiotics	France	French	Explaining phenomena scientifically	S5
S485Q02	Acid Rain	Greece	English/Greek	Explaining phenomena scientifically	S1
S485Q03	Acid Rain	ACER	English	Using scientific evidence	S1
S485Q05	Acid Rain	ACER	English	Identifying scientific issues	S1
S493Q01T	Physical Exercise	Switzerland	French	Explaining phenomena scientifically	S7
S493Q03T	Physical Exercise	Switzerland	French	Explaining phenomena scientifically	S7
S493Q05T	Physical Exercise	Switzerland	French	Explaining phenomena scientifically	S7
S495Q01T	Radiotherapy	France	French	Using scientific evidence	S2
S495Q02T	Radiotherapy	France	French	Using scientific evidence	S2
S495Q03	Radiotherapy	France	French	Using scientific evidence	S2
S495Q04T	Radiotherapy	France	French	Identifying scientific issues	S2
S498Q02T	Experimental Digestion	France	French	Identifying scientific issues	S3
S498Q03	Experimental Digestion	France	French	Identifying scientific issues	S3
S498Q04	Experimental Digestion	France	French	Using scientific evidence	S3
S508Q02T	Genetically Modified Crops	United Kingdom	English	Identifying scientific issues	S1
S508Q03	Genetically Modified Crops	United Kingdom	English	Identifying scientific issues	S1
S510Q01T	Magnetic Hovertrain	Belgium	Dutch	Explaining phenomena scientifically	S4
S510Q04T	Magnetic Hovertrain	Belgium	Dutch	Explaining phenomena scientifically	S4
S514Q02	Development and Disaster	NIER	Japanese	Using scientific evidence	S7
S514Q03	Development and Disaster	NIER	Japanese	Explaining phenomena scientifically	S7
S514Q04	Development and Disaster	NIER	Japanese	Using scientific evidence	S7
S519Q01	Airbags	France	French	Using scientific evidence	S3
S519Q02T	Airbags	France	French	Explaining phenomena scientifically	S3
S519Q03	Airbags	France	French	Identifying scientific issues	S3
S521Q02	Cooking Outdoors	ACER	English	Explaining phenomena scientifically	S2
S521Q06	Cooking Outdoors	ACER	English	Explaining phenomena scientifically	S2
S524Q06T	Penicillin Manufacture	IPN	German	Using scientific evidence	S3
S524Q07	Penicillin Manufacture	IPN	German	Using scientific evidence	S3
S527Q01T	Extinction of the Dinosaurs	Korea	Korean	Using scientific evidence	S1
S527Q03T	Extinction of the Dinosaurs	Korea	Korean	Explaining phenomena scientifically	S1
S527Q04T	Extinction of the Dinosaurs	Korea	Korean	Explaining phenomena scientifically	S1

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



[Part 4/4]

Table A1.3 2006 Main study science item classification (cognitive)

				Item parameters (RP=.50)			Thresholds (RP=.62, PISA scale)		
Item	International % correct	SE % correct	Difficulty	Tau(1)	Tau(2)	Tau(3)	1	2	3
S447Q02	40.5	(0.24)	0.64				588.7		
S447Q03	58.3	(0.23)	-0.31				499.7		
S447Q04	43.0	(0.24)	0.48				574.0		
S447Q05	27.1	(0.22)	1.01	2.01	-2.01		616.4	629.0	
S458Q01	16.3	(0.19)	2.09				724.2		
S458Q02T	56.2	(0.25)	-0.18				512.1		
S465Q01	50.2	(0.22)	0.13	0.09	-0.09		499.7	582.0	
S465Q02	60.9	(0.24)	-0.35				496.8		
S465Q04	36.3	(0.24)	0.89				612.0		
S466Q01T	71.0	(0.22)	-1.01				434.9		
S466Q05	55.7	(0.24)	-0.15				515.1		
S466Q07T	74.9	(0.22)	-1.23				413.8		
S476Q01	70.7	(0.24)	-0.91				444.4		
S476Q02	70.9	(0.22)	-0.91				443.7		
S476Q03	60.1	(0.25)	-0.32				499.1		
S477Q02	74.9	(0.22)	-0.99				436.4		
S477Q03	75.1	(0.22)	-1.05				431.3		
S477Q04	61.7	(0.25)	-0.23				507.1		
S478Q01	42.8	(0.23)	0.53				577.7		
S478Q02T	51.0	(0.25)	0.04				532.5		
S478Q03T	67.7	(0.23)	-0.75				459.0		
S485Q02	57.7	(0.26)	-0.24				506.3		
S485Q03	66.7	(0.25)	-0.74				460.4		
S485Q05	35.5	(0.18)	0.92	-0.97	0.97		513.6	716.9	
S493Q01T	52.6	(0.25)	0.17				544.9		
S493Q03T	82.4	(0.18)	-1.53				386.1		
S493Q05T	45.1	(0.25)	0.57				582.8		
S495Q01T	42.1	(0.24)	0.55				579.9		
S495Q02T	57.6	(0.25)	-0.18				512.1		
S495Q03	38.6	(0.26)	0.82				604.7		
S495Q04T	50.2	(0.25)	0.18				545.7		
S498Q02T	46.9	(0.24)	0.49				574.8		
S498Q03	42.6	(0.24)	0.68				592.3		
S498Q04	59.9	(0.25)	-0.05	1.03	-1.03		507.7	540.9	
S508Q02T	60.9	(0.23)	-0.44				488.1		
S508Q03	73.6	(0.23)	-1.15				421.8		
S510Q01T	53.9	(0.23)	0.08				536.2		
S510Q04T	41.0	(0.24)	0.77				600.3		
S514Q02	85.2	(0.20)	-1.85				356.2		
S514Q03	46.6	(0.25)	0.49				574.0		
S514Q04	52.2	(0.27)	0.14				542.0		
S519Q01	35.3	(0.21)	0.92	-0.01	0.01		569.7	660.4	
S519Q02T	52.6	(0.25)	0.18				545.7		
S519Q03	28.7	(0.22)	1.39				657.9		
S521Q02	55.9	(0.24)	-0.11				518.7		
S521Q06	88.1	(0.18)	-2.14				329.2		
S524Q06T	64.3	(0.24)	-0.40				491.8		
S524Q07	36.5	(0.24)	1.04				625.8		
S527Q01T	16.1	(0.18)	2.10				724.9		
S527Q03T	58.0	(0.25)	-0.25				504.8		
S527Q04T	53.7	(0.24)	-0.03				526.7		

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



[Part 1/1]

Table A1.4 2006 Main study science embedded item classification (interest in learning science topics)

								Item parameters (RP=.50)			Thresholds (RP=.62, PISA scale)		
Item	Name	Source	Language	Cluster	International % correct	SE % correct	Difficulty	Tau(1)	Tau(2)	Tau(3)	1	2	3
S408QNA	Wild Oat Grass	ACER	English	S4	48.3	(0.16)	0.38	-1.22	-0.08	1.30	430.7	557.8	695.4
S408QNB	Wild Oat Grass	ACER	English	S4	45.4	(0.15)	0.52	-1.47	-0.05	1.53	425.5	571.0	725.4
S408QNC	Wild Oat Grass	ACER	English	S4	46.2	(0.15)	0.51	-1.47	-0.11	1.58	424.5	567.6	727.8
S413QNA	Plastic Age	IPN	German	S5	46.1	(0.15)	0.48	-1.69	0.00	1.69	406.0	570.4	735.1
S413QNB	Plastic Age	IPN	German	S5	48.0	(0.15)	0.35	-1.63	-0.04	1.68	398.7	557.1	722.6
S413QNC	Plastic Age	IPN	German	S5	38.1	(0.16)	0.84	-1.27	0.11	1.16	471.2	608.0	729.6
S416QNA	The Moon	IPN	German	S7	55.1	(0.15)	-0.13	-1.40	-0.34	1.73	369.4	497.8	680.7
S416QNB	The Moon	IPN	German	S7	64.6	(0.15)	-0.65	-1.10	-0.36	1.46	343.4	452.4	613.0
S428QNA	Bacteria in Milk	ACER	English	S6.UHS	51.3	(0.14)	0.19	-1.64	-0.19	1.83	382.2	534.1	718.4
S428QNB	Bacteria in Milk	ACER	English	S6.UHS	51.9	(0.15)	0.15	-1.64	-0.09	1.73	379.8	536.2	708.0
S428QNC	Bacteria in Milk	ACER	English	S6.UHS	51.8	(0.15)	0.14	-1.43	-0.07	1.49	395.2	536.9	689.1
S437QNA	Extinguishing Fires	ACER	English	S4	60.5	(0.15)	-0.32	-1.46	-0.28	1.74	348.0	483.1	665.3
S437QNB	Extinguishing Fires	ACER	English	S4	55.0	(0.15)	-0.03	-1.54	-0.15	1.69	370.7	517.3	688.4
S437QNC	Extinguishing Fires	ACER	English	S4	64.2	(0.15)	-0.48	-1.15	-0.14	1.29	358.8	478.2	617.2
S438QNA	Green Parks	ACER	English	S6	39.9	(0.15)	0.80	-1.49	0.02	1.47	449.6	600.5	747.0
S438QNB	Green Parks	ACER	English	S6	37.0	(0.15)	0.95	-1.50	0.07	1.43	463.5	616.1	758.2
S438QNC	Green Parks	ACER	English	S6	43.2	(0.15)	0.61	-1.30	-0.04	1.34	446.8	580.2	719.8
S456QNA	The Cheetah	IPN	German	S2	58.0	(0.14)	-0.29	-1.69	-0.22	1.91	335.1	489.0	682.1
S456QNB	The Cheetah	IPN	German	S2	60.2	(0.14)	-0.41	-1.59	-0.22	1.81	331.6	478.2	662.6
S456QNC	The Cheetah	IPN	German	S2	64.1	(0.15)	-0.58	-1.30	-0.19	1.48	338.6	467.0	622.8
S466QNA	Forest Fires	ACER	English	S6.UHS	59.6	(0.14)	-0.27	-1.61	-0.30	1.90	342.1	485.9	682.1
S466QNB	Forest Fires	ACER	English	S6.UHS	54.6	(0.14)	-0.09	-1.73	-0.03	1.76	351.1	517.3	689.5
S466QNC	Forest Fires	ACER	English	S6.UHS	65.4	(0.15)	-0.60	-1.19	-0.34	1.53	341.4	456.6	622.8
S476QNA	Heart Surgery	IPN	German	S2.UHS	58.8	(0.14)	-0.32	-1.67	-0.22	1.89	333.7	485.9	677.2
S476QNB	Heart Surgery	IPN	German	S2.UHS	57.6	(0.14)	-0.27	-1.50	-0.05	1.54	353.6	501.3	656.6
S476QNC	Heart Surgery	IPN	German	S2.UHS	52.0	(0.15)	0.09	-1.42	-0.06	1.48	391.0	532.7	683.1
S478QNA	Antibiotics	IPN	German	S5	59.2	(0.14)	-0.24	-1.47	-0.15	1.62	357.5	498.5	663.9
S478QNB	Antibiotics	IPN	German	S5	60.2	(0.14)	-0.37	-1.53	-0.07	1.60	342.1	491.2	652.1
S478QNC	Antibiotics	IPN	German	S5	60.9	(0.15)	-0.31	-1.21	-0.12	1.33	369.4	494.3	636.0
S485QNA	Acid Rain	IPN	German	S1	56.7	(0.15)	-0.11	-1.56	-0.17	1.72	362.0	508.9	683.8
S485QNB	Acid Rain	IPN	German	S1	56.2	(0.16)	-0.12	-1.38	0.01	1.37	376.3	517.3	656.9
S485QNC	Acid Rain	IPN	German	S1	48.9	(0.16)	0.27	-1.53	-0.01	1.53	400.4	551.9	704.7
S498QNA	Experimental Digestion	IPN	German	S3	46.8	(0.14)	0.39	-1.62	-0.09	1.71	402.1	557.8	728.2
S498QNB	Experimental Digestion	IPN	German	S3	54.5	(0.14)	-0.04	-1.59	-0.10	1.69	365.8	518.7	687.7
S498QNC	Experimental Digestion	IPN	German	S3	59.3	(0.15)	-0.28	-1.27	-0.27	1.54	365.8	489.0	652.7
S508QNA	Genetically Modified Crops	ACER	English	S1	46.2	(0.15)	0.43	-1.45	-0.14	1.59	417.9	558.5	721.5
S508QNB	Genetically Modified Crops	ACER	English	S1	46.1	(0.15)	0.45	-1.42	-0.15	1.56	422.3	559.9	720.5
S508QNC	Genetically Modified Crops	ACER	English	S1	47.0	(0.16)	0.35	-1.26	-0.10	1.36	425.5	554.3	697.4
S514QNA	Development and Disaster	ACER	English	S7	51.6	(0.15)	0.00	-1.47	-0.05	1.52	379.5	525.4	679.3
S514QNB	Development and Disaster	ACER	English	S7	47.9	(0.15)	0.22	-1.53	-0.03	1.56	395.2	545.9	701.3
S514QNC	Development and Disaster	ACER	English	S7	65.9	(0.15)	-0.71	-1.14	-0.23	1.37	337.2	453.1	601.8
S519QNA	Airbags	ACER	English	S3	71.3	(0.13)	-1.04	-1.49	-0.28	1.77	282.0	418.9	602.5
S519QNB	Airbags	ACER	English	S3	69.4	(0.14)	-0.89	-1.44	-0.21	1.66	299.5	436.4	607.4
S519QNC	Airbags	ACER	English	S3	65.7	(0.14)	-0.66	-1.33	-0.06	1.39	330.9	465.7	610.2
S521QNA	Cooking Outdoors	ACER	English	S2	40.3	(0.14)	0.69	-1.56	0.00	1.55	435.7	589.9	743.5
S521QNB	Cooking Outdoors	ACER	English	S2	41.1	(0.14)	0.65	-1.61	0.01	1.60	427.6	586.1	742.8
S524QNA	Penicillin Manufacture	ACER	English	S3	46.5	(0.15)	0.42	-1.39	-0.09	1.48	422.3	560.6	712.8
S524QNB	Penicillin Manufacture	ACER	English	S3	51.5	(0.16)	0.14	-1.29	-0.17	1.46	403.5	531.3	684.9
S524QNC	Penicillin Manufacture	ACER	English	S3	47.8	(0.15)	0.33	-1.29	-0.12	1.42	421.4	551.2	699.6
S527QNA	Extinction of the Dinosaurs	ACER	English	S1	55.9	(0.15)	-0.13	-1.44	-0.18	1.62	367.9	506.2	673.7
S527QNB	Extinction of the Dinosaurs	ACER	English	S1	68.4	(0.15)	-0.73	-0.99	-0.25	1.24	345.6	451.0	589.9
S527QNC	Extinction of the Dinosaurs	ACER	English	S1	59.0	(0.16)	-0.26	-1.22	-0.07	1.30	374.5	501.3	638.8

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



[Part 1/1]

Table A1.5 2006 Main study science embedded item classification (support for scientific enquiry)

								Item parameters (RP=.50)			Thresholds (RP=.62, PISA scale)		
Item	Name	Source	Language	Cluster	International % correct	SE % correct	Difficulty	Tau(1)	Tau(2)	Tau(3)	1	2	3
S408QSA	Wild Oat Grass	ACER	English	S4	63.7	(0.11)	0.59	-1.60	-0.91	2.51	258.8	405.9	770.3
S408QSB	Wild Oat Grass	ACER	English	S4	60.5	(0.12)	0.70	-1.77	-0.29	2.06	267.8	467.9	738.0
S408QSC	Wild Oat Grass	ACER	English	S4	59.2	(0.12)	0.88	-1.65	-0.58	2.23	294.6	466.0	773.9
S416QSA	The Moon	IPN	German	S7	66.9	(0.11)	0.21	-1.75	-0.72	2.46	205.4	375.4	721.8
S416QSB	The Moon	IPN	German	S7	70.2	(0.11)	0.04	-1.55	-0.58	2.13	206.8	370.0	666.2
S416QSC	The Moon	IPN	German	S7	77.8	(0.12)	-0.19	-0.78	-0.63	1.41	247.1	353.5	564.8
S421QSA	Big and Small	ACER	English	S7.UHS	77.2	(0.11)	-0.25	-0.78	-1.03	1.80	227.3	319.8	595.2
S421QSC	Big and Small	ACER	English	S7.UHS	69.9	(0.12)	0.20	-1.29	-0.70	1.99	246.2	383.5	668.9
S425QSA	Penguin Island	IPN	German	S7	83.1	(0.10)	-0.58	-0.34	-1.15	1.49	217.5	285.7	523.4
S425QSB	Penguin Island	IPN	German	S7	72.3	(0.11)	-0.06	-1.48	-0.67	2.16	198.6	352.1	657.2
S425QSC	Penguin Island	IPN	German	S7	61.7	(0.12)	0.58	-1.84	-0.13	1.96	249.8	466.9	714.6
S426QSA	The Grand Canyon	IPN	German	S1	70.1	(0.11)	0.14	-1.19	-0.98	2.17	240.4	357.5	679.7
S426QSB	The Grand Canyon	IPN	German	S1	64.1	(0.12)	0.43	-1.72	-0.25	1.97	240.8	440.1	697.6
S426QSC	The Grand Canyon	IPN	German	S1	70.6	(0.12)	0.08	-1.27	-0.81	2.08	231.8	362.0	665.3
S438QSA	Green Parks	ACER	English	S6	76.8	(0.12)	-0.10	-0.64	-1.00	1.64	256.5	342.7	596.2
S438QSB	Green Parks	ACER	English	S6	61.1	(0.12)	0.57	-2.14	-0.11	2.25	217.5	466.0	744.2
S438QSC	Green Parks	ACER	English	S6	66.4	(0.12)	0.43	-1.49	-0.37	1.86	261.4	432.8	684.6
S456QSA	The Cheetah	IPN	German	S2	78.6	(0.11)	-0.45	-0.87	-1.08	1.95	196.0	291.1	588.1
S456QSB	The Cheetah	IPN	German	S2	65.2	(0.11)	0.38	-1.84	-0.61	2.45	218.4	403.2	741.6
S456QSC	The Cheetah	IPN	German	S2	70.4	(0.13)	0.11	-1.51	-0.23	1.74	225.6	406.9	637.0
S465QSA	Different Climates	IPN	German	S5	69.7	(0.11)	0.06	-1.66	-0.75	2.41	196.0	357.5	699.4
S465QSB	Different Climates	IPN	German	S5	77.8	(0.11)	-0.25	-0.74	-0.95	1.70	232.3	325.2	584.6
S476QSA	Heart Surgery	ACER	English	S2.UHS	76.4	(0.11)	-0.31	-1.07	-0.85	1.92	202.2	318.4	602.5
S476QSB	Heart Surgery	ACER	English	S2.UHS	87.3	(0.10)	-0.67	-0.51	-0.38	0.88	220.2	316.7	462.5
S476QSC	Heart Surgery	ACER	English	S2.UHS	89.9	(0.09)	-0.91	0.30	-1.23	0.92	219.3	262.8	428.3
S477QSA	Mary Montagu	ACER	English	S3	82.6	(0.11)	-0.44	-0.34	-1.03	1.36	239.0	310.4	527.9
S477QSB	Mary Montagu	ACER	English	S3	62.9	(0.12)	0.38	-2.09	0.08	2.01	203.1	461.5	699.4
S477QSC	Mary Montagu	ACER	English	S3	71.5	(0.11)	-0.02	-1.39	-0.73	2.12	212.1	355.3	658.1
S485QSB	Acid Rain	ACER	English	S1	68.3	(0.11)	0.14	-1.60	-0.68	2.27	212.5	374.1	694.0
S485QSC	Acid Rain	ACER	English	S1	69.5	(0.12)	0.13	-1.40	-0.82	2.22	225.6	364.2	685.1
S498QSA	Experimental Digestion	ACER	English	S3	70.7	(0.11)	0.02	-1.51	-0.73	2.25	204.9	356.6	676.9
S498QSB	Experimental Digestion	ACER	English	S3	73.5	(0.12)	-0.09	-1.25	-0.54	1.78	220.2	362.9	615.0
S519QSA	Airbags	ACER	English	S3	80.7	(0.11)	-0.41	-0.86	-0.61	1.48	214.8	327.8	546.4
S519QSB	Airbags	ACER	English	S3	78.1	(0.11)	-0.39	-1.23	-0.56	1.79	187.5	327.4	581.8
S519QSC	Airbags	ACER	English	S3	84.9	(0.12)	-0.50	-0.65	-0.31	0.96	230.1	338.2	490.3
S527QSB	Extinction of the Dinosaurs	ACER	English	S1	76.7	(0.11)	-0.26	-0.94	-0.80	1.75	220.2	331.1	590.0
S527QSC	Extinction of the Dinosaurs	ACER	English	S1	77.9	(0.12)	-0.23	-0.61	-1.02	1.63	242.6	326.9	579.2



APPENDIX 2 Contrast coding used in conditioning

[Part 1/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
Grade Q1	ST01Q01	7-14 Ungraded	(copy) (mode)	0 1
Study programme Q2	ST02Q01	National categories	If there is at least one school with more than one SP in a country, national study programmes are dummy coded with – national mode = 0 and – other categories = 1	
Age of student	AGE	Value (decimal) Missing	(copy) (mean)	0 1
Gender Q4	ST04Q01	1. Female 2. Male Missing	10 00 11	
Occupational status Mother (SEI)	BMMJ	16-90 Missing	(copy) (mean)	0 1
Occupational status Father (SEI)	BFMJ	16-90 Missing	(copy) (mean)	0 1
Occupational status Self (SEI)	BSMJ	16-90 Missing	(copy) (mean)	0 1
Educational level of mother (ISCED)	MISCED	0. None 1. ISCED 1 2. ISCED 2 3. ISCED 3B, C 4. ISCED 3A, ISCED 4 5. ISCED 5B 6. ISCED 5A, 6 Missing	7 dummy codes with national mode as reference group	
Educational level of father (ISCED)	FISCED	0. None 1. ISCED 1 2. ISCED 2 3. ISCED 3B, C 4. ISCED 3A, ISCED 4 5. ISCED 5B 6. ISCED 5A, 6 Missing	7 dummy codes with national mode as reference group	
Immigration status	IMMIG	1. Native 2. Second-Generation 3. First-Generation Missing	000 100 010 001	
Country arrival age Q11b	ST11Q04	Value Not applicable (born in country) Missing	(copy) 0 (mean)	0 0 1
Language at home Q12	ST12Q01	1. Language of test 2. Other national language 3. Other language Missing	00000 10000 01010 00101	
<Country specific wealth indicator 1>	ST13Q15	1. Yes 2. No Missing	10 00 01	
<Country specific wealth indicator 2>	ST13Q16	1. Yes 2. No Missing	10 00 01	
<Country specific wealth indicator 3>	ST13Q17	1. Yes 2. No Missing	10 00 01	
How many books at home Q15	ST15Q01	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	6 dummy codes with national mode as reference group	

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



[Part 2/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
Sci info – Photosynthesis – none Q20a	ST20QA1	1. Tick	1	Number of missing values in ST20QA
		2. No Tick	0	
Sci info – Photosynthesis – school Q20a	ST20QA2	1. Tick	1	
		2. No Tick	0	
Sci info – Photosynthesis – media Q20a	ST20QA3	1. Tick	1	
		2. No Tick	0	
Sci info – Photosynthesis – friends Q20a	ST20QA4	1. Tick	1	
		2. No Tick	0	
Sci info – Photosynthesis – family Q20a	ST20QA5	1. Tick	1	
		2. No Tick	0	
Sci info – Photosynthesis – Internet or books Q20a	ST20QA6	1. Tick	1	
		2. No Tick	0	
Sci info – Continents – none Q20b	ST20QB1	1. Tick	1	Number of missing values in ST20QB
		2. No Tick	0	
Sci info – Continents – school Q20b	ST20QB2	1. Tick	1	
		2. No Tick	0	
Sci info – Continents – media Q20b	ST20QB3	1. Tick	1	
		2. No Tick	0	
Sci info – Continents – friends Q20b	ST20QB4	1. Tick	1	
		2. No Tick	0	
Sci info – Continents – family Q20b	ST20QB5	1. Tick	1	
		2. No Tick	0	
Sci info – Continents – Internet or books Q20b	ST20QB6	1. Tick	1	
		2. No Tick	0	
Sci info – Genes – none Q20c	ST20QC1	1. Tick	1	Number of missing values in ST20QC
		2. No Tick	0	
Sci info – Genes – school Q20c	ST20QC2	1. Tick	1	
		2. No Tick	0	
Sci info – Genes – media Q20c	ST20QC3	1. Tick	1	
		2. No Tick	0	
Sci info – Genes – friends Q20c	ST20QC4	1. Tick	1	
		2. No Tick	0	
Sci info – Genes – family Q20c	ST20QC5	1. Tick	1	
		2. No Tick	0	
Sci info – Genes – Internet or books Q20c	ST20QC6	1. Tick	1	
		2. No Tick	0	
Sci info – Soundproofing – none Q20d	ST20QD1	1. Tick	1	Number of missing values in ST20QD
		2. No Tick	0	
Sci info – Soundproofing – school Q20d	ST20QD2	1. Tick	1	
		2. No Tick	0	
Sci info – Soundproofing – media Q20d	ST20QD3	1. Tick	1	
		2. No Tick	0	
Sci info – Soundproofing – friends Q20d	ST20QD4	1. Tick	1	
		2. No Tick	0	
Sci info – Soundproofing – family Q20d	ST20QD5	1. Tick	1	
		2. No Tick	0	
Sci info – Soundproofing – Internet or books Q20d	ST20QD6	1. Tick	1	
		2. No Tick	0	
Sci info – Climate change – none Q20e	ST20QE1	1. Tick	1	Number of missing values in ST20QE
		2. No Tick	0	
Sci info – Climate change – school Q20e	ST20QE2	1. Tick	1	
		2. No Tick	0	
Sci info – Climate change – media Q20e	ST20QE3	1. Tick	1	
		2. No Tick	0	
Sci info – Climate change – friends Q20e	ST20QE4	1. Tick	1	
		2. No Tick	0	
Sci info – Climate change – family Q20e	ST20QE5	1. Tick	1	
		2. No Tick	0	
Sci info – Climate change – Internet or books Q20e	ST20QE6	1. Tick	1	
		2. No Tick	0	



[Part 3/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
Sci info – Evolution – none Q20f	ST20QF1	1. Tick	1	Number of missing values in ST20QF
		2. No Tick	0	
Sci info – Evolution – school Q20f	ST20QF2	1. Tick	1	
		2. No Tick	0	
Sci info – Evolution – media Q20f	ST20QF3	1. Tick	1	
		2. No Tick	0	
Sci info – Evolution – friends Q20f	ST20QF4	1. Tick	1	
		2. No Tick	0	
Sci info – Evolution – family Q20f	ST20QF5	1. Tick	1	
		2. No Tick	0	
Sci info – Evolution – Internet or books Q20f	ST20QF6	1. Tick	1	
		2. No Tick	0	
Sci info – Nuclear energy – none Q20g	ST20QG1	1. Tick	1	Number of missing values in ST20QG
		2. No Tick	0	
Sci info – Nuclear energy – school Q20g	ST20QG2	1. Tick	1	
		2. No Tick	0	
Sci info – Nuclear energy – media Q20g	ST20QG3	1. Tick	1	
		2. No Tick	0	
Sci info – Nuclear energy – friends Q20g	ST20QG4	1. Tick	1	
		2. No Tick	0	
Sci info – Nuclear energy – family Q20g	ST20QG5	1. Tick	1	
		2. No Tick	0	
Sci info – Nuclear energy – Internet or books Q20g	ST20QG6	1. Tick	1	
		2. No Tick	0	
Sci info – Health – none Q20h	ST20QH1	1. Tick	1	Number of missing values in ST20QH
		2. No Tick	0	
Sci info – Health – school Q20h	ST20QH2	1. Tick	1	
		2. No Tick	0	
Sci info – Health – media Q20h	ST20QH3	1. Tick	1	
		2. No Tick	0	
Sci info – Health – friends Q20h	ST20QH4	1. Tick	1	
		2. No Tick	0	
Sci info – Health – family Q20h	ST20QH5	1. Tick	1	
		2. No Tick	0	
Sci info – Health – Internet or books Q20h	ST20QH6	1. Tick	1	
		2. No Tick	0	
Envr info – Air pollution – none Q23a	ST23QA1	1. Tick	1	Number of missing values in ST23QA
		2. No Tick	0	
Envr info – Air pollution – school Q23a	ST23QA2	1. Tick	1	
		2. No Tick	0	
Envr info – Air pollution – media Q23a	ST23QA3	1. Tick	1	
		2. No Tick	0	
Envr info – Air pollution – friends Q23a	ST23QA4	1. Tick	1	
		2. No Tick	0	
Envr info – Air pollution – family Q23a	ST23QA5	1. Tick	1	
		2. No Tick	0	
Envr info – Air pollution – Internet or books Q23a	ST23QA6	1. Tick	1	
		2. No Tick	0	
Envr info – Energy shortages – none Q23b	ST23QB1	1. Tick	1	Number of missing values in ST23QD
		2. No Tick	0	
Envr info – Energy shortages – school Q23b	ST23QB2	1. Tick	1	
		2. No Tick	0	
Envr info – Energy shortages – media Q23b	ST23QB3	1. Tick	1	
		2. No Tick	0	
Envr info – Energy shortages – friends Q23b	ST23QB4	1. Tick	1	
		2. No Tick	0	
Envr info – Energy shortages – family Q23b	ST23QB5	1. Tick	1	
		2. No Tick	0	
Envr info – Energy shortages – Internet or books Q23b	ST23QB6	1. Tick	1	
		2. No Tick	0	

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



[Part 4/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
Envr info – Water shortages – none Q23e	ST23QE1	1. Tick	1	Number of missing values in ST23QE
		2. No Tick	0	
Envr info – Water shortages – school Q23e	ST23QE2	1. Tick	1	
		2. No Tick	0	
Envr info – Water shortages – media Q23e	ST23QE3	1. Tick	1	
		2. No Tick	0	
Envr info – Water shortages – friends Q23e	ST23QE4	1. Tick	1	
		2. No Tick	0	
Envr info – Water shortages – family Q23e	ST23QE5	1. Tick	1	
		2. No Tick	0	
Envr info – Water shortages – Internet or books Q23e	ST23QE6	1. Tick	1	
		2. No Tick	0	
Envr info – Nuclear waste – none Q23f	ST23QF1	1. Tick	1	Number of missing values in ST23QF
		2. No Tick	0	
Envr info – Nuclear waste – school Q23f	ST23QF2	1. Tick	1	
		2. No Tick	0	
Envr info – Nuclear waste – media Q23f	ST23QF3	1. Tick	1	
		2. No Tick	0	
Envr info – Nuclear waste – friends Q23f	ST23QF4	1. Tick	1	
		2. No Tick	0	
Envr info – Nuclear waste – family Q23f	ST23QF5	1. Tick	1	
		2. No Tick	0	
Envr info – Nuclear waste – Internet or books Q23f	ST23QF6	1. Tick	1	
		2. No Tick	0	
Regular lessons – Science Q31a	ST31Q01	1. No time	0	0
		2. Less than 2 hours	1	0
		3. Up to 4 hours	3	0
		4. Up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Out of school – Science Q31b	ST31Q02	1. No time	0	0
		2. Less than 2 hours	1	0
		3. Up to 4 hours	3	0
		4. Up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Self study – Science Q31c	ST31Q03	1. No time	0	0
		2. Less than 2 hours	1	0
		3. Up to 4 hours	3	0
		4. Up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Regular lessons – Mathematics Q31d	ST31Q04	1. No time	0	0
		2. Less than 2 hours	1	0
		3. Up to 4 hours	3	0
		4. Up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Out of school – Mathematics Q31e	ST31Q05	1. No time	0	0
		2. Less than 2 hours	1	0
		3. Up to 4 hours	3	0
		4. Up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Self study – Mathematics Q31f	ST31Q06	1. No time	0	0
		2. Less than 2 hours	1	0
		3. Up to 4 hours	3	0
		4. Up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1

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



[Part 5/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
Regular lessons – Language Q31g	ST31Q07	1. No time	0	0
		2. Less than 2 hours	1	0
		3. 2 up to 4 hours	3	0
		4. 4 up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Out of school – Language Q31h	ST31Q08	1. No time	0	0
		2. Less than 2 hours	1	0
		3. 2 up to 4 hours	3	0
		4. 4 up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Self study – Language Q31i	ST31Q09	1. No time	0	0
		2. Less than 2 hours	1	0
		3. 2 up to 4 hours	3	0
		4. 4 up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Regular lessons – Other Q31j	ST31Q10	1. No time	0	0
		2. Less than 2 hours	1	0
		3. 2 up to 4 hours	3	0
		4. 4 up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Out of school – Other Q31k	ST31Q11	1. No time	0	0
		2. Less than 2 hours	1	0
		3. 2 up to 4 hours	3	0
		4. 4 up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Self study – Other Q31l	ST31Q12	1. No time	0	0
		2. Less than 2 hours	1	0
		3. 2 up to 4 hours	3	0
		4. 4 up to 6 hours	5	0
		5. 6 or more hours	7	0
		Missing	(mean)	1
Course – Comp Sci last year Q33a	ST33Q11	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Comp Sci this year Q33a	ST33Q12	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Sci last year Q33b	ST33Q21	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Sci this year Q33b	ST33Q22	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Comp Bio last year Q33c	ST33Q31	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Comp Bio this year Q33c	ST33Q32	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Bio last year Q33d	ST33Q41	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Bio this year Q33d	ST33Q42	1. Yes	1	0
		2. No	0	0
		Missing	0	1

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



[Part 6/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
Course – Comp Phy last year Q33e	ST33Q51	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Comp Phy this year Q33e	ST33Q52	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Phy last year Q33f	ST33Q61	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Phy this year Q33f	ST33Q62	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Comp Chem last year Q33g	ST33Q71	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Comp Chem this year Q33g	ST33Q72	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Chem last year Q33h	ST33Q81	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Course – Opt Chem this year Q33h	ST33Q82	1. Yes	1	0
		2. No	0	0
		Missing	0	1
Self – Do well Science Q36a	ST36Q01	1. Very important	3	0
		2. Important	2	0
		3. Of little importance	1	0
		4. Not important at all	0	0
		Missing	0	1
Self – Do well Maths Q36b	ST36Q02	1. Very important	3	0
		2. Important	2	0
		3. Of little importance	1	0
		4. Not important at all	0	0
		Missing	0	1
Self – Do well Language Q36c	ST36Q03	1. Very important	3	0
		2. Important	2	0
		3. Of little importance	1	0
		4. Not important at all	0	0
		Missing	0	1
Student information on science-related careers PISA 2006 (WLE)	CARINFO	Value (decimal)	Z-score (national)	0
		Missing	0	1
School preparation for science-related careers PISA 2006 (WLE)	CARPREP	Value (decimal)	Z-score (national)	0
		Missing	0	1
Cultural possessions at home PISA 2006 (WLE)	CULTPOSS	Value (decimal)	Z-score (national)	0
		Missing	0	1
Awareness of environmental issues PISA 2006 (WLE)	ENVAWARE	Value (decimal)	Z-score (national)	0
		Missing	0	1
Environmental optimism PISA 2006 (WLE)	ENVOPT	Value (decimal)	Z-score (national)	0
		Missing	0	1
Perception of environmental issues PISA 2006 (WLE)	ENVPERC	Value (decimal)	Z-score (national)	0
		Missing	0	1
General value of science PISA 2006 (WLE)	GENSCIE	Value (decimal)	Z-score (national)	0
		Missing	0	1
Home educational resources PISA 2006 (WLE)	HEDRES	Value (decimal)	Z-score (national)	0
		Missing	0	1
Index of home possessions PISA 2006 (WLE)	HOMEPOS	Value (decimal)	Z-score (national)	0
		Missing	0	1
Instrumental motivation in science PISA 2006 (WLE)	INSTSCIE	Value (decimal)	Z-score (national)	0
		Missing	0	1



[Part 7/7]

Table A2.1 2006 Main study contrast coding used in conditioning for the student questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
STUDENT QUESTIONNAIRE				
General interest in learning science PISA 2006 (WLE)	INTSCIE	Value (decimal)	Z-score (national)	0
		Missing	0	1
Enjoyment of science PISA 2006 (WLE)	JOYSCIE	Value (decimal)	Z-score (national)	0
		Missing	0	1
Personal value of science PISA 2006 (WLE)	PERSCIE	Value (decimal)	Z-score (national)	0
		Missing	0	1
Responsibility for sustainable development PISA 2006 (WLE)	RESPDEV	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science Teaching – Focus on applications or models PISA 2006 (WLE)	SCAPPLY	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science Teaching – Hands-on activities PISA 2006 (WLE)	SCHANDS	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science activities PISA 2006 (WLE)	SCIEACT	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science self-efficacy PISA 2006 (WLE)	SCIEEFF	Value (decimal)	Z-score (national)	0
		Missing	0	1
Future-oriented science motivation PISA 2006 (WLE)	SCIEFUT	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science Teaching – Interaction PISA 2006 (WLE)	SCINTACT	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science Teaching – Student investigations PISA 2006 (WLE)	SCINVEST	Value (decimal)	Z-score (national)	0
		Missing	0	1
Science self-concept PISA 2006 (WLE)	SCSCIE	Value (decimal)	Z-score (national)	0
		Missing	0	1
Family wealth PISA 2006 (WLE)	WEALTH	Value (decimal)	Z-score (national)	0
		Missing	0	1
Effort B – Effort A	DEFFORT	<0	0	1
		>=0	(copy)	0
		Missing	0	0

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



[Part 1/1]

Table A2.2 2006 Main study contrast coding used in conditioning for the ICT questionnaire variables

Variable	Var. name	Variable coding	Contrast coding	
ICT QUESTIONNAIRE				
How long used computers IC2	IC02Q01	1. Less than 1 year	1	0
		2. 1 to 3 years	2	0
		3. 3 to 5 years	3	0
		4. 5 years or more	4	0
		Missing	(mean)	1
Use computer at home IC3a	IC03Q01	1. Almost every day	4	0
		2. Once or twice a week	3	0
		3. Few times a month	2	0
		4. Once a month or less	1	0
		5. Never	0	0
		Missing	(mean)	1
Use computer at school IC3b	IC03Q02	1. Almost every day	4	0
		2. Once or twice a week	3	0
		3. Few times a month	2	0
		4. Once a month or less	1	0
		5. Never	0	0
		Missing	(mean)	1
Use computer other places IC3c	IC03Q03	1. Almost every day	4	0
		2. Once or twice a week	3	0
		3. Few times a month	2	0
		4. Once a month or less	1	0
		5. Never	0	0
		Missing	(mean)	1
How well – Copy data to CD IC5e	IC05Q05	1. Do well by myself	3	0
		2. Do with help	2	0
		3. Know but can't do	1	0
		4. Don't know	0	0
		Missing	(mean)	1
How well – Move files IC5f	IC05Q06	1. Do well by myself	3	0
		2. Do with help	2	0
		3. Know but can't do	1	0
		4. Don't know	0	0
		Missing	(mean)	1
Self-confidence in ICT high level tasks PISA 2006 (WLE)	HIGHCONF	Value (decimal)	Z-score (national)	0
		Missing	0	1
Self-confidence in ICT Internet tasks PISA 2006 (WLE)	INTCONF	Value (decimal)	Z-score (national)	0
		Missing	0	1
ICT Internet/entertainment use PISA 2006 (WLE)	INTUSE	Value (decimal)	Z-score (national)	0
		Missing	0	1
ICT program/software use PISA 2006 (WLE)	PRGUSE	Value (decimal)	Z-score (national)	0
		Missing	0	1

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



[Part 1/2]

Table A2.3 2006 Main study contrast coding used in conditioning for the parent questionnaire variables and other variables

Variable	Var. name	Variable coding	Contrast coding	
PARENT QUESTIONNAIRE				
Completed Quest – Mother Q1a	PA01Q01	1. Yes	000	
Completed Quest – Father Q1b	PA01Q02	1. Yes	101	
Completed Quest – Other Q1c	PA01Q03	1. Yes	011	
		Missing	001	
Education cost Q9	PA09Q01	1. Less than A	0	0
		2. A or more – less than B	1	0
		3. B or more – less than C	2	0
		4. C or more – less than D	3	0
		5. D leva or more	4	0
		Missing	(median)	1
Father age Q10a	PA10Q01	1. Younger than 36	0	0
		2. 36 to 40 years	1	0
		3. 41 to 45 years	2	0
		4. 46 to 50 years	3	0
		5. 51 years or older	4	0
		Missing	(median)	1
Mother age Q10b	PA10Q02	1. Younger than 36	0	0
		2. 36 to 40 years	1	0
		3. 41 to 45 years	2	0
		4. 46 to 50 years	3	0
		5. 51 years or older	4	0
		Missing	(median)	1
PQ Occupational status Father (SEI)	PQBFMJ	16-90	(copy)	0
		Missing	(mean)	1
PQ Educational level of father (ISCED)	PQFISCED	0. Below ISCED 3A	00000	
		1. ISCED 3A	10000	
		2. ISCED 4	01000	
		3. ISCED 5B	00100	
		4. ISCED 5A or 6	00010	
		Missing	00001	
PQ Occupational status Mother (SEI)	PQBMMJ	16-90	(copy)	0
		Missing	(mean)	1
PQ Educational level of mother (ISCED)	PQMISCED	0. Below ISCED 3A	00000	
		1. ISCED 3A	10000	
		2. ISCED 4	01000	
		3. ISCED 5B	00100	
		4. ISCED 5A or 6	00010	
		Missing	00001	
Household income (relative to median) Q15	PA15Q01	1. Less than < 0.5 median >	0	0
		2. < 0.5 median > or more but less than < 0.75 median >	1	0
		3. < 0.75 median > or more but less than < median >	2	0
		4. < median > or more but less than < 1.25 median >	3	0
		5. < 1.25 median > or more but less than < 1.5 median >	4	0
		6. < 1.5 median > or more	5	0
		Missing	(median)	1

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



[Part 2/2]

Table A2.3 2006 Main study contrast coding used in conditioning for the parent questionnaire variables and other variables

Variable	Var. name	Variable coding	Contrast coding	
PARENT QUESTIONNAIRE				
PQ Perception of environmental issues PISA 2006 (WLE)	PQENPERC	Value (decimal)	(copy)	0
		Missing	(mean)	1
PQ Environmental optimism PISA 2006 (WLE)	PQENVOPT	Value (decimal)	(copy)	0
		Missing	(mean)	1
PQ General value of science PISA 2006 (WLE)	PQGENSCI	Value (decimal)	(copy)	0
		Missing	(mean)	1
PQ Personal value of science PISA 2006 (WLE)	PQPERSCI	Value (decimal)	(copy)	0
		Missing	(mean)	1
Parents reports on science career motivation PISA 2006 (WLE)	PQSCCAR	Value (decimal)	(copy)	0
		Missing	(mean)	1
Parents perception of school quality PISA 2006 (WLE)	PQSCHOOL	Value (decimal)	(copy)	0
		Missing	(mean)	1
Science activities at age 10 PISA 2006 (WLE)	PQSCIACT	Value (decimal)	(copy)	0
		Missing	(mean)	1
Parents view – importance of science PISA 2006 (WLE)	PQSCIMP	Value (decimal)	(copy)	0
		Missing	(mean)	1
OTHER VARIABLES				
Booklet number	BOOKID	1	01 00 00 00 00 00 00 00 00 00 00 00	
		2	00 01 00 00 00 00 00 00 00 00 00 00	
		3	00 00 01 00 00 00 00 00 00 00 00 00	
		4	00 00 00 01 00 00 00 00 00 00 00 00	
		5	00 00 00 00 01 00 00 00 00 00 00 00	
		6	00 00 00 00 00 01 00 00 00 00 00 00	
		7	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	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1	
		12	00 00 00 00 00 00 00 00 00 00 01 00	
		13	00 00 00 00 00 00 00 00 00 00 00 01	
School identification number	SCHOOLID	Unique 5-digit school ID	Total number of schools minus one dummies are created for school membership. A school with the highest number of students is made the reference school in a country (a string of zeros).	



APPENDIX 3 Design effect tables

[Part 1/1]

Table A3.1 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
Albania	3.29	3.08	2.89						
Argentina	9.86	9.38	8.56				7.17	6.24	6.08
Australia	3.52	3.49	3.47	2.13	2.15	2.10	2.06	2.24	2.26
Austria	2.40	2.51	2.55	3.76	3.27	3.44	4.08	3.74	3.92
Azerbaijan							3.12	2.26	2.75
Belgium	3.56	3.90	4.29	2.58	2.29	2.48	3.04	2.95	2.48
Brazil	3.10	3.71	3.26	4.58	4.83	4.35	3.74	2.93	2.79
Bulgaria	4.89	5.66	4.64				6.91	6.13	6.11
Canada	1.56	1.40	1.57	1.75	1.82	2.02	2.44	1.97	2.03
Chile	3.59	3.71	3.48				4.99	4.58	4.32
Colombia							5.08	3.78	3.37
Croatia							2.81	2.37	2.45
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
Hong Kong-China	2.93	3.26	3.01	3.69	4.54	4.26	2.42	2.67	2.47
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
Indonesia	3.99	4.54	3.94	3.38	3.91	3.21	5.92	5.63	5.73
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
Jordan							3.27	3.30	2.84
Korea	2.42	2.76	2.69	3.09	3.24	3.54	3.81	3.76	3.36
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
Luxembourg	1.59	1.99	2.32	1.48	0.97	1.50	1.28	1.07	1.05
Macao-China				2.16	2.89	3.03	1.10	1.30	1.06
Macedonia	1.93	2.72	2.10						
Mexico	3.31	3.36	3.18	4.09	3.64	3.49	3.06	2.93	2.71
Montenegro							1.22	1.37	1.06
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
Peru	4.42	4.33	4.13						
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
Qatar							1.20	1.02	0.86
Romania							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
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
Chinese Taipei							3.38	4.10	3.57
Thailand	3.24	3.60	3.06	2.81	3.00	2.70	2.59	2.34	2.14
Tunisia				2.81	2.54	2.56	4.02	3.96	2.96
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
Uruguay				3.43	3.29	2.90	3.43	2.61	2.75
Central tendency indices on the 35 countries that participated in the three 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**

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



[Part 1/1]

Table A3.2 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
Albania	174	4980	28.6						
Argentina	156	3983	25.5				176	4339	24.7
Australia	231	5176	22.4	321	12551	39.1	356	14170	39.8
Austria	213	4745	22.3	193	4597	23.8	199	4927	24.8
Azerbaijan							171	5184	30.3
Belgium	216	6670	30.9	277	8796	31.8	269	8857	32.9
Brazil	324	4893	15.1	228	4452	19.5	625	9295	14.9
Bulgaria	160	4657	29.1				180	4498	25.0
Canada	1117	29687	26.6	1087	27953	25.7	896	22646	25.3
Chile	179	4889	27.3				173	5233	30.2
Colombia							165	4478	27.1
Croatia							161	5213	32.4
Czech Republic	229	5365	23.4	260	6320	24.3	245	5932	24.2
Denmark	225	4235	18.8	206	4218	20.5	211	4532	21.5
Estonia							169	4865	28.8
Finland	155	4864	31.4	197	5796	29.4	155	4714	30.4
France	177	4673	26.4	170	4300	25.3	182	4716	25.9
Germany	219	5073	23.2	216	4660	21.6	226	4891	21.6
Greece	157	4672	29.8	171	4627	27.1	190	4873	25.6
Hong Kong-China	140	4405	31.5	145	4478	30.9	146	4645	31.8
Hungary	194	4887	25.2	253	4765	18.8	189	4490	23.8
Iceland	130	3372	25.9	129	3350	26.0	139	3789	27.3
Indonesia	290	7368	25.4	346	10761	31.1	352	10647	30.2
Ireland	139	3854	27.7	145	3880	26.8	165	4585	27.8
Israel	165	4498	27.3				149	4584	30.8
Italy	172	4984	29.0	406	11639	28.7	799	21773	27.3
Japan	135	5256	38.9	144	4707	32.7	185	5952	32.2
Jordan							210	6509	31.0
Korea	146	4982	34.1	149	5444	36.5	154	5176	33.6
Kyrgyzstan							201	5904	29.4
Latvia	154	3893	25.3	157	4627	29.5	176	4719	26.8
Liechtenstein	11	314	28.5	12	332	27.7	12	339	28.3
Lithuania							197	4744	24.1
Luxembourg	24	3528	147.0	29	3923	135.3	31	4567	147.3
Macao-China				39	1250	32.1	43	4760	110.7
Macedonia	91	4510	49.6						
Mexico	183	4600	25.1	1124	29983	26.7	1140	30971	27.2
Montenegro							51	4455	87.4
Netherlands	100	2503	25.0	154	3992	25.9	185	4871	26.3
New Zealand	153	3667	24.0	173	4511	26.1	170	4823	28.4
Norway	176	4147	23.6	182	4064	22.3	203	4692	23.1
Peru	177	4429	25.0						
Poland	127	3654	28.8	166	4383	26.4	221	5547	25.1
Portugal	149	4585	30.8	153	4608	30.1	173	5109	29.5
Qatar							131	6265	47.8
Romania							174	5118	29.4
Russian Federation	246	6701	27.2	212	5974	28.2	209	5799	27.7
Serbia				149	4405	29.6	162	4798	29.6
Slovak Republic				281	7346	26.1	189	4731	25.0
Slovenia							361	6595	18.3
Spain	185	6214	33.6	383	10791	28.2	686	19604	28.6
Sweden	154	4416	28.7	185	4624	25.0	197	4443	22.6
Switzerland	282	6100	21.6	445	8420	18.9	510	12192	23.9
Chinese Taipei							236	8815	37.4
Thailand	179	5340	29.8	179	5236	29.3	212	6192	29.2
Tunisia				149	4721	31.7	152	4640	30.5
Turkey				159	4855	30.5	160	4942	30.9
United Kingdom	362	9340	25.8	339	9535	28.1	502	13152	26.2
United States	153	3846	25.1	274	5456	19.9	166	5611	33.8
Uruguay				243	5835	24.0	278	4839	17.4
Central tendency indices on the 35 countries that participated in the three surveys									
Median			**26.4**			**26.8**			
Mean			**30.2**			**29.9**			

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



[Part 1/1]

Table A3.3 School variance estimate by country, by domain and cycle

	PISA 2000			PISA 2003			PISA 2006		
	Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
Albania	4046	3355	2521						
Argentina	5920	6282	4897				6881	5072	4794
Australia	1888	1405	1500	2009	1927	2079	1878	1694	1839
Austria	6417	5173	5241	7566	5250	5823	6861	5785	5464
Azerbaijan							2359	1655	1612
Belgium	7025	6291	6939	7186	7240	5983	6593	5814	5182
Brazil	3379	3548	2453	3416	4159	3182	4555	4342	3711
Bulgaria	6162	5764	3776				7870	5199	6226
Canada	1588	1255	1279	1199	1270	1492	2163	1547	1668
Chile	4968	4268	3813				6011	4800	4740
Colombia							3466	2973	2244
Croatia							3794	2721	3036
Czech Republic	4814	4055	3612	4507	4942	4388	7325	6451	5617
Denmark	1876	1363	1760	1437	1147	1308	1593	1281	1393
Estonia							2217	1594	1437
Finland	1009	410	448	257	343	361	643	489	433
France	4243	3704	5006	4245	3830	5803	6090	5049	5488
Germany	6903	5653	5191	7001	6101	7036	9733	6183	5944
Greece	5060	5576	3786	3976	3357	2723	5493	3877	4369
Hong Kong-China	3318	3955	3198	2949	4573	3915	2605	3420	3072
Hungary	6408	5236	5731	4919	5710	5424	7164	6181	5453
Iceland	696	430	572	382	319	365	1220	725	898
Indonesia	2019	2253	1704	1991	2720	1605	2422	2746	1745
Ireland	1566	816	1242	1712	1218	1408	2010	1310	1539
Israel	5109	5673	4953				5641	4668	3926
Italy	4844	3578	4188	5009	4915	5701	6210	4951	4758
Japan	3377	3727	3646	4998	5400	5543	5459	4474	4867
Jordan							2629	1660	1792
Korea	1840	2889	2574	2475	3607	3870	3205	3494	2869
Kyrgyzstan							4334	3159	2763
Latvia	3305	2836	2775	1666	1761	1778	2183	1537	1316
Liechtenstein	3456	3395	3171	2998	3461	3510	3452	2921	3176
Lithuania							2671	2687	2308
Luxembourg	3069	2056	2474	2656	2673	3018	2817	2777	2738
Macao-China				1105	1455	1356	1708	1733	1739
Macedonia	3994	3025	2350						
Mexico	3969	3467	2429	2818	2496	1934	3296	2580	2293
Montenegro							2715	1752	1812
Netherlands	3984	3873	4262	4316	5508	5743	5567	4880	5359
New Zealand	1892	1702	1732	1916	1781	1922	2108	1406	1930
Norway	1111	726	845	819	578	846	1385	942	964
Peru	5992	4786	3179						
Poland	6127	5483	4684	1351	1035	1489	1580	1121	1108
Portugal	3457	2492	2427	3315	2620	2733	3449	2746	2502
Qatar							7141	5015	4240
Romania							4658	3614	3182
Russian Federation	3079	3896	3034	2034	2558	2086	3121	2325	2166
Serbia				2305	2566	1978	3941	3723	3086
Slovak Republic				3538	3794	4560	5567	4541	3690
Slovenia							6634	4674	5811
Spain	1473	1445	1595	1700	1489	1677	1271	1240	1151
Sweden	793	691	679	873	970	1046	1694	1215	1091
Switzerland	4421	3970	4024	2608	3165	3314	3101	3283	3375
Chinese Taipei							3194	5020	4120
Thailand	1848	2324	1789	2120	2602	2176	2863	2480	2294
Tunisia				3024	2807	2549	4636	4003	2904
Turkey				4772	5915	4732	4047	4557	3653
United Kingdom	2114	1865	2195	1815	1829	2048	2234	1726	2200
United States	3236	3127	3637	2481	2345	2270		2201	2626
Uruguay				5553	4618	4108	6018	3926	3525
Central tendency indices on the 35 countries that participated in the three surveys									
Median	**3305**	**3127**	**2574**	**2481**	**2620**	**2270**	**2982**	**2746**	**2502**
Mean	**3303**	**2990**	**2909**	**2935**	**2997**	**3017**	**3628**	**3006**	**2931**

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



[Part 1/1]

Table A3.4 Intraclass correlation by country, by domain and cycle

	PISA 2000			PISA 2003			PISA 2006		
	Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
Albania	0.41	0.29	0.28						
Argentina	0.51	0.43	0.41				0.45	0.51	0.48
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
Azerbaijan							0.46	0.57	0.50
Belgium	0.60	0.55	0.55	0.56	0.56	0.50	0.54	0.53	0.52
Brazil	0.44	0.36	0.30	0.28	0.45	0.34	0.46	0.53	0.47
Bulgaria	0.56	0.47	0.40				0.56	0.51	0.54
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
Colombia							0.30	0.37	0.30
Croatia							0.47	0.38	0.40
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
Hong Kong-China	0.47	0.45	0.45	0.42	0.47	0.45	0.39	0.40	0.37
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
Indonesia	0.43	0.34	0.33	0.36	0.44	0.37	0.50	0.50	0.43
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
Jordan							0.31	0.25	0.23
Korea	0.37	0.40	0.39	0.36	0.42	0.38	0.40	0.40	0.35
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
Luxembourg	0.31	0.24	0.27	0.27	0.31	0.28	0.29	0.32	0.30
Macao-China				0.23	0.19	0.17	0.27	0.23	0.26
Macedonia	0.45	0.31	0.34						
Mexico	0.53	0.50	0.41	0.36	0.39	0.28	0.41	0.42	0.40
Montenegro							0.33	0.25	0.28
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
Peru	0.58	0.39	0.36						
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
Qatar							0.54	0.53	0.53
Romania							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
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
Chinese Taipei							0.46	0.49	0.47
Thailand	0.31	0.33	0.30	0.34	0.37	0.32	0.42	0.36	0.37
Tunisia				0.33	0.42	0.33	0.47	0.48	0.42
Turkey				0.53	0.55	0.53	0.48	0.53	0.53
United Kingdom	0.22	0.23	0.24	0.21	0.22	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
Uruguay				0.36	0.44	0.33	0.41	0.40	0.40
Central tendency indices on the 35 countries that participated in the three 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**



[Part 1/1]

Table A3.5 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
Albania	0.26	0.19	0.19						
Argentina	0.40	0.33	0.31				0.37	0.43	0.40
Australia	0.13	0.12	0.12	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
Azerbaijan							0.37	0.53	0.42
Belgium	0.42	0.39	0.38	0.31	0.29	0.26	0.33	0.32	0.32
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
Canada	0.13	0.14	0.13	0.12	0.13	0.12	0.20	0.18	0.16
Chile	0.33	0.26	0.22				0.33	0.38	0.32
Colombia							0.29	0.36	0.30
Croatia							0.22	0.17	0.17
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
Hong Kong-China	0.47	0.44	0.44	0.42	0.46	0.45	0.38	0.39	0.36
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
Indonesia	0.38	0.28	0.29	0.33	0.40	0.33	0.46	0.44	0.38
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
Jordan							0.26	0.21	0.19
Korea	0.17	0.13	0.13	0.18	0.21	0.20	0.27	0.25	0.20
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
Luxembourg	0.31	0.22	0.27	0.25	0.28	0.25	0.13	0.15	0.14
Macao-China				0.22	0.16	0.16	0.19	0.17	0.19
Macedonia	0.31	0.19	0.19						
Mexico	0.48	0.44	0.36	0.30	0.33	0.23	0.29	0.31	0.29
Montenegro							0.27	0.23	0.24
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
Peru	0.49	0.30	0.28						
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
Qatar							0.20	0.20	0.20
Romania							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
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
Chinese Taipei							0.37	0.40	0.38
Thailand	0.26	0.29	0.24	0.28	0.32	0.26	0.31	0.29	0.27
Tunisia				0.33	0.42	0.33	0.18	0.19	0.13
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
Uruguay				0.22	0.31	0.22	0.25	0.22	0.21
Central tendency indices on the 35 countries that participated in the three surveys									
Median	**0.21**	**0.19**	**0.18**	**0.18**	**0.20**	**0.14**	**0.27**	**0.22**	**0.20**
Mean	**0.23**	**0.20**	**0.19**	**0.18**	**0.19**	**0.16**	**0.26**	**0.25**	**0.22**

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



[Part 1/1]

Table A3.6 Percentages of school variance explained by explicit stratification variables, by domain and cycle

	PISA 2000			PISA 2003			PISA 2006		
	Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
Albania	0.48	0.43	0.41						
Argentina	0.34	0.36	0.35				0.29	0.26	0.27
Australia	0.33	0.35	0.35	0.36	0.35	0.37	0.43	0.38	0.43
Austria	0.90	0.84	0.86	0.55	0.60	0.60	0.65	0.69	0.62
Azerbaijan							0.30	0.16	0.27
Belgium	0.51	0.48	0.50	0.65	0.68	0.64	0.58	0.57	0.56
Brazil	0.04	0.03	0.03	0.48	0.48	0.51	0.16	0.21	0.19
Bulgaria	0.32	0.32	0.33				0.27	0.26	0.31
Canada	0.28	0.25	0.24	0.26	0.24	0.23	0.18	0.18	0.20
Chile	0.60	0.58	0.59				0.50	0.52	0.54
Colombia							0.04	0.04	0.03
Croatia							0.69	0.67	0.70
Czech Republic	0.88	0.78	0.85	0.50	0.53	0.55	0.68	0.72	0.71
Denmark	0.07	0.12	0.03	0.07	0.07	0.05	0.02	0.01	0.02
Estonia							0.39	0.32	0.39
Finland	0.11	0.12	0.29	0.18	0.11	0.19	0.15	0.11	0.11
France	0.77	0.76	0.79	0.76	0.77	0.77	0.67	0.73	0.71
Germany				0.29	0.26	0.28	0.41	0.24	0.27
Greece	0.28	0.36	0.27	0.07	0.08	0.09	0.33	0.42	0.43
Hong Kong-China	0.03	0.03	0.03	0.01	0.01	0.01	0.04	0.04	0.04
Hungary	0.29	0.18	0.24	0.78	0.75	0.80	0.65	0.65	0.68
Iceland	0.14	0.13	0.11	0.23	0.28	0.20	0.11	0.13	0.17
Indonesia	0.19	0.24	0.19	0.15	0.18	0.16	0.17	0.21	0.19
Ireland	0.06	0.05	0.04	0.13	0.11	0.12	0.12	0.14	0.18
Israel	0.24	0.24	0.18				0.28	0.35	0.26
Italy	0.58	0.52	0.50	0.75	0.74	0.76	0.77	0.76	0.80
Japan	0.10	0.09	0.11	0.07	0.08	0.08	0.16	0.13	0.14
Jordan							0.18	0.19	0.21
Korea	0.66	0.77	0.76	0.59	0.62	0.60	0.44	0.51	0.52
Kyrgyzstan							0.58	0.55	0.56
Latvia	0.23	0.19	0.16	0.13	0.14	0.11	0.13	0.20	0.17
Liechtenstein									
Lithuania							0.51	0.52	0.52
Luxembourg	0.01	0.10	0.02	0.08	0.17	0.15	0.62	0.62	0.61
Macao-China				0.08	0.14	0.10	0.36	0.29	0.35
Macedonia	0.45	0.48	0.54						
Mexico	0.20	0.20	0.18	0.24	0.23	0.24	0.41	0.39	0.39
Montenegro							0.24	0.10	0.22
Netherlands	0.79	0.78	0.79	0.72	0.74	0.79	0.65	0.75	0.77
New Zealand	0.06	0.07	0.06	0.03	0.07	0.05	0.01	0.03	0.02
Norway	0.09	0.09	0.07	0.03	0.03	0.01	0.06	0.08	0.07
Peru	0.30	0.34	0.31						
Poland	0.80	0.75	0.75	0.06	0.06	0.07	0.13	0.18	0.17
Portugal	0.08	0.08	0.08	0.16	0.17	0.16	0.57	0.62	0.64
Qatar							0.79	0.77	0.77
Romania							0.53	0.49	0.53
Russian Federation	0.23	0.26	0.25	0.42	0.44	0.46	0.34	0.35	0.36
Serbia				0.08	0.07	0.08	0.19	0.22	0.22
Slovak Republic				0.19	0.21	0.20	0.41	0.40	0.49
Slovenia							0.79	0.77	0.80
Spain	0.44	0.44	0.47	0.45	0.43	0.41	0.42	0.48	0.44
Sweden	0.27	0.24	0.31	0.19	0.18	0.17	0.24	0.19	0.17
Switzerland	0.30	0.30	0.28	0.22	0.23	0.21	0.34	0.35	0.33
Chinese Taipei							0.31	0.32	0.30
Thailand	0.22	0.18	0.25	0.25	0.20	0.27	0.37	0.28	0.36
Tunisia				0.02	0.02	0.01	0.75	0.74	0.80
Turkey				0.50	0.44	0.43	0.24	0.17	0.17
United Kingdom	0.04	0.04	0.04	0.05	0.06	0.04	0.07	0.08	0.08
United States				0.11	0.10	0.12			
Uruguay				0.48	0.43	0.44	0.53	0.58	0.60
Central tendency indices on the 35 countries that participated in the three surveys									
Median	**0.23**	**0.22**	**0.24**	**0.23**	**0.22**	**0.21**	**0.34**	**0.28**	**0.34**
Mean	**0.31**	**0.31**	**0.31**	**0.29**	**0.30**	**0.30**	**0.34**	**0.34**	**0.34**

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



APPENDIX 4 Changes to core questionnaire items from 2003 to 2006

[Part 1/2]

Table A4.1 Student questionnaire

PISA 2006 Question number	PISA 2006 Variable name	PISA 2003 Question number	PISA 2003 Variable name	PISA 2006 English version	Summary of changes from PISA 2003
Q1	ST01Q01	Q1a	ST01Q01	What <grade> are you in?	*Unchanged*
Q2	ST02Q01	Q1b	ST01Q02	Which one of the following <programmes> are you in?	*Unchanged*
Q3		Q2		On what date were you born?	
	ST03Q02	b	ST02Q02	Month	*Unchanged*
	ST03Q03	c	ST02Q03	Year	*Unchanged*
Q4	ST04Q01	Q3	ST03Q01	Are you female or male?	*Unchanged*
				Female	
				Male	
Q5a	ST05Q01	Q7	ST07Q01	What is your mother's main job? (*e.g.* school teacher, kitchen-hand, sales manager)	*Identical except for the example jobs: "(e.g. <school teacher, nurse, sales manager>)"*
Q5b		Q8	ST08Q01	What does your mother do in her main job? (*e.g.* teaches high school students, helps the cook prepare meals in a restaurant, manages a sales team)	*Identical except for the example jobs: "(e.g. <teaches high school students, cares for patients, manages a sales team>)"*
Q6	ST06Q01	Q11		What is the <highest level of schooling> completed by your mother?	**In PISA 2003: "Which of the following did your mother complete at <school>?"**
				(Please tick only one box)	*(Please <tick> as many boxes as apply.)*
		a	ST11Q01	<ISCED level 3A>	
		b	ST11Q02	<ISCED level 3B. 3C>	
		c	ST11Q03	<ISCED level 2>	
		d	ST11Q04	<ISCED level 1>	
		e	ST11Q05	She did not complete <ISCED level 1>	In PISA 2003: "None of the above"
Q7		Q12		Does your mother have any of the following qualifications?	*Wording is identical but response format changed. In PISA 2006 there were two boxes: "Yes" or "No". In PISA 2003 there was only one box for each item.*
a	ST07Q01	a	ST12Q01	<ISCED level 5A. 6>	
b	ST07Q02	b	ST12Q02	<ISCED level 5B>	
c	ST07Q03	c	ST12Q03	<ISCED level 4>	
Q8a	ST08Q01	Q9	ST09Q01	What is your father's main job? (*e.g.* school teacher, kitchen-hand, sales manager)	*Identical except for the example jobs: "(e.g. <school teacher. carpenter. sales manager>)"*
Q8b		Q10	ST10Q01	What does your father do in his main job? (*e.g.* teaches high school students, helps the cook prepare meals in a restaurant, manages a sales team)	*Identical except for the example jobs: "(e.g. <teaches high school students, builds houses, manages a sales team>)"*
Q9	ST09Q01	Q13		What is the <highest level of schooling> completed by your father?	**In PISA 2003: "Which of the following did your father complete at <school>?"**
				(Please tick only one box)	*(Please <tick> as many boxes as apply.)*
		a	ST13Q01	<ISCED level 3A>	
		b	ST13Q02	<ISCED level 3B, 3C>	
		c	ST13Q03	<ISCED level 2>	
		d	ST13Q04	<ISCED level 1>	
		e	ST13Q05	He did not complete <ISCED level 1>	In PISA 2003: "None of the above"
Q10		Q14		Does your father have any of the following qualifications?	*Wording is identical but response format changed. In PISA 2006 there were two boxes: "Yes" or "No". In PISA 2003 there was only one box for each item.*
a	ST10Q01	a	ST14Q01	<ISCED 5A. 6>	
b	ST10Q02	b	ST14Q02	<ISCED 5B>	
c	ST10Q03	c	ST14Q03	<ISCED 4>	

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



[Part 2/2]

Table A4.1 Student questionnaire

PISA 2006 Question number	PISA 2006 Variable name	PISA 2003 Question number	PISA 2003 Variable name	PISA 2006 English version	Summary of changes from PISA 2003
Q11a		**Q15a**		**In what country were you and your parents born?**	
	ST11Q01	**a**	**ST15Q01**	You	*Unchanged*
	ST11Q02	**b**	**ST15Q02**	Mother	*Unchanged*
	ST11Q03	**c**	**ST15Q03**	Father	*Unchanged*
Q11b	**ST11Q04**	**Q15b**	**ST15Q04**	**If you were NOT born in <country of test>, how old were you when you arrived in <country of test>?**	*Unchanged*
Q12	**ST12Q01**	**Q16**	**ST16Q01**	What language do you speak at home most of the time?	*Unchanged*
Q13		**Q17**		Which of the following are in your home?	*Wording is identical but response format changed. In PISA 2006 there were two tick-boxes: "Yes" or "No". In PISA 2003 there was only one tick-box for each item.*
a	**ST13Q01**	**a**	**ST17Q01**	A desk to study at	
b	**ST13Q02**	**b**	**ST17Q02**	A room of your own	
c	**ST13Q03**	**c**	**ST17Q03**	A quiet place to study	
d	**ST13Q04**	**d**	**ST17Q04**	A computer you can use for school work	
e	**ST13Q05**	**e**	**ST17Q05**	Educational software	
f	**ST13Q06**	**f**	**ST17Q06**	A link to the Internet	
g	**ST13Q07**	**g**	**ST17Q07**	Your own calculator	
h	**ST13Q08**	**h**	**ST17Q08**	Classic literature (*e.g.* <Shakespeare>)	
i	**ST13Q09**	**I**	**ST17Q09**	Books of poetry	
j	**ST13Q10**	**j**	**ST17Q10**	Works of art (*e.g.* paintings)	
k	**ST13Q11**	**k**	**ST17Q11**	Books to help with your school work	
l	**ST13Q12**	**l**	**ST17Q12**	A dictionary	
m	**ST13Q13**	**m**	**ST17Q13**	A dishwasher	
Q15	**ST15Q01**	**Q19**	**ST19Q01**	How many books are there in your home? 0-10 books 11-25 books 26-100 books 101-200 books 201-500 books More than 500 books	*Unchanged*
Q30	**ST30Q01**	**Q8**	**EC08Q01**	What kind of job do you expect to have when you are about 30 years old? Write the job title	*Unchanged*



[Part 1/1]

Table A4.2 ICT familiarity questionnaire

PISA 2006 Question number	PISA 2006 Variable name	PISA 2003 Question number	PISA 2003 Variable name	PISA 2006 English version	Summary of changes from PISA 2003
Q1	**IC01Q01**	**Q2**	**IC02Q01**	**Have you ever used a computer?** *If you answered Yes to the above question, please continue. If you answered No, please stop here.*	*Identical except in PISA 2003 it was not a filter question.*
Q2		**Q3**	**IC03Q01**	**How long have you been using computers?**	**How long have you been using computers?**
				Less than one year	Less than one year
				One year or more but less than three years	One to three years
				Three years or more but less than five years	Three to five years
				Five years or more	More than five years
Q3		**Q4**		**How often do you use a computer at these places?**	**How often do you use a computer at these places?**
a	**IC03Q01**	**a**	**IC04Q01**	**At home**	**At home**
b	**IC03Q02**	**b**	**IC04Q02**	**At school**	**At school**
c	**IC03Q03**	**c**	**IC04Q03**	**At other places**	**At other places**
				Almost every day	Almost every day
				Once or twice a week	A few times each week
				A few times a month	Between once a week and once a month
				Once a month or less	Less than once a month
				Never	Never
Q4		**Q5**		**How often do you use computers for the following reasons?**	**How often do you use:**
a	**IC04Q01**	**a**	**IC05Q01**	Browse the Internet for information about people, things, or ideas	the Internet to look up information about people, things, or ideas?
b	**IC04Q02**	**b**	**IC05Q02**	Play games	games on a computer?
c	**IC04Q03**	**c**	**IC05Q03**	Write documents (*e.g.* with <Word® or WordPerfect®>)	Word processing (*e.g.* <Word® or WordPerfect®>)?
d	**IC04Q04**	**d**	**IC05Q04**	Use the Internet to collaborate with a group or team	the Internet to collaborate with a group or team?
e	**IC04Q05**	**e**	**IC05Q05**	Use spreadsheets (*e.g.* <Lotus 1 2 3® or Microsoft Excel®>)	spreadsheets (*e.g.* <Lotus 1 2 3® or Microsoft Excel®>)?
f	**IC04Q06**	**f**	**IC05Q06**	Download software from the Internet (including games)	the Internet to download software (including games)?
g	**IC04Q07**	**g**	**IC05Q07**	Drawing, painting or using graphics programmes	drawing, painting or graphics programmes on a computer?
h	**IC04Q08**	**h**	**IC05Q08**	Use educational software such as Mathematics programmes	educational software such as Mathematics programmes?
I	**IC04Q09**	**j**	**IC05Q10**	Download music from the Internet	the Internet to down-load music?
j	**IC04Q10**	**k**	**IC05Q11**	Writing computer programmes	the computer for programming?
k	**IC04Q11**	**l**	**IC05Q12**	For communication (*e.g.* E-mail or "chat rooms")	a computer for electronic communication (*e.g.* e-mail or "chat rooms")?
				Almost every day	Almost every day
				A few times each week	A few times each week
				Between once a week and once a month	Between once a week and once a month
				Less than once a month	Less than once a month
				Never	Never
Q5		**Q6**		**How well can you do each of these tasks on a computer?**	*Unchanged*
b	**IC05Q02**	**b**	**IC06Q02**	Use software to find and get rid of computer viruses	*Unchanged*
f	**IC05Q06**	**k**	**IC06Q11**	Move files from one place to another on a computer	*Unchanged*
h	**IC05Q08**	**m**	**IC06Q13**	Download files or programmes from the Internet	*Unchanged*
i	**IC05Q09**	**n**	**IC06Q14**	Attach a file to an e-mail message	*Unchanged*
k	**IC05Q11**	**p**	**IC06Q16**	Use a spreadsheet to plot a graph	*Unchanged*
l	**IC05Q12**	**q**	**IC06Q17**	Create a presentation (*e.g.* using <Microsoft PowerPoint®>)	*Unchanged*
m	**IC05Q13**	**s**	**IC06Q19**	Download music from the Internet.	*Unchanged*
n	**IC05Q14**	**t**	**IC06Q20**	Create a multi-media presentation (with sound, pictures, video)	*Unchanged*
o	**IC05Q15**	**v**	**IC06Q22**	Write and send E-mails	*Unchanged*
p	**IC05Q16**	**w**	**IC06Q23**	Construct a web page	*Unchanged*
				I can do this very well by myself	*Unchanged*
				I can do this with help from someone	*Unchanged*
				I know what this means but I cannot do it	*Unchanged*
				I don't know what this means	*Unchanged*

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



[Part 1/3]

Table A4.3 School questionnaire

PISA 2006 Question number	PISA 2006 Variable name	PISA 2003 Question number	PISA 2003 Variable name	PISA 2006 English version	Summary of changes from PISA 2003
Q1		**Q2**		**As at <February 1. 2006>, what was the total school enrolment (number of students)?**	*Unchanged*
a	**SC01Q01**	**a**	**SC02Q01**	Number of boys:	*Unchanged*
b	**SC01Q02**	**b**	**SC02Q02**	Number of girls:	*Unchanged*
Q2	**SC02Q01**	**Q3**	**SC03Q01**	**Is your school a public or a private school?**	*Unchanged*
				A public school	
				A private school	
Q3		**Q4**		**About what percentage of your total funding for a typical school year comes from the following sources?**	
a	**SC03Q01**	**a**	**SC04Q01**	Government (includes departments, local, regional, state and national)	*Unchanged*
b	**SC03Q02**	**b**	**SC04Q02**	Student fees or school charges paid by parents	*Unchanged*
c	**SC03Q03**	**c**	**SC04Q03**	Benefactors, donations, bequests, sponsorships, parent fund raising	*Unchanged*
d	**SC03Q04**	**d**	**SC04Q04**	Other	*Unchanged*
Q4		**Q5**		**Do you have the following <grade levels> in your school?**	
a	**SC04Q01**	**a**	**SC05Q01**	<Grade 1>	*Unchanged*
b	**SC04Q02**	**b**	**SC05Q02**	<Grade 2>	*Unchanged*
c	**SC04Q03**	**c**	**SC05Q03**	<Grade 3>	*Unchanged*
d	**SC04Q04**	**d**	**SC05Q04**	<Grade 4>	*Unchanged*
e	**SC04Q05**	**e**	**SC05Q05**	<Grade 5>	*Unchanged*
f	**SC04Q06**	**f**	**SC05Q06**	<Grade 6>	*Unchanged*
g	**SC04Q07**	**g**	**SC05Q07**	<Grade 7>	*Unchanged*
h	**SC04Q08**	**h**	**SC05Q08**	<Grade 8>	*Unchanged*
I	**SC04Q09**	**I**	**SC05Q09**	<Grade 9>	*Unchanged*
j	**SC04Q10**	**j**	**SC05Q10**	<Grade 10>	*Unchanged*
k	**SC04Q11**	**k**	**SC05Q11**	<Grade 11>	*Unchanged*
l	**SC04Q12**	**l**	**SC05Q12**	<Grade 12>	*Unchanged*
m	**SC04Q13**	**m**	**SC05Q13**	<Grade 13>	*Unchanged*
n	**SC04Q14**	**n**	**SC05Q14**	<Ungraded school>	*Unchanged*
Q5		**Q6**		**About what percentage of students in your school repeated a <grade>, at these <ISCED levels>, last academic year?**	*Unchanged. However, in PISA 2003 there was a checkbox labeled "Not Applicable". In PISA 2006 the checkbox was labeled "<ISCED level> not available in this school".*
a	**SC05Q01**	**a**	**SC06Q01**	The approximate percentage of students repeating a <grade> at <ISCED 2> in this school last year was:	
b	**SC05Q02**	**b**	**SC06Q02**	The approximate percentage of students repeating a <grade> at <ISCED 3> in this school last year was:	
Q7	**SC07Q01**	**Q1**	**SC01Q01**	**Which of the following best describes the community in which your school is located?**	*Unchanged*
				A village, hamlet or rural area (fewer than 3 000 people)	
				A small town (3 000 to about 15 000 people)	
				A town (15 000 to about 100 000 people)	
				A city (100 000 to about 1 000 000 people)	
				A large city (with over 1 000 000 people)	
Q8		**Q16**		**Some schools organise instruction differently for students with different abilities. What is your school's policy about this for students in <national modal grade for 15-year-olds>?**	**Schools sometimes organise instruction differently for students with different abilities and interests in Mathematics. Which of the following options describe what your school does for 15-year-old students in Mathematics classes?**
b	**SC08Q02**	**c**	**SC16Q03**	Students are grouped by ability within their classes	Students are grouped by ability within their Mathematics classes.
				For all subjects	For all classes
				For some subjects	For some classes
				Not for any subjects	Not for any classes

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



[Part 2/3]

Table A4.3 School questionnaire

PISA 2006 Question number	PISA 2006 Variable name	PISA 2003 Question number	PISA 2003 Variable name	PISA 2006 English version	Summary of changes from PISA 2003
Q9		**Q18**		**How many of the following are on the staff of your school?**	*Unchanged*
a		**a**		Teachers in TOTAL	*Unchanged*
	SC09Q11		**SC18Q11**	Full time	
	SC09Q12		**SC18Q21**	Part time	
b		**b**		Teachers fully certified by <the appropriate authority>	*Unchanged*
	SC09Q21		**SC18Q12**	Full time	
	SC09Q22		**SC18Q22**	Part time	
c		**c**		Teachers with an <ISCED 5A> qualification	*In PISA 2003: "Teachers with an <ISCED5A> qualification in <pedagogy>"*
	SC09Q31		**SC18Q13**	Full time	
	SC09Q32		**SC18Q23**	Part time	
Q11		**Q26**		**Regarding your school, who has a considerable responsibility for the following tasks?**	**In your school, who has the main responsibility for:**
a	**SC11Q01**	**a**	**SC26Q01**	Selecting teachers for hire	selecting teachers for hire?
b	**SC11Q02**	**b**	**SC26Q02**	Firing teachers	firing teachers?
c	**SC11Q03**	**c**	**SC26Q03**	Establishing teachers' starting salaries	establishing teachers' starting salaries?
d	**SC11Q04**	**d**	**SC26Q04**	Determining teachers' salaries increases	determining teachers' salary increases?
e	**SC11Q05**	**e**	**SC26Q05**	Formulating the school budget	formulating the school budget?
f	**SC11Q06**	**f**	**SC26Q06**	Deciding on budget allocations within the school	deciding on budget allocations within the school?
g	**SC11Q07**	**g**	**SC26Q07**	Establishing student disciplinary policies	establishing student disciplinary policies?
h	**SC11Q08**	**h**	**SC26Q08**	Establishing student assessment policies	establishing student assessment policies?
I	**SC11Q09**	**i**	**SC26Q09**	Approving students for admission to the school	approving students for admittance to the school?
j	**SC11Q10**	**j**	**SC26Q10**	Choosing which textbooks are used	choosing which textbooks are used?
k	**SC11Q11**	**k**	**SC26Q11**	Determining course content	determining course content?
l	**SC11Q12**	**l**	**SC26Q12**	Deciding which courses are offered	deciding which courses are offered?
				Principal or teachers	Not a main responsibility of the school
				<School governing board>	School's <governing board>
				<Regional or local education authority>	Principal
				National education authority	<Department Head>
					Teacher(s)
Q12		**Q27**		**Regarding your school, which of the following bodies exert a direct influence on decision making about staffing, budgeting, instructional content and assessment practices?**	**In your school. which of the following <bodies> exert a direct influence on decision making about staffing, budgeting, instructional content and assessment practises?**
a	**SC12Q01**	**a**	**SC27Q01**	Regional or national education authorities	*Unchanged*
				(*e.g.* inspectorates)	
b	**SC12Q02**	**b**	**SC27Q02**	The school's <governing board>	*Unchanged*
c	**SC12Q03**	**d**	**SC27Q04**	Parent groups	*Unchanged*
d	**SC12Q04**	**e**	**SC27Q05**	Teacher groups	*Unchanged*
				(*e.g.* Staff Association, curriculum committees, trade union)	
e	**SC12Q05**	**f**	**SC27Q06**	Student groups	*Unchanged*
				(*e.g.* Student Association, youth organisation)	
f	**SC12Q06**	**g**	**SC27Q07**	External examination boards	*Unchanged*
				Staffing	*Unchanged*
				Budgeting	*Unchanged*
				Instructional content	*Unchanged*
				Assessment practices	*Unchanged*
		Q9			**In your school, about how many computers are:**
Q13a	**SC13Q01**	**a**	**SC09Q01**	**About how many computers are available in the school altogether?**	in the school altogether?
Q13b	**SC13Q02**	**b**	**SC09Q02**	**About how many of these computers are available for instruction?**	available to 15-year-old students?
Q13c	**SC13Q03**	**e**	**SC09Q05**	**About how many computers in the school are connected to the Internet/World Wide Web?**	connected to the Internet/World Wide Web?

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



[Part 3/3]

Table A4.3 School questionnaire

PISA 2006 Question number	PISA 2006 Variable name	PISA 2003 Question number	PISA 2003 Variable name	PISA 2006 English version	Summary of changes from PISA 2003
Q14		**Q8**		**Is your school's capacity to provide instruction hindered by any of the following?**	**Is your school's capacity to provide instruction hindered by a shortage or inadequacy of any of the following?**
a	**SC14Q01**	**b**	**SC08Q02**	A lack of qualified science teachers	Availability of qualified Science teachers
b	**SC14Q02**	**a**	**SC08Q01**	A lack of qualified mathematics teachers	Availability of qualified Mathematics teachers
c	**SC14Q03**	**c**	**SC08Q03**	A lack of qualified <test language> teachers	Availability of qualified <test language> teachers
f	**SC14Q06**	**h**	**SC08Q08**	A lack of other support personnel	Availability of support personnel.
g	**SC14Q07**	**t**	**SC08Q20**	Shortage or inadequacy of science laboratory equipment	Science laboratory equipment and materials
h	**SC14Q08**	**l**	**SC08Q09**	Shortage or inadequacy of instructional materials (*e.g.* textbooks)	Instructional materials (*e.g.* textbooks)
i	**SC14Q09**	**o**	**SC08Q15**	Shortage or inadequacy of computers for instruction	Computers for instruction
k	**SC14Q11**	**p**	**SC08Q16**	Shortage or inadequacy of computer software for instruction	Computer software for instruction
l	**SC14Q12**	**r**	**SC08Q18**	Shortage or inadequacy of library materials	Library materials
m	**SC14Q13**	**s**	**SC08Q19**	Shortage or inadequacy of audio-visual resources	Audio-visual resources
				Not at all	
				Very little	
				To some extent	
				A lot	
Q17		**Q13**		**In your school, are achievement data used in any of the following <accountability procedures>?**	**In your school, are assessments of <15-year-old students> used for any of the following purposes?**
c	**SC17Q03**	**f**	**SC13Q06**	Achievement data are used in evaluation of teachers' performance	To make judgements about teachers' effectiveness
Q19		**Q10**		**How much consideration is given to the following factors when students are admitted to your school?**	*Unchanged*
a	**SC19Q01**	**a**	**SC10Q01**	Residence in a particular area	*Unchanged*
b	**SC19Q02**	**b**	**SC10Q02**	Student's academic record (including placement tests)	*Unchanged*
c	**SC19Q03**	**c**	**SC10Q03**	Recommendation of feeder schools	*Unchanged*
d	**SC19Q04**	**d**	**SC10Q04**	Parents' endorsement of the instructional or religious philosophy of the school	*Unchanged*
e	**SC19Q05**	**e**	**SC10Q05**	Student's need or desire for a special programme	*Unchanged*
f	**SC19Q06**	**f**	**SC10Q06**	Attendance of other family members at the school (past or present)	*Unchanged*
				Prerequisite	
				High priority	
				Considered	
				Not considered	



APPENDIX 5 Mapping of ISCED to years

[Part 1/1]

Table A5.1 Mapping of ISCED to accumulated years of education

		ISCED 1	ISCED 2	ISCED 3B or 3C	ISCED 3A or 4	ISCED 5B	ISCED 5A or 6
OECD	Australia	6.0	10.0	11.0	12.0	14.0	15.0
	Austria	4.0	9.0	12.0	12.5	15.0	17.0
	Belgium	6.0	9.0	12.0	12.0	14.5	17.0
	Canada	6.0	9.0	12.0	12.0	15.0	17.0
	Czech Republic	5.0	9.0	11.0	13.0	16.0	16.0
	Denmark	6.0	9.0	12.0	12.0	15.0	17.0
	England, Wales & North. Ireland	6.0	9.0	12.0	13.0	15.0	16.0
	Finland	6.0	9.0	12.0	12.0	14.5	16.5
	France	5.0	9.0	12.0	12.0	14.0	15.0
	Germany	4.0	10.0	13.0	13.0	15.0	18.0
	Greece	6.0	9.0	11.5	12.0	15.0	17.0
	Hungary	4.0	8.0	10.5	12.0	13.5	16.5
	Iceland	7.0	10.0	13.0	14.0	16.0	18.0
	Ireland	6.0	9.0	12.0	12.0	14.0	16.0
	Italy	5.0	8.0	12.0	13.0	16.0	17.0
	Japan	6.0	9.0	12.0	12.0	14.0	16.0
	Korea	6.0	9.0	12.0	12.0	14.0	16.0
	Luxembourg	6.0	9.0	12.0	13.0	16.0	17.0
	Mexico	6.0	9.0	12.0	12.0	14.0	16.0
	Netherlands	6.0	10.0		12.0		16.0
	New Zealand	5.5	10.0	11.0	12.0	14.0	15.0
	Norway	6.0	9.0	12.0	12.0	14.0	16.0
	Poland		8.0	11.0	12.0	15.0	16.0
	Portugal	6.0	9.0	12.0	12.0	15.0	17.0
	Scotland	7.0	11.0	13.0	13.0	16.0	16.0
	Slovak Republic	4.5	8.5	12.0	12.0	13.5	17.5
	Spain	5.0	8.0	10.0	12.0	13.0	16.5
	Sweden	6.0	9.0	11.5	12.0	14.0	15.5
	Switzerland	6.0	9.0	12.5	12.5	14.5	17.5
	Turkey	5.0	8.0	11.0	11.0	13.0	15.0
	United States	6.0	9.0		12.0	14.0	16.0
Partners	Argentina	6.0	10.0	12.0	12.0	14.5	17.0
	Azerbaijan	4.0	9.0	11.0	11.0	14.0	17.0
	Brazil	4.0	8.0	11.0	11.0	14.5	16.0
	Bulgaria	4.0	8.0	12.0	12.0	15.0	17.5
	Chile	6.0	8.0	12.0	12.0	16.0	17.0
	Colombia	5.0	9.0	11.0	11.0	14.0	15.5
	Croatia	4.0	8.0	11.0	12.0	15.0	17.0
	Estonia	4.0	9.0	12.0	12.0	15.0	16.0
	Hong Kong-China	6.0	9.0	11.0	13.0	14.0	16.0
	Indonesia	6.0	9.0	12.0	12.0	14.0	15.0
	Israel	6.0	9.0	12.0	12.0	15.0	15.0
	Jordan	6.0	10.0	12.0	12.0	14.5	16.0
	Kyrgyzstan	4.0	8.0	11.0	10.0	13.0	15.0
	Latvia	3.0	8.0	11.0	11.0	16.0	16.0
	Liechtenstein	5.0	9.0	11.0	13.0	14.0	17.0
	Lithuania	3.0	8.0	11.0	11.0	15.0	16.0
	Macao-China	6.0	9.0	11.0	12.0	15.0	16.0
	Montenegro	4.0	8.0	11.0	12.0	15.0	16.0
	Qatar	6.0	9.0	12.0	12.0	15.0	16.0
	Romania	4.0	8.0	11.5	12.5	14.0	16.0
	Russian Federation	4.0	9.0	11.5	12.0		15.0
	Serbia	4.0	8.0	11.0	12.0	14.5	17.0
	Slovenia	4.0	8.0	11.0	12.0	15.0	16.0
	Chinese Taipei	6.0	9.0	12.0	12.0	14.0	16.0
	Thailand	6.0	9.0	12.0	12.0	14.0	16.0
	Tunisia	6.0	9.0	12.0	13.0	16.0	17.0
	Uruguay	6.0	9.0	12.0	12.0	15.0	17.0

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



APPENDIX 6 National household possession items

[Part 1/2]

Table A6.1 National household possession items

		St13q15	St13q16	ST13Q17
OECD	Australia	Cable/Pay TV	Digital Camera	Plasma TV
	Austria	MP3 Player	Digital Camera	Digital Video Camera
	Belgium (Flemish region)	Home Cinema	Alarm System	Plasma or LCD TV
	Belgium (French and German regions)	Home Cinema (LCD screen…)	Alarm System	Housekeeper
	Canada	MP3 Player/iPod	Subscription to a Daily Newspaper	Central Air Conditioning
	Czech Republic	Digital Camera (not part of a mobile phone)	Digital Video Camera	Personal Discman or MP3 Player
	Denmark	Colour Printer	MP3 Player	Digital Camera
	Finland	Digital Camera	Wide Screen TV	Fitness Equipment (*e.g.* exercise bike, rowing machine)
	France	Flat Screen TV	Digital Camera (not part of a mobile phone)	Laptop Computer
	Germany	Subscription to a Newspaper	Video Camera	ISDN Connection
	Greece	Home Cinema	Cable TV (Nova, Filmnet, etc.)	Alarm System
	Hungary	Automatic Washing Machine	Video Camera	Digital Camera (not part of a mobile phone)
	Iceland	Security Service or Security System	Satellite Dish	Plasma TV or TV Projector
	Ireland	MP3 Player (*e.g.* iPod)	Bedroom with an En-suite Bathroom	Premium Cable TV Package (*e.g.* Sky Movies, Sky Sports)
	Italy	Antique Furniture	Plasma TV Set	Air Conditioning
	Japan	Digital Camera	Plasma/Liquid Crystal TV	Clothing Dryer
	Korea	Air Conditioning	Digital Camera	Water Purifier
	Luxembourg	Digital Camera	MP3 Player	Flat Screen TV
	Mexico	Pay TV	Telephone Line	Microwave Oven
	Netherlands	Digital Camera (not part of mobile phone or laptop computer)	Piano	Laptop
	New Zealand	Broadband Internet Connection	Digital Camera (not part of mobile phone)	Clothes Dryer
	Norway	Cleaner	Plasma/LCD TV	Spa Bath
	Poland	Cable TV with at least 30 channels	Digital Camera	Telescope or Microscope
	Portugal	Cable TV or Satellite Dish	Plasma or LCD Screen TV	Central Heating or Air Conditioning Equipment
	Slovak Republic	Video Camera	Digital Camera (not part of mobile phone)	Satelite Receiver or Cable TV
	Spain	Video Camera	Satellite Dish or Digital TV Set	Home Cinema Set
	Sweden	Piano	Video Camera	Wall TV
	Switzerland & Liechtenstein	MP3 Player or iPod	Digital Camera	Digital Video Camera
	Turkey	Air-Conditioning-type Heating and Cooling System	Treadmill (fitness equipment device)	Home Cinema System (5+1)
	United Kingdom (England, Wales & Northern Ireland)	Digital TV	Digital Camcorder	Swimming Pool
	United Kingdom (Scotland)	Video Camera	Plasma Screen TV	Broadband Internet Connection
	United States	Guest Room	High-Speed Internet Connection	iPod or MP3 Player

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



[Part 2/2]

Table A6.1 National household possession items

		ST13Q15	ST13Q16	ST13Q17
Partners	Argentina	Cable TV (Direct TV, Cablevision. etc.)	Telephone Line	Refrigerator with Freezer
	Azerbaijan	Satellite Dish	Video Camera	Colour Printer
	Brazil	Personal Mobile Phone	Cable TV	Video Game
	Bulgaria	Air Conditioning	Freezer	Digital Camera
	Chile	Hot Water	Washing Machine	Digital Video Camera
	Colombia	Refrigerator	Cable TV or Direct to Home TV	Encyclopedia
	Croatia	Video Camera	Clothes Dryer	Air Conditioning
	Estonia	Video Camera	Hi-Fi	Broadband Internet Connection
	Hong Kong-China	Digital Camera / Video Recorder	Musical Instrument (*e.g.* piano, violin)	Pay TV Channel
	Indonesia	Washing Machine	Motorcycle	Air Conditioning
	Israel	Home Alarm System	Digital Camera	Home Movie Theatre
	Jordan	Central Heating	Satellite Dish	Digital Camera
	Kyrgyzstan	Camera	Vacuum Cleaner	Imported Clothes Washing Machine such as an Ariston or an Indesit
	Latvia	Bicycle	Snowboard	Digital Camera
	Switzerland & Liechtenstein	MP3 Player or iPod	Digital Camera	Digital Video Camera
	Lithuania	Digital Camera	Press Subscription Edition (newspaper, magazine)	MP3 Player
	Macao-China	Video Game Machine	Digital Camera	MP3 Player
	Montenegro	Cable TV	Jacuzzi	Digital Camera
	Qatar	MP3 Walkman	Digital Video Camera	X-Box
	Romania	Video Camera / Digital Photo Camera	Cable TV	Air Conditioning
	Russian Federation	Digital Camera or Video Camera	Home Cinema	Satellite Antenna
	Serbia	Digital Camera	Laundry Drying Machine	Cable TV
	Slovenia	Digital Camera or Video Camera	Personal MP3 Player	Sauna
	Chinese Taipei	Musical Instrument	iPod	Jacuzzi Bath
	Thailand	Air Conditioning	Washing Machine	Microwave Oven
	Tunisia	Satellite Dish	Digital Camera	Washing Machine
	Uruguay	Television Subscription	Washing Machine	Microwave Oven

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



APPENDIX 7 Exploratory and confirmatory factor analyses for the embedded items

[Part 1/2]

Table A7.1 Exploratory and confirmatory factor analyses (EFA and CFA) for the embedded items

	EFA		CFA				
	PROMAX ROTATED LOADINGS		Two-dimensional				
Item	1	2	Loadings	Explained variance	Unexplained variance	Conquest fit	Item labels
S408RNA	**0.59**	0.07	0.613	0.38	0.62	1.05	Wild Oat Grass
S408RNB	**0.62**	0.09	0.62	0.38	0.62	0.94	Wild Oat Grass
S408RNC	**0.64**	0.07	0.624	0.39	0.61	0.92	Wild Oat Grass
S413RNA	**0.72**	–0.16	0.515	0.27	0.73	1.04	Plastic Age
S413RNB	**0.72**	–0.12	0.552	0.30	0.70	1.00	Plastic Age
S413RNC	**0.70**	–0.10	0.579	0.34	0.66	1.02	Plastic Age
S416RNA	**0.41**	0.22	0.519	0.27	0.73	1.06	The Moon
S416RNB	**0.37**	0.26	0.517	0.27	0.73	1.08	The Moon
S428RNA	**0.66**	0.00	0.577	0.33	0.67	0.95	Bacteria In Milk
S428RNB	**0.65**	–0.02	0.563	0.32	0.68	0.98	Bacteria In Milk
S428RNC	**0.65**	0.01	0.622	0.39	0.61	0.97	Bacteria In Milk
S437RNA	**0.57**	–0.01	0.485	0.24	0.76	1.09	Extinguishing Fires
S437RNB	**0.59**	–0.01	0.512	0.26	0.74	1.05	Extinguishing Fires
S437RNC	**0.57**	0.04	0.545	0.30	0.70	1.07	Extinguishing Fires
S438RNA	**0.81**	–0.20	0.591	0.35	0.65	0.94	Green Parks
S438RNB	**0.80**	–0.18	0.594	0.35	0.65	0.92	Green Parks
S438RNC	**0.79**	–0.13	0.656	0.43	0.57	0.89	Green Parks
S456RNA	**0.40**	**0.32**	0.533	0.28	0.72	1.00	The Cheetah
S456RNB	**0.38**	**0.36**	0.552	0.30	0.70	0.99	The Cheetah
S456RNC	**0.36**	**0.36**	0.555	0.31	0.69	1.04	The Cheetah
S466RNA	**0.55**	0.05	0.507	0.26	0.74	1.04	Forest Fires
S466RNB	**0.61**	0.05	0.57	0.32	0.68	0.96	Forest Fires
S466RNC	**0.51**	0.15	0.549	0.30	0.70	1.05	Forest Fires
S476RNA	**0.48**	0.19	0.523	0.27	0.73	0.98	Heart Surgery
S476RNB	**0.49**	0.17	0.556	0.31	0.69	1.01	Heart Surgery
S476RNC	**0.43**	0.19	0.528	0.28	0.72	1.11	Heart Surgery
S478RNA	**0.58**	0.11	0.587	0.34	0.66	0.97	Antibiotics
S478RNB	**0.58**	0.09	0.58	0.34	0.66	0.99	Antibiotics
S478RNC	**0.56**	0.14	0.622	0.39	0.61	0.97	Antibiotics
S485RNA	**0.43**	0.24	0.538	0.29	0.71	1.02	Acid Rain
S485RNB	**0.46**	0.25	0.598	0.36	0.64	1.00	Acid Rain
S485RNC	**0.50**	0.12	0.538	0.29	0.71	1.06	Acid Rain
S498RNA	**0.63**	0.07	0.597	0.36	0.64	0.93	Experimental Digestion
S498RNB	**0.63**	0.10	0.623	0.39	0.61	0.92	Experimental Digestion
S498RNC	**0.56**	0.18	0.625	0.39	0.61	0.94	Experimental Digestion
S508RNA	**0.52**	0.15	0.585	0.34	0.66	0.98	Genetically Modified Crops
S508RNB	**0.57**	0.12	0.602	0.36	0.64	0.95	Genetically Modified Crops
S508RNC	**0.50**	0.17	0.599	0.36	0.64	1.02	Genetically Modified Crops
S514RNA	**0.72**	–0.07	0.601	0.36	0.64	0.99	Development And Disaster
S514RNB	**0.73**	–0.05	0.614	0.38	0.62	0.94	Development And Disaster
S514RNC	**0.58**	0.10	0.582	0.34	0.66	1.01	Development And Disaster
S519RNA	**0.35**	0.24	0.407	0.17	0.83	1.10	Airbags
S519RNB	**0.34**	0.22	0.416	0.17	0.83	1.15	Airbags
S519RNC	**0.41**	0.18	0.482	0.23	0.77	1.14	Airbags
S521RNA	**0.63**	–0.04	0.526	0.28	0.72	0.99	Cooking Outdoors

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



[Part 2/2]

Table A7.1 Exploratory and confirmatory factor analyses (EFA and CFA) for the embedded items

	EFA		CFA				
	PROMAX ROTATED LOADINGS		Two-dimensional				
Item	1	2	Loadings	Explained variance	Unexplained variance	Conquest fit	Item labels
S521RNB	**0.61**	–0.04	0.504	0.25	0.75	1.04	Cooking Outdoors
S524RNA	**0.64**	0.07	0.625	0.39	0.61	0.94	Penicillin Manufacture
S524RNB	**0.61**	0.10	0.639	0.41	0.59	0.95	Penicillin Manufacture
S524RNC	**0.63**	0.07	0.637	0.41	0.59	0.95	Penicillin Manufacture
S527RNA	**0.28**	**0.32**	0.464	0.22	0.78	1.16	Extinction Of The Dinosaurs
S527RNB	0.19	**0.45**	0.475	0.23	0.77	1.18	Extinction Of The Dinosaurs
S527RNC	**0.33**	**0.33**	0.552	0.30	0.70	1.09	Extinction Of The Dinosaurs
S408RSA	0.13	**0.44**	0.38	0.14	0.86	0.97	Wild Oat Grass
S408RSB	0.04	**0.50**	0.43	0.18	0.82	0.98	Wild Oat Grass
S408RSC	0.14	**0.42**	0.41	0.16	0.84	1.01	Wild Oat Grass
S416RSA	–0.02	**0.44**	0.29	0.09	0.91	1.01	The Moon
S416RSB	–0.06	**0.49**	0.32	0.10	0.90	1.00	The Moon
S416RSC	–0.02	**0.51**	0.36	0.13	0.87	1.01	The Moon
S421RSA	0.03	**0.55**	0.40	0.16	0.84	0.92	Big And Small
S421RSC	0.08	**0.51**	0.42	0.18	0.82	0.95	Big And Small
S425RSA	–0.05	**0.53**	0.32	0.10	0.90	0.97	Penguin Island
S425RSB	–0.04	**0.53**	0.35	0.12	0.88	0.96	Penguin Island
S425RSC	0.01	**0.37**	0.31	0.10	0.90	1.08	Penguin Island
S426RSA	0.05	**0.52**	0.37	0.14	0.86	0.97	The Grand Canyon
S426RSB	0.03	**0.36**	0.30	0.09	0.91	1.12	The Grand Canyon
S426RSC	0.04	**0.54**	0.40	0.16	0.84	0.95	The Grand Canyon
S438RSA	0.03	**0.31**	0.25	0.06	0.94	1.09	Green Parks
S438RSB	0.07	**0.29**	0.28	0.08	0.92	1.06	Green Parks
S438RSC	0.11	**0.29**	0.31	0.10	0.90	1.07	Green Parks
S456RSA	0.03	**0.56**	0.38	0.14	0.86	0.93	The Cheetah
S456RSB	0.07	**0.46**	0.36	0.13	0.87	0.98	The Cheetah
S456RSC	–0.01	**0.44**	0.35	0.13	0.87	1.05	The Cheetah
S465RSA	–0.02	**0.39**	0.26	0.07	0.93	1.04	Different Climates
S465RSB	0.04	**0.40**	0.30	0.09	0.91	1.04	Different Climates
S476RSA	–0.07	**0.51**	0.31	0.10	0.90	0.99	Heart Surgery
S476RSB	–0.15	**0.49**	0.25	0.06	0.94	1.03	Heart Surgery
S476RSC	–0.17	**0.56**	0.25	0.06	0.94	0.96	Heart Surgery
S477RSA	–0.07	**0.52**	0.31	0.10	0.90	0.98	Mary Montagu
S477RSB	–0.08	**0.41**	0.29	0.08	0.92	1.09	Mary Montagu
S477RSC	–0.11	**0.52**	0.31	0.10	0.90	0.99	Mary Montagu
S485RSB	–0.01	**0.51**	0.36	0.13	0.87	0.97	Acid Rain
S485RSC	–0.04	**0.55**	0.38	0.14	0.86	0.94	Acid Rain
S498RSA	0.02	**0.46**	0.33	0.11	0.89	1.00	Experimental Digestion
S498RSB	–0.03	**0.47**	0.34	0.11	0.89	1.03	Experimental Digestion
S519RSA	–0.12	**0.51**	0.29	0.08	0.92	1.02	Airbags
S519RSB	–0.15	**0.55**	0.31	0.10	0.90	0.99	Airbags
S519RSC	–0.08	**0.44**	0.26	0.07	0.93	1.06	Airbags
S527RSB	–0.14	**0.59**	0.34	0.12	0.88	0.99	Extinction Of The Dinosaurs
S527RSC	–0.09	**0.64**	0.40	0.16	0.84	0.93	Extinction Of The Dinosaurs
			RMSEA	0.025			

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



APPENDIX 8 PISA consortium, staff and consultants

PISA Technical Advisory Group

Keith Rust (Chair) (Westat, United States)
Ray Adams (International Project Director, ACER)
Aletta Grisay (Consultant, France)
John de Jong (Language Testing Services, The Netherlands)
Norman Verhelst (CITO, The Netherlands)
Christian Monseur (Université de Liège, Belgium)
Thierry Rocher (Ministere de l'Éducation Nationale, France) (From October 2005)
David Kaplan (University of Wisconsin, United States) (From May 2005)
Kentaro Yamamoto (ETS – New Jersey, United States) (From July 2006)
Larry Hedges (Northwestern University, United States) (To July 2006)
Rebecca Zwick (University of California – Santa Barbara, United States) (From March 2007)
Steve May (Ministry of Education, New Zealand) (To October 2005)
Pierre Foy (IEA Data Processing Centre, Germany) (To October 2005)
J. Douglas Willms (University of New Brunswick, Canada) (To May 2005)
Eugene Johnson (American Institutes for Research, United States) (To October 2005)

PISA Expert Groups

Science Expert Group

Rodger Bybee (Chair) (BSCS, Colorado Springs, United States)
Ewa Bartnik (Warsaw University, Poland)
Peter Fensham (Queensland University of Technology, Australia)
Paulina Korsnakova (National Institute for Education, Slovak Republic)
Robert Laurie (University of New Brunswick, Canada)
Svein Lie (University of Oslo, Norway)
Pierre Malléus (Ministère de l'Education nationale, de l'enseignement supérieur et de la recherche, France)
Michaela Mayer (INVALSI, Italy)
Robin Millar (University of York, UK)
Yasushi Ogura (National Institute for Educational Policy Research, Japan)
Manfred Prenzel (University of Kiel, Germany)
Andrée Tiberghien (Université de Lyon, France)

Mathematics Expert Group

Jan de Lange (Chair) (Freudenthal Institute, Utrecht University, The Netherlands)
Werner Blum (University of Kassel, Germany)
John Dossey (Consultant, United States)
Mogens Niss (University of Roskilde, Denmark)
Zbigniew Marciniak (University of Warsaw, Poland)
Yoshi Shimizu (Tsukuba University, Japan)

Reading Expert Group

John de Jong (Chair from September 2005) (Language Testing Services, The Netherlands)
Irwin Kirsch (Chair to September 2005) (ETS – Princeton, United States)
Dominique Lafontaine (Université de Liège, Belgium)
Pirjo Linnakylä (University of Jyväskylä, Finland)
Martine Rémond (Université de Paris 8 et IUFM de Créteil, France)
Alan Davies (University of Edinburgh, UK)
Marilyn Binkley (National Centre for Educational Statistics, United States)
Stan Jones (Statistics Canada, Canada)

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



Questionnaire Expert Group

David Baker (Pennsylvania State University, United States)
Rodger Bybee (BSCS, Colorado Springs, United States)
Aletta Grisay (Consultant, France)
David Kaplan (University of Wisconsin – Madison, United States)
John Keeves (Flinders University, Australia)
Reinhard Pekrun (University of Munich, Germany)
Erich Ramseier (Abteilung Bildungsplanung und Evaluation, Switzerland)
J. Douglas Willms (University of New Brunswick, Canada)

ACER

Ray Adams (International Project Director)
Alla Berezner (Data management and analysis)
Yan Bibby (Data processing and analysis)
Wei Buttress (Project administration, quality monitoring)
Mary Blackwood (Science test development)
Renee Chow (Data processing and analysis)
Judith Cosgrove (Data processing and analysis, national centre support)
George Doukas (Data processing and analysis, computer-based assessment)
Eveline Gebhardt (Data management and analysis)
Sam Haldane (IT services, computer-based assessment)
Dewi Handayani (Data processing, field operations)
John Harding (Science test development)
Jennifer Hong (Data processing, sampling)
Marten Koomen (Management, computer-based assessment)
Dulce Lay (Data processing, field operations, sampling)
Le Tu Luc (Data processing and analysis)
Tom Lumley (Reading test development)
Helen Lye (Science test development)
Greg Macaskill (Data management and processing, sampling)
Fran Maher (Science test development)
Ron Martin (Mathematics test development)
Barry McCrae (Maths, Science, Reading test development)
Pippa McKelvie (Project administration, data processing, quality monitoring)
Joy McQueen (Reading test development)
Juliette Mendelovits (Reading test development)
Esther Michael (Administrative support)
Martin Murphy (Field operations and sampling)
Van Nguyen (Data processing and analysis)
Gayle O'Connor (Science test development)
Alla Routitsky (Data management and analysis), ACER – Data Analysis
Dara Searle (Reading test development)
Wolfram Schulz (Questionnaire development and analysis)
Fionnuala Shortt (Data processing, quality monitoring)
Ross Turner (Management, Mathematics test development)
Daniel Urbach (Data processing and analysis)
Maurice Walker (Sampling, questionnaire development and data analysis)
Wahyu Wardono (Project administration, computer-based assessment)

CITO

Janny Harmsen (Project administration)
Kees Lagerwaard (Mathematics test development)
Ger Limpens (Mathematics test development)
Norman Vorhelst (Technical advice, data analysis)
Jose Bruens (Science test development)
Joop Hendricx (Science test development)
Annemarie de Knecht (Management)

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



Educational Testing Service (ETS)

Irwin Kirsch (Reading framework and test development)

NIER

Hanako Senuma (Mathematics test development)
Yasushi Ogura (Science test development)

Westat

Keith Rust (Director of the PISA Consortium for sampling and weighting, Chair of TAG)
Sheila Krawchuk (Sampling, weighting and quality monotoring)
Eugene Brown (Weighting)
Ming Chen (Weighting)
Fran Cohen (Weighting)
Joseph Croos (Weighting)
Susan Fuss (Sampling, weighting and quality monotoring)
Ismael Flores-Cervantes (Quality monotoring)
Amita Gopinath (Weighting)
Sharon Hirabayashi (Weighting)
John Lopdell (Weighting)
Shawn Lu (Weighting)
Christian Monseur (Consultant, sampling, weighting and quality monotoring)
Merl Robinson (Quality Monitoring)
William Wall (Weighting)
Erin Wilson (Sampling and weighting)

Other Consultants

Steve Dept (cApStAn Linguistic Quality Control, Belgium) (Translation and verification services)
Andrea Farrari (cApStAn Linguistic Quality Control, Belgium) (Translation and verification services)
Oystein Guttersrud (ILS, University of Oslo, Norway) (Science test development)
Marit Kjaernsli (ILS, University of Oslo, Norway) (Science test development)
Svein Lie (ILS, University of Oslo, Norway) (Science test development)
Rolf V. Olsen (ILS, University of Oslo, Norway) (Science test development)
Steffen Brandt (IPN, University of Kiel, Germany) (Science test development)
Claus Carstensen (IPN, University of Kiel, Germany) (Science test development)
Barbara Dreschel (IPN, University of Kiel, Germany) (Science test development)
Marcus Hammann (IPN, University of Kiel, Germany) (Science test development)
Michael Komorek (IPN, University of Kiel, Germany) (Science test development)
Manfred Prezel (IPN, University of Kiel, Germany) (Science test development, Questionnaire framework development)
Peter Nentwig (IPN, University of Kiel, Germany) (Science test development)
Tina Seidel (IPN, University of Kiel, Germany) (Questionnaire framework development)
Martin Senkbeil (IPN, University of Kiel, Germany) (Science test development)
Béatrice Halleux (Consultant, Belgium) (Translation/verification referee, French source development)
Aletta Grisay (Consultant, France) (Technical advice, French source development, questionnaire development)
Anne-Laure Monnier (Consultant, France) (French source development)
Christian Monseur (Université de Liège, Belgium) (Technical advice, data analysis)
Eve Recht (Consultant, Australia) (Editorial services)
Peter Watson (Consultant, Australia) (Science test development)
Alexander Wiseman (University of Tulsa, United States) (Questionnaire framework development)

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

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(98 2009 04 1 P) ISBN 978-92-64-04808-9 – No. 56393 2009

PISA 2006 Technical Report

The *PISA 2006 Technical Report* describes the methodology underlying the PISA 2006 survey. 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, methodologies used to analyse the data, technical features of the project and quality control mechanisms.

The full text of this book is available on line via this link:
www.sourceoecd.org/education/9789264048089

Those with access to all OECD books on line should use this link:
www.sourceoecd.org/9789264048089

SourceOECD is the OECD's online library of books, periodicals and statistical databases.
For more information about this award-winning service and free trials ask your librarian, or write to us at **SourceOECD@oecd.org**.

OECD*publishing*
www.oecd.org/publishing
www.pisa.oecd.org

ISBN 978-92-64-04808-9
98 2009 04 1 P

