



# ICCS 2009 Technical Report

Edited by  
**Wolfram Schulz**  
**John Ainley**  
**Julian Fraillon**



**Università degli Studi Roma Tre**  
**Laboratorio di Pedagogia sperimentale**

# **ICCS 2009 Technical Report**

Edited by

**Wolfram Schulz**

**John Ainley**

**Julian Fraillon**

Contributors

**John Ainley**

**Falk Brese**

**Julian Fraillon**

**Tim Friedman**

**Michael Jung**

**David Kerr**

**Bruno Losito**

**Barbara Malak**

**Wolfram Schulz**

**Caroline Vandenplas**

**Alana Yu**

**Olaf Zuehlke**



**Università degli Studi Roma Tre**  
*Laboratorio di Pedagogia sperimentale*

Copyright © 2011 International Association for the Evaluation of Educational Achievement (IEA)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise without permission in writing from the copyright holder.

ISBN/EAN: 978-90-79549-00-9

Copies of this publication can be obtained from:

The Secretariat

International Association for the Evaluation of Educational Achievement

Herengracht 487

1017 BT Amsterdam, The Netherlands

Telephone + 31 20 625 3625

Fax + 31 20 420 7136

Email: [Department@IEA.nl](mailto:Department@IEA.nl)

Website: [www.iea.nl](http://www.iea.nl)

The International Association for the Evaluation of Educational Achievement, known as IEA, is an independent, international consortium of national research institutions and governmental research agencies, with headquarters in Amsterdam. Its primary purpose is to conduct large-scale comparative studies of educational achievement with the aim of gaining more in-depth understanding of the effects of policies and practices within and across systems of education.

*Copyedited by Paula Wagemaker Editorial Services, Christchurch, New Zealand*

*Design and production by Becky Bliss Design and Production, Wellington, New Zealand*

*Printed by MultiCopy Netherlands b.v.*

# Contents

<i>List of tables and figures</i>	7
<b>Chapter 1: Overview of the IEA International Civic and Citizenship Education Study</b>	<b>15</b>
Introduction	15
ICCS assessment of knowledge and perceptions	16
Contexts for civic and citizenship education	17
Instruments	17
Measures	18
Populations and samples	19
Outline of this report	19
<b>Chapter 2: ICCS test development</b>	<b>21</b>
Introduction	21
Test scope and format	21
Test-development process	23
Field trial test design and content	26
Main survey	28
Released test items	30
Summary	31
<b>Chapter 3: ICCS questionnaire development</b>	<b>33</b>
Introduction	33
Conceptual framework for questionnaire development	33
Student questionnaire development	36
Teacher questionnaire development	38
School-questionnaire development	39
Development and implementation of the national contexts survey	41
Summary	43
<b>Chapter 4: The development of regional instruments</b>	<b>45</b>
Introduction	45
European test and questionnaire development	45
Latin American test and questionnaire development	47
Asian questionnaire development	48
Summary	49
<b>Chapter 5: Translation and national adaptations of ICCS 2009 instruments</b>	<b>51</b>
Introduction	51
Translation of ICCS 2009 instruments	51
International verification of the instruments	55
Layout verification	57
Summary	58
<b>Chapter 6: Sampling design and implementation</b>	<b>59</b>
Introduction	59
Target-population definitions	59
Coverage and exclusions	60
Sample size requirements	61

School sampling design	63
Within-school sampling design	67
Teacher sampling	68
Summary	68
<b>Chapter 7: Sampling weights and participation rates</b>	<b>69</b>
Introduction	69
Within-school participation requirements	69
Calculating student weights	70
Calculating teacher weights	73
Calculating school weights	75
Calculating participation rates	76
ICCS standards for sampling participation	80
Summary	87
<b>Chapter 8: ICCS survey operations procedures</b>	<b>89</b>
Introduction	89
The role of the national research coordinators	89
Documentation and software	89
Survey tracking forms	90
Contacting schools and sampling classes	91
Linking students to classes and schools and teachers to schools	93
Preparing the test instruments for data collection	93
Administering the ICCS assessment	94
Quality control	95
Scoring the ICCS assessment	96
Documenting scoring reliability	96
Online data collection for school principal and teacher questionnaires	98
ICCS field trial	100
Summary	101
<b>Chapter 9: Quality assurance in the ICCS data collection</b>	<b>103</b>
Introduction	103
Quality control observations of the ICCS data collection	103
Survey activities questionnaire	112
Assembling and printing the ICCS assessment materials	115
Administering the booklets and questionnaires	116
Summary	119
<b>Chapter 10: Data management and creation of the ICCS international database</b>	<b>121</b>
Introduction	121
Confirming the integrity of the ICCS international database	121
Data checks at the IEA Data Processing and Research Center	122
Data-cleaning quality control	122
Preparing national data files	123
National cleaning documentation	126
The ICCS international database	127
Summary	128



<b>Chapter 11: Scaling procedures for ICCS test items</b>	<b>129</b>
Introduction	129
The scaling model	129
Test coverage and item dimensionality	130
Assessment of item fit	131
Assessment of scorer reliabilities	134
Differential item functioning by gender	134
Cross-national measurement equivalence	136
Missing data issues	137
International item adjudication	139
International item calibration and test reliability	140
International ability estimates	143
Development of proficiency levels for civic knowledge	147
Estimation of changes between 1999 and 2009	148
Regional cognitive items and scales	152
Summary	155
<b>Chapter 12: Scaling procedures for ICCS questionnaire items</b>	<b>157</b>
Introduction	157
Simple indices	157
Scaling procedures	160
Scaled indices	164
Summary	258
<b>Chapter 13: The reporting of ICCS results</b>	<b>261</b>
Introduction	261
Estimation of sampling variance	261
Estimation of imputation variance for civic knowledge scores	263
Reporting of differences	265
Multiple regression modeling	268
Hierarchical linear modeling	269
Missing data treatment in ICCS multivariate analyses	270
Summary	280
<b>Appendices</b>	<b>281</b>
Appendix A: Organizations and individuals involved in ICCS	281
Appendix B: Characteristics of national samples	286
Appendix C: Descriptions for cognitive test items and allocations to proficiency levels	315
Appendix D: Tables with coding information	319





# List of Tables and Figures

## Tables

Table 2.1:	Test development processes and timeline	22
Table 2.2:	Composition of the field trial test instrument by item type and origin	26
Table 2.3:	Field trial cluster composition	27
Table 2.4:	Field trial test booklet design	27
Table 2.5:	Planned field trial item mapping to assessment framework	27
Table 2.6:	Field trial item mapping to assessment framework	28
Table 2.7:	Composition of the main study test instrument by item type and origin	29
Table 2.8:	Main study cluster composition	29
Table 2.9:	Main study test booklet design	29
Table 2.10:	Main study item mapping to assessment framework	30
Table 2.11:	Characteristics and mapping of ICCS released test items	30
Table 3.1:	Mapping of variables to contextual framework (examples)	35
Table 5.1:	Languages used for the ICCS 2009 survey instruments	53
Table 6.1:	Population coverage and exclusion rates	62
Table 6.2:	School, student, and teacher sample sizes	64
Table 7.1:	Unweighted participation rates in the student survey	78
Table 7.2:	Weighted participation rates in the student survey	79
Table 7.3:	Unweighted participation rates in the teacher survey	81
Table 7.4:	Weighted participation rates in the teacher survey	82
Table 7.5:	Categories into which countries were placed with respect to student sampling participation rates	84
Table 7.6:	Categories into which countries were placed with respect to teacher sampling participation rates	85
Table 7.7:	Participation by country in the student and teacher surveys	86
Table 8.1:	Hierarchical identification (ID) system	93
Table 8.2:	Time allowed for administering the ICCS student instruments	95
Table 8.3:	Weighted percentages of online mode administration for school and teacher questionnaires	100
Table 9.1:	Percentages of IQCM responses for preliminary activities of the test administrator	105
Table 9.2:	Percentages of IQCM responses for administration of the ICCS achievement test	107
Table 9.3:	Percentages of IQCM responses for administration of the ICCS student questionnaire	108
Table 9.4:	Percentages of IQCM responses for administration of the ICCS regional module	109
Table 9.5:	Percentages of IQCM responses for observations of student behavior	110
Table 9.6:	Percentages of IQCM responses for general observations	111
Table 9.7:	Percentages of IQCM responses from interviews with the school coordinator and/or test administrator: overall impressions	112
Table 9.8:	Percentages of IQCM responses from interviews with the school coordinator and/or test administrator: receipt of materials and test administration	113





Table 9.9:	Numbers of NRC responses to the survey activities questionnaire: sampling	114
Table 9.10:	Numbers of NRC responses to the survey activities questionnaire: contacting schools and recruiting school coordinators	114
Table 9.11:	Numbers of NRC responses to the survey activities questionnaire: adapting and translating the ICCS assessment materials	115
Table 9.12:	Numbers of NRC responses to the survey activities questionnaire: assembling and printing the ICCS assessment materials	116
Table 9.13:	Numbers of NRC responses to the survey activities questionnaire: administering the booklets and questionnaires	116
Table 9.14:	Numbers of NRC responses to the survey activities questionnaire: national quality control monitoring	117
Table 9.15:	Numbers of NRC responses to the survey activities questionnaire: scoring open-ended response items and coding occupation data	118
Table 9.16:	Numbers of NRC responses to the survey activities questionnaire: entering and submitting data	118
Table 9.17:	Numbers of NRC responses to the survey activities questionnaire: other experience	119
Table 11.1:	Item total-score correlations and weighted item fit for international calibration sample	132
Table 11.2:	Percentages of scorer agreement for open-ended ICCS test items	135
Table 11.3:	Gender DIF estimates for retained test items	136
Table 11.4:	Percentages of omitted and invalid responses for test items	138
Table 11.5:	Example of item statistics provided to national centers	141
Table 11.6:	National items excluded from scaling	142
Table 11.7:	Final item parameters used to scale the international test items	144
Table 11.8:	Estimated ICCS booklet parameters	145
Table 11.9:	Median test reliabilities across booklets for national samples (Cronbach's alpha)	146
Table 11.10:	Proficiency level cut-points and percentage of students at each level	148
Table 11.11:	National items excluded from the ICCS equating procedures	150
Table 11.12:	Item parameters and average percentage correct for link items	151
Table 11.13:	Test reliabilities for link items (Cronbach's alpha)	152
Table 11.14:	Item parameters, item fit, and item-total correlations from tentative calibration of European regional test items	153
Table 11.15:	Reliabilities for European test items	154
Table 11.16:	Item parameters, item fit, and item-total correlations for Latin American regional test items	154
Table 11.17:	Reliabilities for Latin American test items	155
Table 12.1:	Reliabilities for scales reflecting students' behaviors	166
Table 12.2:	Item parameters for scales reflecting students' behaviors	167
Table 12.3:	Reliabilities for scales reflecting students' perceptions of the school context	169
Table 12.4:	Item parameters for scales reflecting students' perceptions of the school context	170
Table 12.5:	Item parameters for scales reflecting students' support for democratic value beliefs	173



Table 12.6:	Item parameters for scale reflecting students' democratic value beliefs	174
Table 12.7:	Reliabilities for scales reflecting students' perceptions of the importance of citizenship behaviors	175
Table 12.8:	Item parameters for scales reflecting students' perceptions of public service	176
Table 12.9:	Reliabilities for scales reflecting students' civic-related self-beliefs	178
Table 12.10:	Item parameters for scales reflecting students' civic-related self-beliefs	179
Table 12.11:	Reliabilities for scales reflecting students' attitudes toward equal rights	181
Table 12.12:	Item parameters for scales reflecting students' attitudes toward equal rights	182
Table 12.13:	Reliabilities for scales reflecting students' attitudes toward civic institutions and their country	184
Table 12.14:	Item parameters for scales reflecting students' attitudes toward civic institutions and their country	185
Table 12.15:	Reliabilities for scales reflecting students' expected participation in political protest	187
Table 12.16:	Item parameters for scales reflecting students' expected participation in political protest	188
Table 12.17:	Reliabilities for scales reflecting students' expected political participation	190
Table 12.18:	Item parameters for scales reflecting students' expected political participation	191
Table 12.19:	Reliabilities for scale reflecting students' attitudes toward the influence of religion in society	192
Table 12.20:	Item parameters for scales reflecting students' attitudes toward the influence of religion in society	193
Table 12.21:	Factor loadings and reliabilities for the national index of students' socioeconomic background	194
Table 12.22:	Reliabilities for scales reflecting teachers' perceptions of school governance	196
Table 12.23:	Reliabilities for scales reflecting teachers' perceptions of school governance	197
Table 12.24:	Reliabilities for scales reflecting teachers' perceptions of teaching in classes	198
Table 12.25:	Item parameters for scales reflecting teachers' perceptions of teaching in classes	199
Table 12.26:	Reliabilities for scales reflecting teachers' perceptions of participation in the community	202
Table 12.27:	Item parameters for scales reflecting teachers' perceptions of participation in the community	203
Table 12.28:	Reliabilities for scales reflecting teachers' perceptions of classroom climate	205
Table 12.29:	Item parameters for scales reflecting teachers' perceptions of classroom climate	206
Table 12.30:	Reliabilities for scales reflecting teachers' perceptions of classroom climate	209
Table 12.31:	Item parameters for scales reflecting teachers' perceptions of classroom climate	210
Table 12.32:	Reliabilities for scales reflecting principals' reports on school governance	213
Table 12.33:	Item parameters for scales reflecting principals' reports on school governance	214
Table 12.34:	Reliabilities for scales reflecting principals' reports on the local community	217
Table 12.35:	Item parameters for scales reflecting principals' reports on the local community	218
Table 12.36:	Reliabilities for scales reflecting principals' reports on school climate	221



Table 12.37: Item parameters for scales reflecting principals' reports on school climate	222
Table 12.38: Reliabilities for scales reflecting principals' reports on sense of belonging to school	223
Table 12.39: Item parameters for scales reflecting principals' reports on sense of belonging to school	224
Table 12.40: Reliabilities for scale reflecting students' perceptions of European identity	226
Table 12.41: Item parameters for scale reflecting students' perceptions of European identity	226
Table 12.42: Reliabilities for scales reflecting students' reports on activities related to Europe	228
Table 12.43: Item parameters for scales reflecting students' reports on activities related to Europe	229
Table 12.44: Reliabilities for scale reflecting students' attitudes toward learning other European languages	231
Table 12.45: Item parameters for scale reflecting students' attitudes toward learning other European languages	231
Table 12.46: Reliabilities for scales reflecting students' attitudes toward migration within Europe	233
Table 12.47: Item parameters for scales reflecting students' attitudes toward migration within Europe	234
Table 12.48: Reliabilities for scales reflecting students' attitudes toward European integration	236
Table 12.49: Item parameters for scales reflecting students' attitudes toward European integration	237
Table 12.50: Reliabilities for scale reflecting students' self-reported knowledge about the European Union	240
Table 12.51: Item parameters for scale reflecting students' self-reported knowledge about the European Union	240
Table 12.52: Reliabilities for scale reflecting students' perceptions of Latin American identity	241
Table 12.53: Item parameters for scale reflecting students' perceptions of Latin American identity	241
Table 12.54: Reliabilities for scales reflecting students' perceptions of government and law in Latin America	242
Table 12.55: Item parameters for scales reflecting students' perceptions of government and law in Latin America	243
Table 12.56: Reliabilities for scales related to Latin American students' perceptions of peaceful coexistence	246
Table 12.57: Item parameters for scales related to Latin American students' perceptions of peaceful coexistence	246
Table 12.58: Reliabilities for scale reflecting Latin American students' reports on frequency of discussions at school about civic issues	249
Table 12.59: Item parameters for scale reflecting Latin American students' reports on frequency of discussions at school about civic issues	249
Table 12.60: Reliabilities for scales reflecting students' perceptions of government and law in Asia	251
Table 12.61: Item parameters for scales reflecting students' perceptions of government and law in Asia	251



Table 12.62: Scale reliabilities for scales reflecting students' perceptions of identity, citizenship, and culture in Asia	253
Table 12.63: Item parameters for scales reflecting students' perceptions of identity, citizenship, and culture in Asia	254
Table 12.64: Reliabilities for scales reflecting students' perceptions of public service in Asia	256
Table 12.65: Item parameters for scales reflecting students' perceptions of public service in Asia	257
Table 13.1: Numbers of sampling zones in national samples	262
Table 13.2: Example of computation of replicate weights	264
Table 13.3: National averages for civic knowledge with standard deviations, sampling, and overall errors	266
Table 13.4: Coefficients of missing indicators in hierarchical linear analysis of ICCS data	271
Table 13.5: ICCS students included in hierarchical linear models	273
Table 13.6: Coefficients of missing indicators multiple regression analysis of expected electoral participation	275
Table 13.7: Coefficients of missing indicators multiple regression analysis of expected political participation	277
Table 13.8: ICCS students included in multiple regression models	279

#### *Figures*

Figure 3.1: Contexts for the development of learning outcomes related to civics and citizenship	35
Figure 6.1: Systematic PPS sampling of schools	66
Figure 8.1: Procedures for working with schools to prepare for test administration	92
Figure 10.1: Overview of data processing at the IEA DPC	123
Figure 11.1: Mapping of student abilities and item difficulties	130
Figure 11.2: Item characteristic curve by category for Item CI2MOM1	133
Figure 11.3: Item characteristic curve by category for Item CI2HRM2	133
Figure 11.4: Example of item-by-country interaction graph for Item CI2HRM2	137
Figure 11.5: Percentages of not-reached responses for groups of countries for Test Booklet 1	139
Figure 11.6: Scatterplot for link item parameter estimates from CIVED and for the ICCS trend sample	150
Figure 12.1: Summed category probabilities for fictitious item	164
Figure 12.2: Example of questionnaire item-by-score map	165
Figure 12.3: Confirmatory factor analysis of items measuring students' behaviors	168
Figure 12.4: Confirmatory factor analysis of items measuring students' perceptions of the school context	172
Figure 12.5: Confirmatory factor analysis of items measuring students' democratic values	174
Figure 12.6: Confirmatory factor analysis of items measuring students' perceptions of good citizenship	176



Figure 12.7:	Confirmatory factor analysis of items measuring students' civic-related self-beliefs	180
Figure 12.8:	Confirmatory factor analysis of items measuring students' attitudes toward equal rights	183
Figure 12.9:	Confirmatory factor analysis of items measuring students' attitudes toward civic institutions and their country	186
Figure 12.10:	Confirmatory factor analysis of items measuring students' expected participation in protest activities	188
Figure 12.11:	Confirmatory factor analysis of items measuring students' expected political participation	191
Figure 12.12:	Confirmatory factor analysis of items measuring students' attitudes toward the influence of religion in society	193
Figure 12.13:	Confirmatory factor analysis of items measuring teachers' perceptions of school governance	197
Figure 12.14:	Confirmatory factor analysis of items measuring teachers' perceptions of teaching in classrooms	201
Figure 12.15:	Confirmatory factor analysis of items measuring teachers' perceptions of participation in the community	204
Figure 12.16:	Confirmatory factor analysis of items measuring teachers' perceptions of school climate	207
Figure 12.17:	Confirmatory factor analysis of items reflecting teachers' reports on civic and citizenship education	211
Figure 12.18:	Confirmatory factor analysis of items reflecting teachers' reports on school governance	215
Figure 12.19:	Confirmatory factor analysis of items measuring principals' perceptions of the local community	219
Figure 12.20:	Confirmatory factor analysis of items measuring principals' perceptions of school climate	222
Figure 12.21:	Confirmatory factor analysis of items measuring principals' perceptions of sense of belonging to school	225
Figure 12.22:	Confirmatory factor analysis of items measuring students' sense of European identity	227
Figure 12.23:	Confirmatory factor analysis of items reflecting students' reports on activities related to Europe	230
Figure 12.24:	Confirmatory factor analysis of items measuring students' attitudes toward learning European languages	232
Figure 12.25:	Confirmatory factor analysis of items measuring students' attitudes toward migration within Europe	235
Figure 12.26:	Confirmatory factor analysis of items measuring students' attitudes toward European integration	239
Figure 12.27:	Confirmatory factor analysis of items measuring students' self-reported knowledge about the European Union	241
Figure 12.28:	Confirmatory factor analysis of items measuring students' sense of Latin American identity	242
Figure 12.29:	Confirmatory factor analysis of items measuring students' attitudes toward government and law in Latin America	245
Figure 12.30:	Confirmatory factor analysis of items measuring Latin American students' perceptions of peaceful coexistence	248



Figure 12.31: Confirmatory factor analysis of items measuring Latin American students' reports on frequency of discussions at school about civic issues	250
Figure 12.32: Confirmatory factor analysis of items measuring students' perceptions of government and law in Asia	252
Figure 12.33: Confirmatory factor analysis of items measuring students' perceptions of identity, citizenship, and culture in Asia	255
Figure 12.34: Confirmatory factor analysis of items measuring students' perceptions of public service in Asia	258

#### *Tables appendices*

Table B.1.1: Allocation of student sample in Austria	286
Table B.1.2: Allocation of teacher sample in Austria	286
Table B.2.1: Allocation of student sample in Belgium (Flemish)	287
Table B.2.2: Allocation of teacher sample in Belgium (Flemish)	287
Table B.3: Allocation of student sample and teacher sample in Bulgaria	288
Table B.4: Allocation of student sample and teacher sample in Chile	288
Table B.5.1: Allocation of student sample in Chinese Taipei	289
Table B.5.2: Allocation of teacher sample in Chinese Taipei	289
Table B.6.1: Allocation of student sample in Colombia	290
Table B.6.2: Allocation of teacher sample in Colombia	290
Table B.7.1: Allocation of student sample in Cyprus	291
Table B.7.2: Allocation of teacher sample in Cyprus	291
Table B.8.1: Allocation of student sample in the Czech Republic	291
Table B.8.2: Allocation of teacher sample in the Czech Republic	292
Table B.9.1: Allocation of student sample in Denmark	292
Table B.9.2: Allocation of teacher sample in Denmark	292
Table B.10: Allocation of student sample and teacher sample in the Dominican Republic	293
Table B.11.1: Allocation of student sample in England	293
Table B.11.2: Allocation of teacher sample in England	293
Table B.12.1: Allocation of student sample in Estonia	294
Table B.12.2: Allocation of teacher sample in Estonia	294
Table B.13.1: Allocation of student sample in Finland	295
Table B.13.2: Allocation of teacher sample in Finland	295
Table B.14: Allocation of student sample in Greece	296
Table B.15: Allocation of student sample and teacher sample in Guatemala	296
Table B.16.1: Allocation of student sample in Hong Kong, SAR	297
Table B.16.2: Allocation of teacher sample in Hong Kong, SAR	297
Table B.17.1: Allocation of student sample in Indonesia	297
Table B.17.2: Allocation of teacher sample in Indonesia	298
Table B.18.1: Allocation of student sample in Ireland	298
Table B.18.2: Allocation of teacher sample in Ireland	299
Table B.19.1: Allocation of student sample in Italy	299
Table B.19.2: Allocation of teacher sample in Italy	299
Table B.20.1: Allocation of student sample in Korea	300
Table B.20.2: Allocation of teacher sample in Korea	300





Table B.21.1:	Allocation of student sample in Latvia	301
Table B.21.2:	Allocation of teacher sample in Latvia	301
Table B.22:	Allocation of student sample and teacher sample in Liechtenstein	301
Table B.23:	Allocation of student sample and teacher sample in Lithuania	302
Table B.24.1:	Allocation of student sample in Luxembourg	302
Table B.24.2:	Allocation of teacher sample in Luxembourg	303
Table B.25:	Allocation of student sample and teacher sample in Malta	303
Table B.26.1:	Allocation of student sample in Mexico	303
Table B.26.2:	Allocation of teacher sample in Mexico	304
Table B.27.1:	Allocation of student sample in the Netherlands	304
Table B.27.2:	Allocation of teacher sample in the Netherlands	304
Table B.28.1:	Allocation of student sample in New Zealand	305
Table B. 28.2:	Allocation of teacher sample in New Zealand	305
Table B. 29.1:	Allocation of student sample in Norway	306
Table B.29.2:	Allocation of teacher sample in Norway	306
Table B.30.1:	Allocation of student sample in Paraguay	306
Table B.30.2:	Allocation of teacher sample in Paraguay	307
Table B.31:	Allocation of student sample and teacher sample in Poland	307
Table B.32:	Allocation of student sample and teacher sample in the Russian Federation	308
Table B.33.1:	Allocation of student sample in the Slovak Republic	309
Table B.33.2:	Allocation of teacher sample in the Slovak Republic	309
Table B.34.1:	Allocation of student sample in Slovenia	310
Table B.34.2:	Allocation of teacher sample in Slovenia	310
Table B.35:	Allocation of student sample and teacher sample in Spain	311
Table B.36.1:	Allocation of student sample in Sweden	312
Table B.36.2:	Allocation of teacher sample in Sweden	312
Table B.37.1:	Allocation of student sample in Switzerland	313
Table B.37.2:	Allocation of teacher sample in Switzerland	313
Table B.38:	Allocation of student sample and teacher sample in Thailand	314
Table C.1:	Descriptions for cognitive test items and allocations to proficiency levels	315
Table D.1:	List of international and regional student-level conditioning variables for ICCS 2009	319
Table D.2:	Years of parental education corresponding to ISCED levels per country	329
Table D.3:	Years of further schooling beyond target grade corresponding to ISCED levels per country	330
Table D.4:	Transformation parameters for international and regional questionnaire scales (means and standard deviations of original IRT scores)	331



## CHAPTER 1:

# Overview of the IEA International Civic and Citizenship Education Study

*John Ainley and Wolfram Schulz*

## Introduction

The International Civic and Citizenship Education Study (ICCS) studied the ways in which young people are prepared to undertake their roles as citizens. ICCS was based on the premise that preparing students for citizenship roles involves developing relevant knowledge and understanding as well as helping them form positive attitudes toward being a citizen and participating in activities related to civic and citizenship education. It also examined differences among countries in relation to these outcomes of civic and citizenship education, and it explored how differences among countries relate to student characteristics, school and community contexts, and national characteristics.

ICCS builds on the previous IEA (International Association for the Evaluation of Educational Achievement) studies of civic education and is a response to the challenge of educating young people for civic participation in the 21st century. The first IEA study of civic education was conducted as part of the Six Subject Study, with data collected in 1971 (Torney, Oppenheim, & Farnen, 1975). The second study, the IEA Civic Education Study (CIVED), was carried out in 1999 (Torney-Purta, Lehmann, Oswald, & Schulz, 2001; Torney-Purta, Schwillie, & Amadeo, 1999). An additional survey, of upper secondary students, was undertaken in 2000 (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002).

Results from ICCS have been reported in a brief report of first findings (Schulz, Ainley, Fraillon, Kerr, & Losito, 2010a) and in an international report (Schulz, Ainley, Fraillon, Kerr, & Losito, 2010b). These reports document differences among countries in relation to a wide range of civic-related learning outcomes and civic-related dispositions and behaviors. They also document differences in the relationships between those outcomes, student characteristics, school contexts, and characteristics of countries. ICCS results have also been reported in a series of regional reports concerned with Europe (Kerr, Sturman, Schulz, & Burge, 2010), Latin America (Schulz, Ainley, Friedman, & Lietz, 2011), and Asia (Fraillon, Ainley, & Schulz, forthcoming). This technical report provides information about the data and analytic procedures that provided the basis for those reports.

## General design

ICCS was based around six research questions concerned with (1) variations in civic knowledge, (2) changes in content knowledge since 1999, (3) the interest and dispositions of students to engage in public and political life, (4) perceptions of threats to civil society, (5) features of education systems, schools, and classrooms that are related to civic and citizenship education, and (6) aspects of student background associated with the outcomes of civic and citizenship education. The design of ICCS was elaborated in the ICCS assessment framework (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008).

ICCS was a comparative international survey of students in their eighth year of school (usually Grade 8).<sup>1</sup> In addition to collecting and analyzing data on student outcomes, the ICCS research team gathered information by conducting surveys of samples of teachers and the principals in the participating schools. These survey data were complemented by information about the national contexts for civic and citizenship education gathered by the national research centers of the participating countries.



1 Provided that the average age of students in Grade 8 was 13.5 years or above at the time of the assessment.



ICCS collected data from more than 140,000 Grade 8 (or equivalent) students in more than 5,300 schools from 38 countries. These student data were augmented with data from more than 62,000 teachers in those schools and further contextual data collected from school principals and national research centers.

## ICCS assessment of knowledge and perceptions

Unlike learning and teaching in areas such as mother tongue, mathematics, and science, civics and citizenship rarely has a well-defined place in a core curriculum. The ICCS assessment framework provided a conceptual underpinning for the international instrumentation for ICCS and a point of reference for the development of regional instruments. Within this overall framework, a civics and citizenship framework outlined the outcome measures addressed through the cognitive test and the student perceptions questionnaire.

The civics and citizenship framework consisted of the following:

- A content dimension specifying the subject matter to be assessed within civics and citizenship (with regard to both affective-behavioral and cognitive aspects);
- An affective-behavioral dimension which described the student perceptions and activities that were measured; and
- A cognitive dimension that described the thinking processes assessed.

The four content domains in the ICCS assessment framework were *civic society and systems*,<sup>2</sup> *civic principles*,<sup>3</sup> *civic participation*,<sup>4</sup> and *civic identities*.<sup>5</sup> Each of these content domains was made up of a set of subdomains that incorporated elements referred to as aspects and key concepts.

Student perceptions and behaviors relevant to civics and citizenship were drawn from four affective-behavioral domains: *value beliefs*, *attitudes*, *behavioral intentions*, and *behaviors*. Value beliefs relate to fundamental beliefs about democracy and citizenship. They are, relative to attitudes, more constant over time, more deeply rooted, and broader. Attitudes include self-cognitions related to civics and citizenship, attitudes toward the rights and responsibilities of groups in society, and attitudes toward institutions. Behavioral intentions refer to expectations of future civic action, and include constructs such as preparedness to participate in forms of civic protest, anticipated future political participation as adults, and anticipated future participation in citizenship activities. Behaviors reflect present or past participation in civic-related activities at school or in the wider community.

The two cognitive processes included in the ICCS framework were *knowing* and *reasoning and analyzing*. Knowing refers to the learned civic and citizenship information that students use when engaging in cognitive tasks that help them to make sense of their civic worlds. Reasoning and analyzing refers to the ways in which students use civic and citizenship information to reach conclusions. These ways typically involve integrating perspectives that apply to more than a single concept and are applicable in a range of contexts.

2 Civic society and systems consists of three subdomains: *citizens* (roles, rights, responsibilities, and opportunities), *state institutions* (central to civic governance and legislation), and *civil institutions* (these mediate citizens' contact with state institutions and allow citizens to pursue many of their roles in their societies).

3 Civic principles consists of three subdomains: *equity* (all people having the right to fair and just treatment), *freedom* (of belief and speech, and from fear and want), and *social cohesion* (sense of belonging, connectedness, and common vision held by individuals and communities within a society).

4 Civic participation consists of three subdomains: *decision-making* (organizational governance and voting), *influencing* (debating, demonstrating, developing proposals, and selective purchasing), and *community participation* (volunteering, participating in organizations, keeping informed).

5 Civic identities consists of two subdomains: *civic self-image* (individuals' experience of place in each of their civic communities) and *civic connectedness* (sense of connection to different civic communities; refers also to the civic roles individuals play within each community).



## Contexts for civic and citizenship education

ICCS took account of the context in which civic learning takes place. Civic and citizenship outcomes are influenced by students' wider communities (local, regional, national, supra-national), schools and classrooms (the instruction provided, the school culture experienced, and the general school environment), home environments (the direct home background and the social out-of-school environment), and individual characteristics (those that shape the way students respond to learning about civics and citizenship).

ICCS gathered information about national contexts because the ways students develop civic-related dispositions and competencies and acquire understandings with regard to their role as citizens are influenced by country-level factors. Interpreting the results from an international assessment of civic and citizenship education requires taking into account historical background, the political system, the structure of education, and the curriculum. The national context survey was designed to systematically collect relevant data on the structure of the education system, education policy, civic and citizenship education, teacher qualifications for civic and citizenship education, and the extent of current debates and reforms in this area. The survey also collected data on process at the national level regarding assessment of and quality assurance not only with respect to civic and citizenship education but also with respect to school curriculum approaches.

## Instruments

Several instruments were administered as part of the ICCS. Those completed by students focused on the outcomes of civic and citizenship education and on student background (the individual context) and included the following:

- *The international student cognitive test*: this consisted of 80 items measuring civic and citizenship knowledge, analysis, and reasoning. The assessment items were assigned to seven booklets (each of which contained three of a total seven item-clusters) according to a balanced rotated design. Each student completed one of the 45-minute booklets. The cognitive items were generally presented with contextual material that served as a brief introduction to each item or set of items. ICCS incorporated a cluster of test items that had been used in the IEA CIVED study in 1999. Fifteen of these 17 items performed in a comparable way on both occasions and therefore were used to assess change on the CIVED subdomain of *civic content knowledge*.
- *A 45-minute international student questionnaire*: this was used to obtain student perceptions about civics and citizenship as well as information about each student's background.
- *A set of regional instruments*: these were directed toward particular issues associated with civics and citizenship in Asia, Europe, and Latin America. Each instrument took students between 15 and 30 minutes to complete. The Asian regional instrument was a 15-minute questionnaire. The European regional instrument consisted of a 12-minute region-specific cognitive test and a 17-minute region-specific questionnaire (29 minutes total). The Latin American regional instrument consisted of a 15-minute region-specific cognitive test and a 15-minute region-specific questionnaire (30 minutes total).

ICCS also included a set of instruments designed to gather information from and about teachers, schools, and education systems. These included:

- *A 30-minute teacher questionnaire*: respondents provided information about their perceptions of civic and citizenship education in their schools and their schools' organization and culture as well their own teaching assignments and backgrounds.



- *A 30-minute school questionnaire:* principals provided information about school characteristics, school culture and climate, and the provision of civic and citizenship education in the school.

National research coordinators (NRCs) coordinated the information procured from national experts in response to an online national contexts survey. This information covered the structure of the education system, civic and citizenship education in the national curricula, and recent developments in civic and citizenship education.

## Measures

Student scores on the cognitive test were derived, via the Rasch model (Rasch, 1960), from 79 (of the original 80) items. Summary student achievement statistics were derived using plausible value methodology with full conditioning, in which five separate estimates were generated for each student. The final reporting scale was set to a metric with a mean of 500 and a standard deviation of 100 for the equally weighted national samples.

The ICCS-described achievement scale was developed in line with the contents and scaled difficulties of the assessment items. An analysis of the item map and student achievement data led to the establishment of proficiency levels with a width of 84 scale points and boundaries at 395, 479, and 563 scale points. Students who scored below 395 scale points were deemed to have civic and citizenship knowledge below the level targeted by the assessment instrument.

The proficiency-level descriptions are syntheses of the item descriptors within each level.<sup>6</sup> They describe a hierarchy of civic knowledge in terms of increasing sophistication of content knowledge and cognitive process. The scale reflects a broad range of development, from dealing with concrete, familiar, and mechanistic elements of civics and citizenship through to the wider policy and institutional processes that determine the shape of our civic communities.

Seventeen of the ICCS countries that took part in IEA CIVED collected data from comparable samples on both occasions, making it possible to measure change in *civic content knowledge*. ICCS test data for the CIVED link items were scaled using the same item parameters as in CIVED and then transformed to the metric used in CIVED to report *civic content knowledge*. This metric had an average of 100 and a standard deviation of 20 scale points for the equally weighted 28 countries participating in CIVED.

Responses to the questionnaires were either reported as scale scores based on sets of theoretically and empirically related items or (in a few cases) as discrete item responses. The scale scores were based on IRT methods and calculated as weighted likelihood estimates. For reporting purposes, the scale scores were transformed to a mean of 50 (for equally weighted national samples) and a standard deviation of 10.



6 Proficiency Level 1 is characterized by engagement with the fundamental principles and broad concepts that underpin civics and citizenship and by a mechanistic working knowledge of the operation of civic, civil, and political institutions. Proficiency Level 2 is characterized by knowledge and understanding of the main civic and citizenship institutions, systems, and concepts as well as by an understanding of the interconnectedness of civic and civil institutions and relevant operational processes. Proficiency Level 3 is characterized by the ability to apply knowledge and understanding in order to evaluate or justify policies, practices, and behaviors related to civics and citizenship.

## Populations and samples

The ICCS student population comprised students in Grade 8 (i.e., students who, on average, are approximately 14 years of age), provided that the average age of students in this grade at the time of testing was 13.5 years or above.

The samples were designed as two-stage cluster samples. In the first stage of sampling, schools were sampled within each country using PPS (probability proportional to size as measured by the number of students enrolled in a school). The numbers required in the sample to achieve the necessary precision were estimated on the basis of national characteristics. However, as a guide, a minimum sample size of 150 schools was planned in each country. Within each sampled school, an intact class from the target grade was sampled randomly, and all students in that class were surveyed.

The population for the ICCS teacher survey was defined as all teachers teaching regular school subjects to the students in the target grade (mostly Grade 8) at each sampled school. Up to 15 teachers were selected at random from all teachers teaching the target grade at each sampled school. In schools with 20 or fewer such teachers, all teachers were invited to participate. In schools with more than 20 teachers, 15 of those teachers were sampled at random.

## Outline of this report

This report is structured so as to provide technical detail about each aspect of ICCS. This overview is followed by a series of chapters that provide detail about different aspects of ICCS.

Chapters 2, 3, 4, and 5 are concerned with the instruments. Chapter 2 provides information about the development and properties of the ICCS cognitive test. Chapter 3 details the development and properties of the ICCS international questionnaires. These include the international student questionnaire, the teacher questionnaire, and the school questionnaire, as well as the national contexts survey. Chapter 4 is concerned with the development of the ICCS regional instruments: the European test and questionnaire, the Latin American test and questionnaire, and the Asian questionnaire. Chapter 5 describes the procedures used to translate and adapt ICCS instruments.

Chapters 6 and 7 are concerned with aspects of sampling. Chapter 6 details the sampling design and implementation and Chapter 7 documents the weighting procedures that were used to ensure the results from ICCS represented the defined populations in each country.

Chapters 8, 9, and 10 focus on the survey implementation. Chapter 8 details the field operation procedures and the process of preparation of data files. Chapter 9 documents the quality control protocols and procedures that were used in the ICCS data collection. Chapter 10 provides an account of data management in ICCS and the creation of the ICCS database.

Chapters 11, 12, and 13 describe the psychometric and statistical analyses used in ICCS. Chapter 11 reports on the scaling procedures for the ICCS cognitive test and how responses to the test items were used to construct scores on the ICCS civic knowledge scale. Chapter 12 describes the methods used to form scales from the ICCS questionnaire items, while Chapter 13 details how the ICCS results were reported and gives an account of the conventions adopted for the construction of tables.

The ICCS technical report also contains a set of appendices. These list the organizations and individuals involved in ICCS, describe the characteristics of the national samples, provide descriptions of the cognitive test items (including allocations of those items to proficiency levels), and include tables featuring the coding information for the items in the questionnaires and the test.



## References

- Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper-secondary students in sixteen countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Fraillon, J., Ainley, J., & Schulz, W. (forthcoming). *ICCS 2009 Asian report: Civic knowledge and attitudes among lower-secondary students in five Asian countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Kerr, D., Sturman, L., Schulz, W., & Burge, B. (2010). *ICCS 2009 European report: Civic knowledge, attitudes, and engagement among lower-secondary students in 24 European countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010a). *Initial findings from the IEA International Civic and Citizenship Study*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010b). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Friedman, T., & Lietz, P. (2011). *ICCS 2009 Latin American report: Civic knowledge and attitudes among lower-secondary students in six Latin American countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International civic and citizenship education study: Assessment framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Torney, J., Oppenheim, A. N., & Farnen, R. F. (1975). *Civic education in ten countries: An empirical study*. New York: John Wiley and Sons.
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Torney-Purta, J., Schwille, J., & Amadeo, J. A. (1999). *Civic education across countries: Twenty-four national case studies for the IEA Civic Education Project*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).



# ICCS test development

Julian Fraillon

## Introduction

The ICCS civic knowledge assessment was developed over an 18-month period from October 2006 to April 2008. Most of the ICCS test-item development was conducted by the international study center (ISC) at the Australian Council for Educational Research (ACER) in collaboration with the study's national research coordinators (NRCs) and the Project Advisory Committee (PAC).

This chapter provides a detailed description of the test-development process, review procedures, and the test design implemented for the ICCS field trial and main survey. Table 2.1 provides an overview of the test-development processes and timeline.

## Test scope and format

### *ICCS assessment framework*

The cognitive test items for this study were developed with reference to the ICCS assessment framework (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008) and designed to measure a single trait labeled civic knowledge in the international reports on ICCS (Schulz, Ainley, Fraillon, Kerr, & Losito, 2010a, b). The manner in which civic knowledge was expressed through the ICCS test items required students to apply the cognitive processes to the civics and citizenship content as described in the assessment framework.

Each test item developed for ICCS was mapped to both the civics and citizenship content and the cognitive process that students required to respond correctly to the item. The assessment framework was designed to subsume and broaden the conceptual model underpinning IEA's 1999 Civic Education Study (CIVED) test items (see Schulz et al., 2008, pp. 12–13; Torney-Purta, Lehmann, Oswald, & Schulz, 2001, pp. 20–22), thereby making it possible to map the CIVED secure trend items to the ICCS assessment framework domains.

The ICCS assessment framework includes four content and two cognitive domains. The four content domains are:

- Civic society and systems;
- Civic principles;
- Civic participation;
- Civic identities.

The two cognitive domains are:

- Knowing;
- Reasoning and analyzing.

### *Test-item descriptions*

The test items were presented in units consisting, in most cases, of some form of stimulus material (such as text or an image) followed by one or more items relating to the context established by the stimulus. On average, there were 1.4 items per unit in the main survey test instrument.

Two item formats were used: the majority (approximately 93%) of test items had a multiple-choice format with four response options. The remaining items were constructed-response items for which students were required to write between one and three sentences.





Table 2.1: Test development processes and timeline

Date	Group and activity
October 2006	<i>ICCS International Study Center</i> Drafting of preliminary item ideas
October 2006	<i>Project Advisory Committee (Amsterdam)</i> Review of proposed test development process and preliminary item ideas <i>First meeting of NRCs (Amsterdam)</i> Item development workshop
November 2006	<i>ICCS International Study Center</i> Drafting, review and refinement of test items Call for item submissions by NRCs and PAC members
March 2007	<i>National Research Coordinators and Project Advisory Committee</i> Web-based item review
March 2007	<i>ICCS International Study Center</i> Piloting of draft test items in English with a small convenience sample of students in Australia, Colombia, England, Italy, the Netherlands, and New Zealand.
April 2007	<i>ICCS International Study Center</i> Item revision based on web-based review
June 2007	<i>Project Advisory Committee (Rome)</i> Review of items proposed for inclusion in field trial test and confirmation of test design <i>Second meeting of NRCs (Rome)</i> Review of items proposed for inclusion in field trial test and confirmation of test design
July 2007	<i>ICCS Scoring Trainers</i> Review of field trial scoring guides for constructed-response items (as part of scoring training)
July 2007	<i>ICCS International Study Center</i> Revision of field trial scoring guides for constructed-response items
March 2008	<i>ICCS International Study Center</i> Analysis of field trial item data and recommendations for items to be included in main survey test (field trial analysis report)
June 2008	<i>Project Advisory Committee (Windsor)</i> Review of field trial analysis report and recommendations for test design and items proposed for inclusion in main survey <i>Third meeting of NRCs (Windsor)</i> Review of field trial analysis report and recommendations for test design and items proposed for inclusion in main survey
November 2008	<i>ICCS Scoring Trainers</i> Review of main survey scoring guides for constructed-response items (as part of scoring training)
November 2008	<i>ICCS International Study Center</i> Revision of main survey scoring guides for constructed-response items



The ICCS test of civic knowledge included a link to the CIVED survey in 1999 through the inclusion of secure items from the CIVED item pool. The inclusion of these 17 multiple-choice items made it possible to measure changes in performance for countries that participated in both ICCS and CIVED and had assessed comparable student populations.

## Test-development process

The cognitive test-item and instrument development process consisted of a series of stages. These stages followed one another sequentially. However, the iterative and collaborative nature of the overall process meant that some materials were reviewed and revised within particular stages more than once.

### *NRC test-item development workshop*

The item development process began formally with an item development workshop at the first meeting of NRCs in October 2006 in Amsterdam. At this workshop, national representatives were provided with information about the framework and procedures for ICCS test development. Participants drafted items in small working groups. The workshop involved the following activities:

- A review of the content of the assessment framework to ensure a common understanding of the fundamental civics and citizenship constructs;
- A mapping of the CIVED trend items against the assessment framework to guide the development of new items;
- Confirmation of the necessary properties of test-item stimuli, including issues relating to ensuring cultural sensitivity and avoiding potential biases (such as cultural or gender bias);
- Confirmation of item types, scoring-guide formats, and test-development systems, including the online item-review process;
- An introduction to the principles of cognitive test-item development;
- The opportunity to discuss and consider any cognitive test items that NRCs had brought to the workshop;
- Test-item development in small groups; and
- An invitation for NRCs to make arrangements for cognitive test items to be developed and submitted to the ISC for consideration.

### *Stimulus selection and preliminary item development*

The focus of this preliminary stage of item development was on establishing authentic, viable, and relevant contexts for items to assess the content specified in the assessment framework. Stimulus materials, contexts, and item ideas were developed internally at the ISC. NRCs and members of the PAC submitted a small number of contexts and ideas, some of which had been created at the item-development workshop.

These materials were submitted to the ISC, where the test-development team assessed them for their suitability for further development. This work included evaluating the extent to which these materials were appropriate for civic and citizenship assessment purposes and their suitability for the target student population. The team also reviewed these materials with respect to the range of contexts and themes that they covered.

Materials selected for further development were subsequently refined (as required), and the related items were developed by a test developer responsible for a particular *unit* (stimulus, items, and scoring guides). Once the project researchers had developed their respective units to the degree that they considered them to be complete, they submitted them to a quality control process called “paneling.”





### *Paneling*

Paneling is a team-based approach to reviewing assessment materials. This rigorous quality-control mechanism is employed during the development of assessment materials. Paneling is a process that recognizes the importance of exposing material to multiple viewpoints. During this process, a small group (between three and six) of test developers jointly review material that one or more of them has developed. The review leads to acceptance, modification, or rejection.

Panel participants compare their answers to the questions and raise issues about the questions and the material. Discussion is robust because of the need to ensure that the selected items perform as intended.

The following questions provide a summary of the issues that formed the focus of the evaluation of the item material developed for ICCS. The relevance of each evaluation issue varied according to the individual characteristics of the material under consideration.

### **Content validity**

- How did the material relate to the ICCS test specifications?
- Did the questions test the content and cognitive processes described in the assessment framework?
- Did the questions relate to the essence of the stimulus or did they focus on trivial side-issues?
- How would this material stand up to public scrutiny (including staff involved in the project as well as members of the wider community)?

### **Clarity and context**

- Was the material coherent, unambiguous, and clear?
- Was the material interesting, worthwhile, and relevant?
- Did the material assume prior knowledge and, if so, was this assumed to be acceptable or part of what the test intended to measure?
- Was the reading load as low as possible?
- Were there idioms or syntactical structures likely to prove difficult to translate into other languages?

### **Format**

- Was the proposed format the most suitable for the content and process being assessed by the item?
- Was the key (the correct answer to a multiple-choice question) indisputably correct?
- Were the distractors (the incorrect options to a multiple-choice question) plausible but also irrefutably incorrect?

### **Test-takers**

- Did the test-item material match the expected range of ability levels, age, and maturity of the ICCS target population?
- Did the material appear to be cross-culturally relevant and sensitive?
- Were items likely to be easier or harder for certain subgroups in the target population for reasons other than differences in the ability measured by the test?
- Did the constructed-response items provide clear guidance as to the expected answers to the test question?



### Scoring

- Was the proposed scoring consistent with the underlying ability measured by the test and would test respondents with higher ability levels always score better than those with lower ones?
- Were there different possible student responses that might receive the same score, and did these responses represent equivalent or different levels of proficiency?
- Were there other kinds of answers that had not been anticipated in the marking guide (e.g., any that did not fall within the “correct” answer category description but appeared to be equally correct)?
- Were the scoring criteria sufficiently clear for coders to allow them to distinguish the different levels of performance?

The reviews and evaluations conducted during the paneling process provided the participating test developers with extensive notes on each stimulus piece, item, and scoring guide (for the constructed-response items). The item material deemed appropriate for further development was subsequently refined on the basis of the panel’s feedback.

### *Refinement of item material*

During the process of refinement, all revised materials were shown to at least one test developer who had not previously seen them. The purpose of this additional check was to ensure that the revision of items had not created additional technical problems.

### *External review*

All ICCS draft test material (stimulus items and scoring guides) were placed on the ICCS website for review by members of the PAC, NRCs, and other consultant experts. The web-based review process allowed reviewers to post comments on each component of the materials and to complete a brief rating (1 to 4) of the suitability of the materials for inclusion in the test. The rating categories were:

- 4: Include item without change;
- 3: Include item, changes recommended;
- 2: Include item only if changes have been made;
- 1: Do not include item.

The web-based external review took place in March and April 2007, after which the test draft material was further revised in accordance with the feedback arising out of the review.

### *Piloting*

The draft test items were piloted at the same time as the web-based item review. Two short test forms, each with 42 items, were created using the draft item material and delivered to smaller convenience samples of students in Australia, Colombia, England, Italy, the Netherlands, and New Zealand. Overall, 436 students participated in the pilot study. In addition to completing the test items, Australian students participated in small group discussions that centered on their experiences when completing the piloted test items. Students were asked to comment on their perceptions of the difficulty of the items, the wording of the items (with a focus on any ambiguities or difficulty they had in terms of understanding each item), and the content (context as well as the civics and citizenship content) of the items.

The results from the pilot study provided a first empirical basis for further development of the ICCS test material. The student responses to the open-ended test items in the pilot were particularly useful for informing the further development of both items and scoring guides.



### *Development of constructed-response scoring guides*

The scoring guides are essential parts of the constructed-response test items. The scoring guides were drafted and refined using the same processes as the test items. These processes were also informed by the student responses from the pilot study.

An international training session for scorers was conducted before both the field trial and the main survey. National center representatives who attended these meetings subsequently trained the national center staff in charge of scoring student responses in their respective countries. The field trial scorer training was the first opportunity that country representatives had to meet and discuss the scoring guides with ISC staff. Feedback from these training sessions was used to further refine the scoring guides.

The scoring guides for the field trial included a “dummy scoring code,” which allowed for student responses that appeared to be worthy of credit but were not clearly accounted for by the scoring guides. National center staff communicated the nature of these student responses to the ISC in order to inform the ongoing development of the scoring guides in preparation for the main survey. Some additional valid scoring categories were developed on the basis of the experience during the field trial. ISC staff reviewed and discussed the layout and description of these categories with country representatives at the scorer training for the main survey. The final scoring guides for the main survey items were distributed after completion of the international main survey scorer training.

## **Field trial test design and content**

### *Test design*

The field trial test consisted of 98 items, including 19 items that were secure trend items from the IEA CIVED study. These items were included to facilitate comparisons in achievement between CIVED and ICCS for those countries that participated in both studies and had assessed comparable student populations. Table 2.2 shows the composition of the field trial test instrument by item type and item origin.

*Table 2.2: Composition of the field trial test instrument by item type and origin*

Item format	New items	CIVED items	Total	Total score points	Percentage of score points
Multiple choice	65	19	84	84	82
True/False	6	–	6	2	2
Constructed response	8	–	8	16	16
Total	79	19	98	102	100

The items were allocated to six clusters that were presented in a fully balanced rotated test design across six test booklets. One cluster of items comprised the set of secure CIVED items. Students had 60 minutes to answer each booklet. Table 2.3 shows the cluster composition. Table 2.4 shows the rotated cluster test design used in the field trial. The shaded part of Table 2.4 denotes the CIVED trend-item cluster.



Table 2.3: Field trial cluster composition

Cluster	Items
C1	16 (15 MCQ, 1 CRQ)
C2	18 (11 MCQ, 6 T/F, 1 CRQ)
C3	16 (15 MCQ, 1 CRQ)
C4	15 (14 MCQ, 2 CRQ)
C5	14 (12 MCQ, 2 CRQ)
C6 (CIVED)	19 MCQ

**Note:** MCQ = multiple-choice question; CRQ = constructed-response question; T/F = True/False.

Table 2.4: Field trial test booklet design

Booklet	Position		
	1	2	3
1	C1	C2	C4
2	C2	C3	C5
3	C3	C4	C6
4	C4	C5	C1
5	C5	C6	C2
6	C6	C1	C3

**Note:** CIVED link cluster shaded in grey.

### Coverage of the assessment framework

The newly developed ICCS test items and CIVED trend items were all mapped to the ICCS assessment framework. Table 2.5 shows the original test development plan for ICCS test content according to the ICCS assessment framework content and cognitive domains. Table 2.6 shows the mapping of the final set of field trial test items to the ICCS assessment framework.

Table 2.5: Planned field trial item mapping to assessment framework

Content domain	New items	CIVED items	Total	Percentage of total items
Civic society and systems	25	15	40	41
Civic principles	28	2	30	31
Civic participation	21	–	21	21
Civic identities	5	2	7	7
<b>Total</b>	<b>79</b>	<b>19</b>	<b>98</b>	<b>100</b>
<b>Cognitive domain</b>				
Knowing	13	14	27	28
Reasoning and analyzing	66	5	71	72
<b>Total</b>	<b>79</b>	<b>19</b>	<b>98</b>	<b>100</b>



Table 2.6: Field trial item mapping to assessment framework

Content domain	Percentage of all items
Civic society and systems	40
Civic principles	30
Civic participation	20
Civic identities	10
<b>Total</b>	<b>100</b>
<b>Cognitive domain</b>	
Knowing	30
Reasoning and analyzing	70
<b>Total</b>	<b>100</b>

Because the ICCS framework was broader than the framework for CIVED, most of the CIVED items were mapped to the content domain *civic society and systems* and the cognitive domain *knowing*. The newly developed ICCS items were written to complement the framework coverage of the CIVED items and consequently they refer mainly to the content domains of *civic principles*, *civic participation*, and *civic identities*, and the cognitive domain *reasoning and analyzing*.

The decision to have relatively low proportions of items addressing the content domains *civic participation* and *civic identities* was made because little of this kind of content could reasonably be asked of students in this age group. These two content domains were given a stronger focus in the student questionnaire. Tables 2.5 and 2.6 show that the originally anticipated and final proportions of the field trial items provided good coverage of the assessment framework's content domains.

## Main survey

### *Selection of items*

Evaluation of the measurement properties of the field trial test items was based on the data collected from students in 32 participating countries. The analysis procedures used to review measurement properties are described in Chapter 11. The items with unsatisfactory measurement properties were further reviewed to determine whether they could be revised or whether they needed to be deleted from the item set. Minor revisions were made to 25 items. Items were only modified when there was clear evidence that the revision would improve their measurement properties. Eighteen items were removed from the ICCS test item pool, leaving 80 items for inclusion in the main survey.

### *Test design and content*

The main survey test consisted of 80 items, including 17 items that were secure items from the IEA CIVED study. Table 2.7 shows the composition of the main study test instrument by item type and item origin.

Based on experience with the field trial, the test development team decided to reduce the testing time from 60 to 45 minutes. This decision meant that the number of items within each cluster had to be reduced and the number of booklets increased.



Table 2.7: Composition of the main study test instrument by item type and origin

Item format	New items	CIVED items	Total	Total score points of	Percentage of score points
Multiple choice	57	17	74	74	86
Constructed response	6	–	6	12	14
<b>Total</b>	<b>63</b>	<b>17</b>	<b>80</b>	<b>86</b>	<b>100</b>

For the main survey, test items were allocated to seven clusters that were assembled into a fully balanced rotated test design comprising seven test booklets, each with a testing time of 45 minutes. One cluster of items comprised the set of secure CIVED items. Table 2.8 shows the cluster composition and Table 2.9 shows the rotated cluster test design used in the main survey. The CIVED trend-item cluster is shaded in Table 2.9.

Table 2.8: Main study cluster composition

Cluster	Items
C1	10 (9 MCQ, 1 CRQ)
C2	10 (9 MCQ, 1 CRQ)
C3	10 (9 MCQ, 1 CRQ)
C4	11 (10 MCQ, 1 CRQ)
C5	11 (10 MCQ, 1 CRQ)
C6	11 (10 MCQ, 1 CRQ)
C7 (CIVED)	17 MCQ

**Note:** MCQ = multiple-choice question; CRQ = constructed-response question.

Table 2.9: Main study test booklet design

Booklet	Position		
	1	2	3
1	C1	C2	C4
2	C2	C3	C5
3	C3	C4	C6
4	C4	C5	C7
5	C5	C6	C1
6	C6	C7	C2
7	C7	C1	C3

**Note:** CIVED link cluster shaded in grey.

### Mapping to framework

Table 2.10 shows the mapping of the main survey test items to the ICCS assessment framework.

A comparison of Table 2.10 and Table 2.5 reveals that the final test instrument provided a good coverage of the assessment framework and matched the original test-development plan. However, it needs to be acknowledged that, when compared to the originally anticipated coverage, the main survey test included a somewhat higher proportion of items relating to the cognitive domain *reasoning and analyzing*.



## Released test items

Two clusters of test items have been released following publication of the ICCS international report. The two released clusters were main survey clusters C1 and C7 (C7 was the CIVED secure-item cluster). Table 2.11 provides a summary of the released items by item type and the coverage of the ICCS assessment framework by the released item material.

Table 2.10: Main study item mapping to assessment framework

Content domain	New items	CIVED items	Total	Percentage of total items
Civic society and systems	17	14	31	39
Civic principles	24	2	26	32.5
Civic participation	18	0	18	22.5
Civic identities	4	1	5	6
<b>Total</b>	<b>63</b>	<b>17</b>	<b>80</b>	<b>100</b>
<b>Cognitive domain</b>				
Knowing	5	14	19	24
Reasoning and analyzing	58	3	61	76
<b>Total</b>	<b>63</b>	<b>17</b>	<b>80</b>	<b>100</b>

Table 2.11: Characteristics and mapping of ICCS released test items

	Released cluster C1	Released cluster C7 (CIVED)	Total
<b>Item type</b>			
Multiple choice	9	17	26
Constructed response	1	–	1
<b>Total</b>	<b>10</b>	<b>17</b>	<b>27</b>
<b>Content domain</b>			
Civic society and systems	4	14	18
Civic principles	4	2	6
Civic participation	2	0	2
Civic identities	0	1	1
<b>Total</b>	<b>10</b>	<b>17</b>	<b>27</b>
<b>Cognitive domain</b>			
Knowing	–	14	14
Reasoning and analyzing	10	3	13
<b>Total</b>	<b>10</b>	<b>17</b>	<b>27</b>



## Summary

The test-development process for ICCS was guided by the ICCS assessment framework and was carried out in different stages that involved a large number of reviews and revisions. Test-development staff members at the ISC were responsible for the drafting of item material, which was constantly revised and discussed with national center representatives. The numerous revisions of the ICCS item material were informed by the analysis of the pilot and the field trial data as well as by item paneling, focused discussions with pilot participants, and expert reviews.

## References

- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010a). *Initial findings from the IEA International Civic and Citizenship Study*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010b). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International Civic and Citizenship Education Study: Assessment framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).







# ICCS questionnaire development

Wolfram Schulz, Bruno Losito, and David Kerr

## Introduction

This chapter describes the development of the international questionnaires for students, teachers, schools, and national research centers. The student questionnaire was designed to measure both learning outcomes and contextual information. The teacher questionnaire was designed to gather teacher perspectives on the general school and community environment, teaching methods, and civic and citizenship education. School principals were asked to report on the school context for learning, on school characteristics, including school climate, as well as aspects of civic and citizenship education at their schools. An online questionnaire for national research coordinators (NRCs)—the national contexts survey—was designed to collect contextual information at the national (or subregional) level about the characteristics of education systems, aims, and contexts, implementation of civic and citizenship education, and current developments (reforms, debates) related to this learning area.

## Conceptual framework for questionnaire development

The assessment framework provided a conceptual underpinning for the development of the international instrumentation for ICCS (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008). The assessment framework consisted of two parts:

- *The civics and citizenship framework*: this outlined the outcome measures addressed through the cognitive test and those parts of the student questionnaire designed to measure student perceptions.
- *The contextual framework*: this mapped the context factors expected to influence outcomes and to explain their variation.

The ICCS assessment framework was organized around three dimensions, two of which were relevant for the development of the student questionnaire: a content dimension specifying the subject matter to be assessed within civics and citizenship (with regard to both affective-behavioral and cognitive aspects), and an affective-behavioral dimension describing the types of student perceptions and activities measured.

The four content domains in the ICCS assessment framework were civic society and systems, civic principles, civic participation, and civic identities. Each of these was made up of a set of subdomains that incorporated elements referred to as “aspects” and “key concepts.”

- *Civic society and systems, comprising three subdomains*: (i) citizens (roles, rights, responsibilities, and opportunities), (ii) state institutions (those central to civic governance and legislation), and (iii) civil institutions (the institutions that mediate citizens’ contact with state institutions and allow citizens to pursue many of their roles in their societies).
- *Civic principles, comprising three subdomains*: (i) equity (all people having the right to fair and just treatment), (ii) freedom (of belief, of speech, from fear, and from want), and (iii) social cohesion (sense of belonging, connectedness, and common vision amongst individuals and communities within a society).
- *Civic participation, comprising three subdomains*: (i) decision-making (organizational governance and voting), (ii) influencing (debating, demonstrating, developing proposals, and selective purchasing), and (iii) community participation (volunteering, participating in organizations, keeping informed).



- *Civic identities, comprising two subdomains:* (i) civic self-image (individuals' experience of place in each of their civic communities), and (ii) civic connectedness (sense of connection to different civic communities along with the civic roles individuals play within each community).

The assessment framework identified the different types of student perceptions and behaviors relevant to civics and citizenship. Four affective-behavioral domains were identified: value beliefs, attitudes, behavioral intentions, and behaviors.

- *Value beliefs:* these relate to fundamental beliefs about democracy and citizenship; they are more constant over time, more deeply rooted, and broader than attitudes.
- *Attitudes:* these include self-cognitions related to civics and citizenship, attitudes toward the rights and responsibilities of groups in society, and attitudes toward institutions.
- *Behavioral intentions:* these refer to expectations of future civic action. They include constructs such as preparedness to participate in forms of civic protest, anticipated future political participation as adults, and anticipated future participation in citizenship activities.
- *Behaviors:* these refer to present or past participation in civic-related activities at school or in the wider community.

The contextual framework identified the context variables that reflect the environment in which civic learning takes place. It assumes that young people develop their understandings about their roles as citizens through a number of activities and experiences that take place in the home, school, classroom, and wider community.

Students' knowledge, competencies, dispositions, and self-beliefs are influenced by their wider community (at local, regional, national, and supranational levels), their schools and classrooms (the instruction they receive, the school culture they experience, and their general school environment), their home environments (their direct home background and their social out-of-school environment), and their individual characteristics. The latter shape the way students *respond* to learning about civics and citizenship.

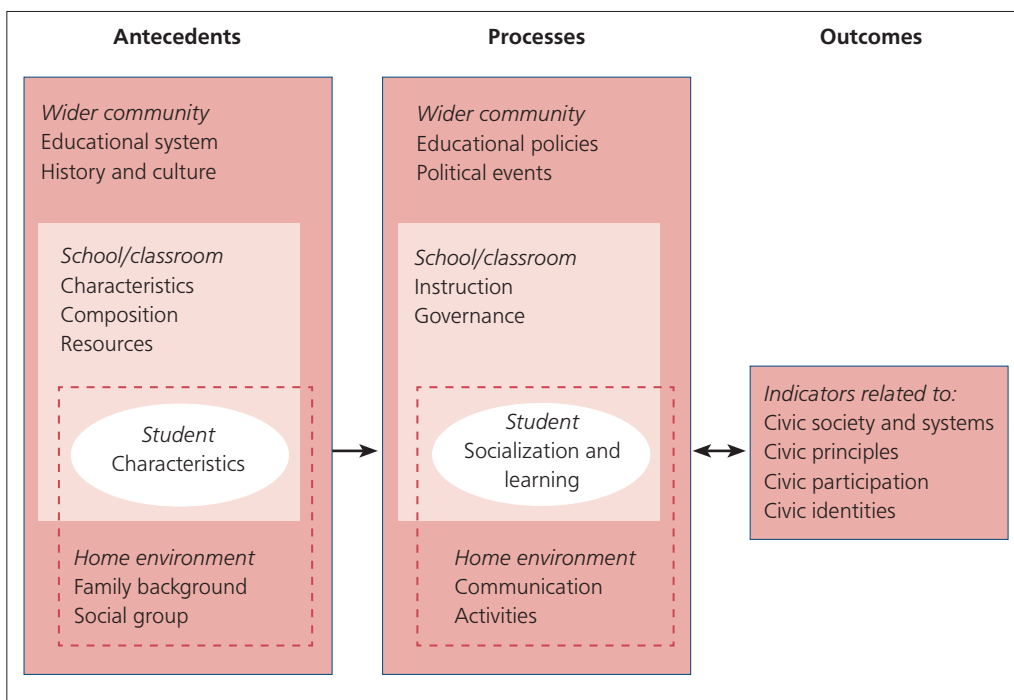
Contextual influences on civic and citizenship education act as either antecedents or processes. Antecedents refer to the historical background that affects how civics and citizenship learning takes place (e.g., through historical factors and policies that shape how learning is provided). Processes contemporaneously shape civic and citizenship education (e.g., the extent of civic understanding and engagement among students can influence the way schools teach this area of educational provision).

Figure 3.1 illustrates the contextual factors that influence the learning outcomes of civic and citizenship education. The (double-headed) arrow between *processes* and *outcomes* signals a reciprocal relationship. Feedback occurs between civic-related learning outcomes and processes. Students with higher levels of civic knowledge and engagement are the students most likely to participate in activities (at school, at home, and within the community) that promote these outcomes. The (single-headed) arrow between antecedents and processes describes the relationship between factors that are unidirectional.

Table 3.1 maps the variables (or groups of variables) that the ICCS researchers collected through their use of the various ICCS instruments. Variables related to the context of nation/community were collected primarily through the online national contexts survey. Variables related to the context of schools and classrooms were collected through the school and teacher questionnaires. The student background questionnaire provided information on the antecedents of the individual student and the home environment as well as about some process-related variables (e.g., learning activities). The student test and the student perceptions questionnaire



Figure 3.1: Contexts for the development of learning outcomes related to civics and citizenship



were used to collect data on outcomes. The student questionnaire also included questions about student participation in civic-related activities, the answers to which were used as indicators of active citizenship.

Table 3.1: Mapping of variables to contextual framework (examples)

Level of ...	Antecedents	Processes	Outcomes
National and other communities	<b>NCS and other sources:</b> Democratic history Structure of education	<b>NCS and other sources:</b> Intended curriculum Political developments	<b>StT and StQ:</b> Test results Student perceptions Student behaviors
School/classroom	<b>ScQ and TQ:</b> School characteristics Resources	<b>ScQ and TQ:</b> Implemented curriculum Policies and practices	
Student	<b>StQ:</b> Gender Age	<b>StQ:</b> Learning activities Practiced engagement	
Home environment	<b>StQ:</b> Parental SES Ethnicity Language Country of birth	<b>StQ:</b> Communication Peer-group activities	

**Note:** NCS = national contexts survey; ScQ = school questionnaire; TQ = teacher questionnaire; StQ = student questionnaire; StT = student test; SES = socioeconomic status.



The context of the wider community can be viewed as multilayered: there is the local community, comprising the students' school and home environment, which, in turn, is embedded within the broader regional, national, and (possibly) supranational contexts. Within the scope of ICCS, the level of the local community and the level of the national context were the most relevant levels.

## Student questionnaire development

The student questionnaire was initially conceived as consisting of two separate instruments, one capturing contextual variables and the other one measuring student perceptions related to value beliefs, attitudes, behavioral intentions, and behaviors. Given that some of the measures could be conceived of as either contextual or affective-behavioral variables (e.g., student reports on participation in community activities), it was decided to develop and administer the final student questionnaire as a single instrument.

The development and implementation of the student questionnaire was coordinated by the International Study Centre (ISC) at the Australian Council for Educational Research (ACER) in liaison with their partner institutions. This work also included, during different stages of the project, extensive reviews and discussions with experts and national centers.

The development process took place in three phases:

- *Phase 1*: the first phase included reviews of first-draft material by national centers and experts and the subsequent piloting of a draft questionnaire in six countries.
- *Phase 2*: the second phase involved another round of review by national centers and experts and an international field trial, this time conducted in 32 participating ICCS countries. The aim of this phase was to finalize the content of the student questionnaire.
- *Phase 3*: this, the final phase, involved discussion of the field trial results with the national center staff and experts followed by a final selection of main survey items.

During each of these phases, the national centers and experts used the following criteria to select proposed item material:

- Relevance with regard to the ICCS assessment framework;
- Appropriateness for the national contexts of the participating countries;
- Psychometric properties of items designed to measure latent traits postulated in the initial formulation and found in the pilot and/or field trial data.

The six countries involved in the piloting of the international student questionnaire material were Australia, Colombia, England, Italy, the Netherlands, and New Zealand. Two forms of the questionnaire, each of which took about 20 minutes to complete, were administered to convenience samples of target-grade students. Taken together, the two draft forms included a total of 252 affective-behavioral questionnaire items. The results from this pilot study were used in conjunction with feedback from the national centers and experts to elaborate a draft student questionnaire for the field trial.

The ICCS field trial questionnaire material included a total of 272 student items and was administered to samples in 32 participating countries. Three different questionnaire forms were used to trial a pool of questionnaire items that was larger than would have been possible in a single form; these items were allocated in a way that allowed analysis of all possible combinations of item-sets and item-scales.

The analyses of the field trial data were designed to provide empirical evidence for the selection of the main survey material; particular emphasis was placed on reviewing the cross-national validity of measures derived from the ICCS questionnaires (see Schulz, 2009). The field trial



outcomes were discussed with the national coordinators and experts, as was the draft student questionnaire proposed for use in the final data collection. This process led to formulation of the questionnaire that would be used during the final data-collection stage of ICCS.

The final international student questionnaire consisted of 173 items. Fifty-two of these items were designed to capture student-background information and 121 were designed to measure the affective-behavioral domains specified in the assessment framework. Another 22 items in the ICCS student questionnaire were optional: national centers could choose to administer or exclude them from their national instrument. The main survey student questionnaire consisted of the following sections:

- *About you:* this section of items included questions about the students' age, gender, and expected education.
- *Your home and your family:* these questions focused on characteristics of the students' homes and their parents.
- *Your activities:* these questions asked students to report on the extent to which they engaged in specified activities at home, at school, and in the wider community.
- *Your school:* students were asked to give their perceptions of different aspects of their schools.
- *Citizens and society:* these questions asked students about aspects of democracy and citizenship behavior.
- *You and society:* this block of questions asked students to give their views of their own relationship with different aspects of society.
- *Rights and responsibilities:* these questions were designed to measure student attitudes toward equal rights for gender groups, ethnic/racial groups, and immigrants.
- *Institutions and society:* this set of questions asked students to give their perceptions of civic institutions and their country.
- *Participating in society:* these questions focused on students' self-confidence with regard to active participation and on their likelihood of engaging in different modes of citizen participation in the future.
- *You and religion:* this block of questions, which each country could elect to include in the questionnaire, asked students about their religious background and practices as well as their attitudes toward the influence of religion on society.

Most of the student questionnaire items were developed at the ISC and at the center's partner institutions. However, national centers also proposed additional student questionnaire material, some of which was included in the final survey instrument.

Optional items were designed to capture variables that some of the participating countries perceived as either not relevant or inappropriate to their national contexts countries but which all of the other countries regarded as crucial. In these cases, single questions or sets of questions were included as international options. This meant that each national center could choose whether to administer this material. In addition, because there was some interest in measuring aspects of the European region (e.g., trust in European institutions) within the context of the international student questionnaire, several optional European items were added to some of the international item-sets.



The international options offered to countries participating in ICCS were:

- Students' ethnicity;
- Composition of students' households (i.e., the people living with student at home); and
- Religion.

Nineteen national centers chose to include the item on ethnicity, 37 national centers opted to include the item on household composition, and 28 chose to include the items measuring student perceptions of religion.

### Teacher questionnaire development

The teacher questionnaire was designed to collect information about school and classroom contexts, connections between schools and local communities, perceived objectives of civic and citizenship education, and approaches to teaching in this learning area.

The instrument was developed to gather data on characteristics of the school context, including school culture, climate and ethos, teachers' participation in school governance, teaching practices, and students' behavior at school and within the classroom. The relationships between schools and local communities included civic-related activities carried out by teachers with their target grade students within the local community as well as teachers' own participation in civic-related activities in the wider community.

Some of these constructs were also assessed through the school questionnaire, with the aim of collecting data on the same issues from the perspective of teachers and school principals. Specifically, questions about school climate and the priority assigned to civic and citizenship education and to its different objectives were included in both teacher and school questionnaires.

The assumption that teaching staff constitute an important factor in determining school climate and culture as well students' school experience meant that the teacher questionnaire was designed so that it could be completed by teachers teaching across all subject areas in the school curriculum for the ICCS target grade. The time that teachers took to complete the questionnaire was about 30 minutes.

The questionnaire also included an international option directed at teachers teaching subjects regarded as directly related to civic and citizenship education in a country. The subjects regarded as related to this learning area were determined by the national centers. Thirty-three of the 38 participating countries chose the international option, which included questions about teaching and assessment approaches to civic and citizenship education.

The development and implementation of the teacher questionnaire was coordinated by ICCS researchers at the Laboratorio di Pedagogia sperimentale at the University Tre of Rome (LPS) in liaison with the ISC. Draft questionnaire material was extensively reviewed by experts and national centers at different stages of the project.

The questionnaire development process involved four phases:

- *Phase 1:* international project staff developed and reviewed the draft material in cooperation with international experts. The first draft of the questionnaire was piloted in a small-scale study carried out in four countries (Colombia, England, Italy, and New Zealand). The results of the pilot were then used to refine the item material and develop closed-format questions. This process also drew on a review of responses from some open-ended questions included in the pilot questionnaire.





- *Phase 2:* during this phase, revised material from the pilot was used to form the first draft of the field trial questionnaire, which the national centers then reviewed.
- *Phase 3:* during this phase, the revised item material underwent a field trial in 32 of the participating countries.
- *Phase 4:* this phase involved reviewing the results from the international field trial and making a final selection of main survey item material. This review was informed by discussions with international experts and national centers.

Similar criteria as those for the student questionnaire (see previous section) were applied during selection of main-survey item material. Item selection involved consideration of the items' appropriateness for the national contexts in participating countries. Consideration was also given to existing differences between education systems and between schools within each participating education system.

The field trial teacher questionnaire consisted of 31 questions with a total of 203 items and was administered to target-grade teachers in the schools selected for the field trial. Twenty-nine of these items were included in the international option for teachers of civic-related subjects. On average, the field trial teacher samples consisted of about 300 teachers in each participating country.

The analyses of the field trial data were designed to provide empirical evidence that would inform the selection of the main survey material. The items included in the main study were selected in addition to and on the basis of the analyses of the field trial outcomes and the national-center reviews.

The final main survey teacher questionnaire consisted of 29 questions (181 items) and was divided into the following five sections:

- *General:* these questions concerned teacher background characteristics.
- *The school:* these questions focused on the school environment and issues related to participation in teaching and learning activities.
- *Civic and citizenship education at school:* these questions asked teachers about the delivery of civic and citizenship education in their respective schools.
- *Teaching of civic and citizenship education:* teachers were asked about subjects directly related to civic and citizenship education. This part of the teacher questionnaire was optional; participating ICCS countries could elect to include these questions.

### **School questionnaire development**

The school questionnaire was designed to collect information about the school context, the context of the local community where the school was located, and the opportunities that schools offered to students with respect to participation in civic-related activities in the wider community.

Factors related to the school context included school characteristics, such as school size and resources, school as a democratic learning environment, school autonomy, student, teacher, and parent participation in the running of the school, school climate and discipline, teachers' and students' sense of belonging to the school, and the approaches to civic and citizenship education adopted at the school level.

Factors related to the local community context centered on resources available to students in the local area as well as issues pertaining to social tension within the local community and within the school.





Some of the constructs measured through the school questionnaire were also assessed through the teacher questionnaire. The aim here was to collect data on the same issues from the perspectives of both teachers and school principals.

The school questionnaire was designed to be completed by school principals. The questions addressed school characteristics as well as school principals' perceptions of school processes that are thought to influence students' civic and citizenship education. The time needed to complete the questionnaire was about 30 minutes.

The development and implementation of the school questionnaire was coordinated by ICCS researchers at the LPS in liaison with members of the ISC. The development of this instrument included extensive reviews by experts and national centers at different stages of the project.

The questionnaire development process took place in four phases:

- *Phase 1:* during this phase, the international project staff developed and reviewed the draft material in cooperation with international experts. The first draft of the questionnaire, which was prepared in Italy, was preceded by interviews with selected school principals.
- *Phase 2:* the second phase involved taking the first-draft material and shaping it, after a review by the national centers, into the first questionnaire draft.
- *Phase 3:* during this phase, the draft material was administered to smaller samples of schools that had agreed to participate in the international field trial, which was undertaken in 32 ICCS participating countries.
- *Phase 4:* this final phase consisted of a review of field trial results. This formed the basis for the final selection of main survey questions and items, a process that took place after discussions with international experts and the national centers.

Each of these phases saw the same criteria that were used to select material for the other questionnaires being used. During the process of instrument development, particular attention was paid to the appropriateness of questionnaire material for the large variety of national contexts in participating countries as well as to existing differences between education systems and between schools within each participating education system. This latter consideration was particularly relevant for those education systems that allow schools to exercise a comparatively high level of autonomy in school curricula development and delivery.

The questionnaire included 23 questions with a total of 133 items and was administered to the principals of schools that were selected in the 32 ICCS countries that participated in the field trial. In most countries, about 25 school principals provided responses to the field trial questionnaire.

The analyses of field trial data were designed to provide empirical evidence for assisting with the selection of the main survey material. Given the relatively small number of responses in each of the participating countries, the analyses that could be carried out with the field trial data gathered with this instrument were somewhat limited in scope. The results of the school-questionnaire field trial were discussed with national coordinators and experts prior to the final selection of item material for the main study. The revisions that were made after the field trial also included a rewording of some of the items.

The final school questionnaire, which consisted of 22 questions with a total of 133 items, was divided into the following five sections:

- *General:* this section included questions about background characteristics of the school principals;



- *The school environment*: this section contained questions about school autonomy and the school as a democratic learning environment;
- *The local community*: this section comprised questions about the resources available to students in the local area and about issues of social tension within the local community and within the school;
- *Civic and citizenship education at school*: this section contained questions about how civic and citizenship education was implemented at the school;
- *School size and resources*: this section included questions about basic school characteristics such as school size, numbers of teachers, and school location.

## Development and implementation of the national contexts survey

The ways in which students develop civic dispositions and acquire knowledge and understanding in their formation as citizens are strongly influenced by factors at the country or *national context* level. These variables include, among others, the historical background, the nature of the political system, the structure of the education system, and the nature of the curriculum. The national contexts survey was designed to collect relevant data and information about both antecedents and processes at the country level. It was also undertaken in two phases: at the outset and toward the conclusion of the study. This procedure was followed so that the survey could be used to inform the process of developing instruments and would provide country- level data on the context for civic and citizenship education.

The development, coordination, analyses, verification, and reporting of the national contexts survey was coordinated by ICCS researchers at the National Foundation for Educational Research (NFER) in the UK in liaison with the ISC. These tasks also involved close working partnerships with the NRCs from the participating countries.

The development process and implementation process consisted of four phases:

- *Phase 1*: during this first phase, agreement was reached on the nature and scope of the contexts and questions to be included in the survey. The national centers and experts reviewed first drafts of the survey and the online version of the survey.
- *Phase 2*: during this phase, initial data from the survey were analyzed. Where necessary, national centers were asked to provide any missing data and clarify inconsistencies in the data.
- *Phase 3*: during this phase, NRCs reviewed the survey data for their country and updated the data in the light of any changes in national contexts since the initial completion of the survey.
- *Phase 4*: this final phase consisted of final reviews and checks of survey data. This process was conducted in close dialogue with national centers in order to ensure consistency, completeness, and comparability.

Consideration of the contexts and questions to be included in the national contexts survey, required the following criteria to be applied across the four phases:

- Relevance with regard to the ICCS assessment framework;
- Relevance and additional value in relation to information about the context of the wider community for civic and citizenship education that was already in the public domain;
- Appropriateness for the national contexts of the participating countries;
- Validity in terms of comparability, analysis, and reporting.



Discussion with national centers and experts led to initial agreement on the following design principles for the national contexts survey:

- Completion by national centers in two stages, with the first stage at the outset of countries' participation in ICCS and the second stage toward the end of the study. This second stage was close to the student main survey period in order to give national centers opportunity to review and update any national contexts in their country;
- A common structure of largely closed, multiple-choice questions and a small number of open-format questions in order to facilitate cross-national comparison;
- Online completion in order to speed up the collection, analysis, verification, and reporting of data;
- Completion of the online questionnaire by NRCs in each country, drawing on further sources and expertise as required;
- Survey administration in English;
- Request for NRCs to draw upon rather than duplicate existing sources of information about the context of the wider community for civic and citizenship education, whenever appropriate; and
- Requirement to keep the collection of relevant data and information manageable so as to avoid the loss of valuable resources, especially that of time, given the other commitments of national center staff.

After further discussion and input, developed a final version for Stage 1 was developed and the national contexts survey, with accompanying notes for guidance, was placed online via the server at the IEA Data Processing and Research Center (DPC) in Hamburg (Germany).

The survey consisted of 46 questions concerning key antecedents and processes in relation to civic and citizenship education in each country and comprised eight sections:

- Education system;
- Education policy and civic and citizenship education;
- Approaches to civic and citizenship education;
- Civic and citizenship education within the context of school curriculum approaches;
- Civic and citizenship education in the school curriculum at the ICCS target grade;
- Teachers and civic and citizenship education;
- Civic and citizenship education and assessment and quality assurance; and
- Current debates and reforms.

The overall completion time was about one hour. However, the online facility enabled national center staff to complete the survey in more than one administration session.

Initial analyses of the national contexts survey data were conducted for 40 questions from 26 participating countries. These analyses were designed to fulfill three purposes: first, to check the consistency and completeness of the data collected in the survey from each participating country; second, to help frame the processes that would guide the review and updating of data by national centers in the second stage of data collection; and, third, to inform decisions about how the contexts data should be reported in order to best provide quick, informative, comparative overviews across the 38 participating countries.

The online data from each of the 26 completed surveys were thoroughly checked for consistency and plausibility. National contexts survey data were used to create national profiles for each country to facilitate both the checking of data and the review of cross-national patterns.



National profiles provided useful summaries of the information provided by the national centers and allowed ICCS researchers to identify issues with respect to missing data and inconsistencies. Country profiles were then reviewed and discussed with the national centers. National centers were asked to provide any missing data, explain for any inconsistencies in their data, and make any changes necessary to reconcile inconsistencies.

Some of the inconsistencies found in the national contexts data related to translation difficulties, given that the national contexts information had to be translated into English. This process of review and refinement by the national centers was crucial in ensuring both the completeness and consistency of the contexts data that the participating countries provided during the first stage of data collection.

During the second stage, and toward the end of the study, national centers were asked to review the online data in their national contexts survey and to update it in terms of any changes in contexts for civic and citizenship education that had occurred in the intervening period. The main changes related to the survey section on current debates and reforms. National center staff indicated the changes they had made in the national contexts survey by using a color and style of text which differed from that used in the first stage.

After the second stage of the national contexts survey, a similar set of analyses was carried out, after which checking and reviewing procedures occurred in the same way as that used in the first stage. This final review used data from all 46 questions and all 38 participating countries, as well as updated national profiles for each country. In instances of any remaining inconsistencies in the updated context information, countries were asked to provide further clarification and to make final adjustments to their online national contexts survey.

## Summary

The ICCS assessment framework (Schulz et al., 2008) was the principal basis for the development of the ICCS questionnaire material because it identified the content to be measured in the surveys of students, teachers, schools, and national centers. The student, teacher, and school surveys were developed in a multistage process that included smaller pilots, a general field trial, and extensive discussions with national centers and international experts.

When developing cross-national survey instruments, it is important to maximize the input from the wide range of stakeholders. Therefore, the contributions from national centers were crucial in the process of developing the ICCS. The proposed questionnaire material was reviewed by national coordinators in several rounds of written consultations as well as in plenary discussions at their international meetings. In addition, national centers proposed some additional item material for the ICCS surveys.

The collection of data on national contexts was conducted in two stages. These informed both the development of ICCS instruments and measures as well as the interpretation of results. Furthermore, the national contexts survey provided a rich database with country-level information on civic and citizenship education.

## References

- Schulz, W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study. *IERI Monograph Series, Volume 2*, 113–135.
- Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International Civic and Citizenship Education Study: Assessment framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).





# The development of regional instruments

David Kerr, Wolfram Schulz, and Julian Fraillon

## Introduction

As described in the first chapter, ICCS included regional modules for Europe, Latin America, and Asia. Thirty-five of the 38 countries that participated in ICCS opted to be involved in these regional modules. Twenty-four countries participated in the European module, six in the Latin American module, and five countries in the Asian module. Regional instruments were developed for each module and were administered to students after they had completed the international assessment. These instruments were designed to address aspects of civic and citizenship education specific to each region. The results of these additional data collections are reported in a series of regional reports (Fraillon, Ainley, & Schulz, forthcoming; Kerr, Sturman, Schulz, & Burge, 2010; Schulz, Ainley, Friedman, & Lietz, 2011).

This chapter describes the development of the three regional-module instruments, namely:

- *The European regional instrument*: this consisted of a 12-minute cognitive test and a 17-minute region-specific questionnaire (29 minutes total);
- *The Latin American regional instrument*: this consisted of a 15-minute cognitive test and a 15-minute region-specific questionnaire (30 minutes total);
- *The Asian regional instrument*: this comprised a 15-minute region-specific questionnaire.

The ICCS assessment framework (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008) provided a reference point for the development of the regional modules. The international study center (ISC) and its associated partner institutions worked with national centers and experts from the three regions to identify aspects that were deemed relevant within each region and therefore appropriate for inclusion in the assessment. The instrument development involved collaborative effort among staff of the national centers as well as experts from the respective regions.

## European test and questionnaire development

The European regional instrument, consisting of a European cognitive test and a European questionnaire, were developed from a regional framework that was linked to the international framework but also identified elements considered pertinent to the region.

The regional framework was drawn up by ICCS researchers, who began the process by identifying potential elements for inclusion from a review of existing developments and mapping these against the ICCS assessment framework. This process of identification and mapping was informed by contributions from individual European countries as well as from cross-national European groups. The regional framework helpfully identified the knowledge, attitudes/values, and competencies to be investigated through the regional instrument in addition to the aspects already measured within the international instruments. ICCS researchers then discussed this framework and mapping in a series of meetings with European national research coordinators (NRCs).

These meetings between the ICCS researchers and European NRCs led to decisions about the scope and focus of the regional instrument. It was decided that the regional student assessment would consist of two components—a cognitive test and a questionnaire. Another decision

---

<sup>1</sup> Norway and the Russian Federation decided not to participate in the European module, although their representatives were involved in preliminary discussions about the module. There was no suitable regional module in which New Zealand could participate.



was that the test would focus on knowledge of the European Union (EU) and its policies, institutions, practices, and processes, including the euro currency. Meanwhile, the questionnaire would concentrate on five specific regional issues considered to be of high importance: European citizenship and identity; intercultural relations in Europe; free movement of citizens in Europe; European political policies, institutions, and participation; and European language learning.

The development of the European instrument was a collaborative effort undertaken under the supervision of ICCS researchers at the National Foundation for Educational Research (NFER, UK) in liaison with the ISC. It also involved NRCs, cross-national European groups, and scholars in the field. The development process consisted of four interrelated phases:

- *Writing the regional test and questionnaire items:* this writing was informed by the ICCS assessment framework and included small-scale pilots in a number of participating countries as well as extensive ongoing consultations with the NRCs and rigorous reviews by expert consultants.
- *Implementation of an international field trial that included the regional instrument and took place in 20 participating countries in the region:* collection of data from smaller samples of students, schools, and teachers occurred during this phase.
- *Further writing, rewriting, and trialing of the regional test items:* this writing was informed by the field trial analysis and included review by expert consultants, a small-scale pilot in some participating countries, and consultation and feedback from the national centers.
- *Final revision of the regional test and questionnaire items:* this work was undertaken by ICCS researchers at NFER working in liaison with the ISC and expert consultants. It was informed by the field trial analysis and results, the small-scale pilot of test items, and further feedback from the national centers.

The ICCS regional field trial instrument comprised 32 cognitive test items and 101 questionnaire items, with an assessment time of 10 minutes for the regional test and 20 minutes for the regional questionnaire. The psychometric properties of the item material were then reviewed against analyses of the field trial data collected via the regional instrument from the 20 European ICCS countries.

The final item selection for the cognitive and questionnaire material was carried out in collaboration with the NRCs from the 24 countries participating in the regional module. The selection and ordering of items on the empirical evidence from the field trial and a further small-scale pilot of test items as well as consideration of the balance of elements and their relationship to the ICCS assessment framework.

The final regional instrument used in the ICCS main survey saw the retention of 20 cognitive items and 83 questionnaire items. The regional module had an overall stipulated assessment time of 29 minutes, consisting of 12 minutes for the regional test and 17 minutes for the regional questionnaire.

The final European instrument addressed the following region-specific cognitive and affective-behavioral aspects:

- Students' knowledge of facts about the European Union (EU) and its institutions, about EU laws and policies, and about the euro currency;
- Students' perceptions of European identity;
- Students' reports of engagement in activities related to Europe;
- Students' attitudes toward learning of European languages;





- Students' attitudes toward migration within Europe (attitudes toward freedom of movement, restriction of migration, and equal opportunities for EU citizens from other countries);
- Students' attitudes toward European integration (attitudes toward common policies, European unification, a common European currency, and further EU expansion); and
- Students' self-reported knowledge about the EU.

### Latin American test and questionnaire development

The regional instrument was designed in line with a regional framework that was developed and linked to the international framework but identified elements deemed particularly relevant to the region. Using as their basis a review of current definitions of such elements, the ICCS regional expert group, consisting of scholars from each participating country, delineated the knowledge, attitudes/values, and competencies to be investigated in addition to the aspect already measured with the international instruments. Within each of these dimensions, the group identified three general themes: "peaceful coexistence," "democratic participation," and "plurality and diversity."

The development of the Latin American instruments was a collaborative effort undertaken under the supervision of the ISC. It involved NRCs, the regional expert group, and consultants on assessment and measurement. The development process comprised three phases:

- *Writing the test and questionnaire items:* this work was guided by the ICCS assessment framework and included smaller pilots in some of the participating countries as well as extensive consultations with the NRCs and the expert consultants.
- *Implementation of an international field trial in all participating countries in the region:* collection of data from smaller samples of schools, students, and teachers also occurred during this phase.
- *A final revision of the material:* this was undertaken at the ISC as well as a number of consultants, and it was conducted in light of the field trial results and further feedback from the national centers and experts.

The ICCS regional field trial instrument consisted of 19 cognitive test items and 105 questionnaire items, with an assessment time of 15 minutes for the regional test and 25 minutes for the regional questionnaire. The field trial data from the six Latin American ICCS countries were then used as the basis upon which to review the psychometric properties of the item material. The final item selection was undertaken in collaboration with the NRCs from the countries participating in the regional module and in line with the empirical evidence derived from the field trial as well as with conceptual considerations.

The final regional instrument used in the ICCS main survey contained 16 retained test and 74 questionnaire items. Each of the two parts of this instrument had a stipulated assessment time of 15 minutes, giving an overall assessment time of 30 minutes.

The final Latin American instrument addressed the following region-specific cognitive and affective-behavioral aspects:

- Students' region-specific civic knowledge;
- Students' perceptions of Latin American identity;
- Students' perceptions of government and law (attitudes toward authoritarian government, corruption, and disobeying the law);
- Students' perceptions regarding peaceful coexistence (attitudes toward neighborhood diversity or violence, feelings of empathy, and experience of aggression); and
- Students' reports on discussion of civic issues at school.





## Asian questionnaire development

The Asian regional questionnaire was developed collaboratively by NRCs from the Asian countries participating in ICCS and a small number of additional experts nominated by the NRCs from within their own national centers. The ISC supported the organization of the questionnaire development and were responsible for quality assurance of the questionnaire, including its fit to the ICCS assessment framework and consistency with the full ICCS suite of instruments.

The development process comprised four phases:

- *Establishing an Asian regional framework:* this began with a meeting where NRCs and experts shared information about their own national experiences of civic and citizenship education research and outlined constructs of interest for inclusion in the Asian regional questionnaire. These constructs were organized according to the structural headings in the ICCS assessment framework (Schulz et al., 2008).
- *Writing the test and questionnaire items:* this work was guided by the ICCS Asian regional framework and included smaller pilots in some of the participating countries as well as extensive consultations with the NRCs and expert consultants.
- *Implementation of an Asian regional questionnaire pilot in all participating countries in the region:* collection of data from smaller samples of schools, students, and teachers also occurred during this phase.
- *The final revision of the material:* this was undertaken by the ISC and consultants in light of the field trial results and further feedback from the national centers and experts.

The Asian regional questionnaire for the pilot study comprised 97 items with an assessment time of 20 minutes. Pilot data from the five Asian ICCS countries were then used to review the psychometric properties of the item material. The final item selection was conducted in collaboration with the Asian ICCS NRCs and was based on empirical evidence derived from the field trial as well as on conceptual considerations.

The ICCS main survey Asian regional module questionnaire comprised 55 items, with a completion time of 15 minutes.

The final Asian instrument addressed the following region-specific cognitive and affective-behavioral aspects:

- Students' perceptions of government and law in Asia (attitudes toward undemocratic government and obedience to authority as well as perceptions of the legal system);
- Students' perceptions of identity, citizenship, and culture in Asia (attitudes toward traditional culture and Asian citizenship as well as perceptions of good citizenship); and
- Students' perceptions of public service (attitudes toward corrupt practices in the public service, personal morality of politicians, and use of connections to hold public office).



## Summary

ICCS established three regional modules for Europe, Latin America, and Asia. These were designed to assess region-specific aspects of civic and citizenship education that were not included in the international data collection. Twenty-four European, six Latin American, and five Asian countries chose to participate in these regional modules.

In each of the regions, national coordinators, national experts, and international project staff worked collaboratively to develop regional instruments that addressed particular research questions derived from region-specific frameworks with links to the ICCS assessment framework. For each of the regional instruments, the development process involved reviews and discussions with national and international experts as well as piloting of item material in participating countries.

## References

- Fraillon, J., Ainley, J., & Schulz, W. (forthcoming). *ICCS 2009 Asian report: Civic knowledge and attitudes among lower-secondary students in five Asian countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Kerr, D., Sturman, L., Schulz, W., & Burge, B. (2010). *ICCS 2009 European report: Civic knowledge, attitudes, and engagement among lower-secondary students in 24 European countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Friedman, T., & Lietz, P. (2011). *ICCS 2009 Latin American report: Civic knowledge and attitudes among lower-secondary students in six Latin American countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International Civic and Citizenship Education Study: Assessment framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).





## CHAPTER 5:

# Translation and national adaptations of ICCS 2009 instruments

Barbara Malak, Alana Yu, Wolfram Schulz, and Tim Friedman

## Introduction

In close collaboration with the participating countries, the international study center (ISC) developed an international English version of the ICCS assessment and questionnaires. These materials were subsequently translated and adapted by countries to their languages of instruction. Throughout this translation and adaptation process, the overarching aim was to create high-quality instruments that were internationally comparable yet also appropriate to each country's national context and education system. Detailed guidelines on translation and adaptation, which were provided to all participating countries, are described in the *ICCS 2009 Survey Operations Procedures, Unit 1* (ICCS International Study Center, 2008).

The ICCS instruments were administered in 31 languages, the most common being English and Spanish. Even when the language of testing was English, adaptations were still required to suit the cultural setting and the version of English being used. Because all countries participating in the Latin American module required Spanish as the assessment language, the instrument was developed in this language for use and adaptation by countries participating in this module.

Given that high-quality translations were crucial to the quality of the ICCS data, all national instruments were subject to a stringent international verification process that included a thorough review of the adaptations, translation, and layout. This process was intended to ensure that national versions were equivalent across countries, to the greatest extent possible. The ISC managed the adaptations' consultation and layout verification; translation verification was coordinated by the IEA Secretariat. These agencies asked the participating countries to submit materials for verification before both the field trial and the main survey data collection.

In general, countries complied very well with the requirements for verification. All participants of the field trial and the main survey submitted their instruments for national adaptations review, translation verification, and layout verification.

## Translation of ICCS 2009 instruments

### *ICCS Instruments to be translated and adapted*

The ICCS instruments requiring translation and/or adaptation were:

- The student cognitive test (including instructions and booklet covers);
- Questionnaires for students, teachers, and school principals (including instructions and covers); and
- The regional student instruments (including instructions and covers), *for countries participating in a regional module*.

ICCS manuals and guides were also translated where necessary. These resources included the following:

- The school coordinator and test administrator manuals; and
- The scoring guides for constructed-response items.



Of these, the survey instruments (cognitive test, questionnaires, and regional module) were subject to the international verification procedure. The ISC provided participating countries with electronic files of all materials to be adapted and/or translated. In addition, the cognitive test and questionnaire items were listed in a single combined document, the national adaptations form (NAF), in which national research coordinators (NRCs) registered their adaptations to the instruments. Reviewers of these documents listed suggestions for changes, which were then commented on by the NRCs.

### *Languages used in ICCS 2009*

For most participating countries, identifying the language that would be used for testing (i.e., the target language) was straightforward. This was typically the dominant language used in public and private arenas of society, including the education system. However, in some countries, there was more than one official language or language of instruction in schools. In these cases, countries prepared instruments in all required languages. Ten countries administered all or parts of the assessment (most commonly, the student instruments) in two or more languages. Table 5.1 shows the list of languages used for the ICCS survey.

Participating countries were strongly encouraged to hire qualified and experienced translators and reviewers to work with the ICCS materials.

National centers were expected to enlist at least one translator (preferably certified) per target language, with the following qualifications:

- Excellent knowledge of English;
- Target language as a native language;
- Knowledge of and experience in the country's present cultural context and, if possible, experience translating texts on social and/or political issues;
- Experience with students in the target grade; and
- Familiarity with test development.

Reviewers were given the task of assessing the translation's readability for the target population. They were required to have the following qualifications:

- Excellent knowledge of English;
- Target language as a native language;
- Knowledge of and experience in the country's present cultural context; and
- Experience with students in the target grade.

Countries that administered the assessment in more than one target language were advised to employ a professional competent in all languages, who could ensure that adaptations were implemented consistently in the different language versions. National centers were permitted to hire more than one translator/reviewer per language (for instance, one person to translate the test, another person to translate the questionnaires), but were responsible for maintaining the consistency of the translations and adaptations within and across instruments.

### *Guidelines for translation and adaptation of the instruments*

The guidelines for translation and adaptation provided to all countries were designed to ensure the international comparability of the national versions of the instruments, while allowing for cultural adaptations when necessary. All of the instruments required some kind of adaptation, and these were subject to a careful documentation and review procedure. The overarching principle of the translation and adaptation process was that students from different countries should receive exactly the same questions.



Table 5.1: Languages used for the ICCS 2009 survey instruments

Country	Language	Instruments				
		Student test	Student questionnaire	Teacher questionnaire	School questionnaire	Regional instrument
Austria	German	•	•	•	•	•
Belgium (Flemish)	Dutch	•	•	•	•	•
Bulgaria	Bulgarian	•	•	•	•	•
Chile	Spanish	•	•	•	•	•
Chinese Taipei	Traditional Chinese	•	•	•	•	•
Colombia	Spanish	•	•	•	•	•
Cyprus	Greek	•	•	•	•	•
Czech Republic	Czech	•	•	•	•	•
Denmark	Danish	•	•	•	•	•
Dominican Republic	Spanish	•	•	•	•	•
England	English	•	•	•	•	•
Estonia	Estonian	•	•	•	•	•
	Russian	•	•	•	•	•
Finland	Finnish	•	•	•	•	•
	Swedish	•	•	•	•	•
Greece	Greek	•	•	•	•	•
Guatemala	Spanish	•	•	•	•	•
Hong Kong SAR	Traditional Chinese	•	•	•	•	•
Indonesia	Indonesian	•	•	•	•	•
Ireland	English	•	•	•	•	•
	Irish	•	•	•	•	•
Italy	Italian	•	•	•	•	•
Korea, Republic of	Korean	•	•	•	•	•
Latvia	Latvian	•	•	•	•	•
	Russian	•	•	•	•	•
Liechtenstein	German	•	•	•	•	•
Lithuania	Lithuanian	•	•	•	•	•
	Polish	•	•	•	•	•
	Russian	•	•	•	•	•
Luxembourg	French	•	•	•	•	•
	German	•	•	•	•	•
Malta	English	•	•	•	•	•
	Maltese	•	•	•	•	•
Mexico	Spanish	•	•	•	•	•
The Netherlands	Dutch	•	•	•	•	•
New Zealand	English	•	•	•	•	•
Norway	Bokmål	•	•	•	•	•
	Nynorsk	•	•	•	•	•
Paraguay	Spanish	•	•	•	•	•
Poland	Polish	•	•	•	•	•
Russian Federation	Russian	•	•	•	•	•
Slovak Republic	Slovak	•	•	•	•	•
Slovenia	Slovene	•	•	•	•	•
Spain	Basque	•	•	•	•	•
	Catalan	•	•	•	•	•
	Galician	•	•	•	•	•
	Spanish (Castilian)	•	•	•	•	•
	Valencian	•	•	•	•	•
Sweden	Swedish	•	•	•	•	•
Switzerland	French	•	•	•	•	•
	German	•	•	•	•	•
	Italian	•	•	•	•	•
Thailand	Thai	•	•	•	•	•

### **Adaptation of the instruments**

In the specific area of civics and citizenship, a number of modifications were required and allowed beyond those necessitated by translation into the target language. They included adaptations to particular concepts that were not common to all countries, such as specific institutions and organizations. For instance, the term “national parliament” (intended to refer to a legislative body at the national level) was adapted to “Parliament” in New Zealand, but “Eerste en Tweede Kamer” (Lower and Upper Chamber) in the bicameral system of the Netherlands. The goal of such adaptations was to make the questions equally familiar to all students, while maintaining the same meaning and level of difficulty.

It was important that the cognitive items not be simplified, clarified, or adapted in such a way as to provide students with a hint or definition of a term that was not given in the international English version. For example, if an item required students to define or identify a particular aspect of democracy, it was essential that the term “democracy” not be translated in such a way as to provide the definition or aspect of democracy in question. It was also important that adaptations be implemented consistently throughout the instruments and, in particular, that the correspondence of text in the stem and options of multiple-choice items be maintained.

The international version of the materials had within them indications of where adaptations were required; any words in angle brackets (carets) needed to be replaced with the country-appropriate term. NRCs were instructed to adapt certain recurring base expressions from the questionnaires according to the particular country context. For example, <country of test> would be replaced with the name of the participating country, and <target grade> would be replaced with the name of the specific target grade in that country. Generic ISCED levels in the student questionnaire were adapted to the equivalent educational terms for each country.

Some references to names of people and fictional places/countries (e.g., <Male name>, <Zedtown>) were also specifically designated for adaptation. These references were adapted to names in the target language that were of similar length, familiarity, and complexity, the aim being to convey the same meaning and style of text as in the international version. When adapting fictional names of countries or towns, translators were explicitly instructed not to use the names of real places or countries so that students’ responses would not be influenced by their knowledge or perceptions of them. Modifications could also be made when necessary to adapt national conventions, such as measurement units and punctuation.

In principle, words not written in carets were not to be adapted. NRCs were provided with detailed notes on all required adaptations, along with the *Operational Manual for ISCED-1997* (UNESCO Institute of Statistics, 1998). These notes clarified what the particular questions were asking so that translators could select the appropriate word or expression to convey the intended meaning.

Participating countries were permitted to add a limited number of national items or categories to the questionnaires, totaling up to five minutes of survey time (approximately 25 items). No national additions were allowed for the cognitive test. NRCs were instructed to place all national items at the end of the questionnaires, and these items were subject to documentation and prior approval for inclusion by the ISC.

### **Test items for overtime comparison**

The student cognitive test for ICCS contained one cluster of 17 test items that was also used in the IEA Civic Education Study (CIVED) of 1999. These items provided the basis for comparisons of changes in achievement from the earlier assessment, for those countries that participated in both studies with comparable populations.



Countries were expected to use the same translation for the ICCS items as was used in the previous assessment. In some cases, however, national centers considered modifications to the translations from CIVED absolutely necessary. In these cases, the changes were carefully documented and referenced during the data analysis. If the changes seemed to have altered the performance of any item, this item was not included in the scaling of link items to measure changes in civic content knowledge since 1999 for that particular participating country.

### **International verification of the instruments**

In addition to the internal review of translations carried out by each national center, all survey instruments went through a rigorous three-part international verification process: (i) adaptations negotiation, (ii) translation verification, and (iii) layout verification. An independent review of the translation verification record was also conducted by international quality-control monitors, as part of the ICCS quality assurance program.

#### *Documentation in the national adaptations form*

When translating and adapting the international version of the instruments for national use, national centers needed to make certain changes, selections, and adaptations to the survey instruments. In doing so, they were required to keep in mind that the objective of the study was to create an international database containing comparable data from all participating countries with complete documentation. Consequently, any change that was made had to be recorded electronically on the NAF. This form was used not only for documentation purposes, but were also when national data were added to the international database.

The NAF in Microsoft® Excel format consisted of several worksheets for the cognitive test, student questionnaire, teacher questionnaire, school questionnaire, regional test/questionnaire, and additional country documentation (language information, version number, and inclusion status of international optional questions). National centers were required to complete a NAF for each survey language used in their country. National centers were also requested to document whether they intended to include any of the international options or any national items and categories (together with a description of their content in both the national language and English). For those countries measuring overtime comparisons in achievement, adaptations for existing CIVED items had to be documented in the NAF, in the same way as for the other national adaptations.

The NAFs were completed and reviewed at various stages of the verification process. Detailed instructions on how to work with the NAFs and how to adapt the data-entry software were therefore provided to national center staff at data management seminars preceding the field trial and main survey.

#### *Adaptations negotiation*

NRCs were required to consult with ISC staff when reviewing all proposed national adaptations. In particular, they were strongly encouraged to discuss any adaptation that might result in a serious deviation from the items in the international instruments.

National centers began completing the NAF (Version I) after reviewing the international version of the survey instruments. They submitted the NAF to the ISC and consulted with them on the form's contents. Following the review process, the ISC provided the national centers with feedback on their adaptations and, where appropriate, suggested improvements.

National centers were requested to take the recommendations into account and update the forms accordingly so that these updated forms (Version II) could be used during the translation verification process to evaluate the quality and accuracy of the translations.





### *International translation verifiers*

The IEA Secretariat enlisted the assistance of an independent translation company, cApStAn Linguistic Quality Control (Brussels, Belgium), to verify the translations for each country. International translation verifiers for ICCS were required to have the target language as their first language, have formal credentials as translators working in English, be educated at university level and, if possible, live and work in the target country (or be in close contact with it).

Verifiers attended a training seminar where they received detailed instructions for reviewing the survey instruments and registering deviations from the international version. They also received general information about the study and design of the instruments, together with a description of the translation procedures used by the national centers.

### *International translation verification*

The primary task of the translation verifiers was to evaluate the accuracy and comparability of the national versions of ICCS instruments. The instructions given to verifiers emphasized the importance of maintaining the meaning and difficulty level of each test and questionnaire item. Specifically, verifiers had to ensure the following:

- The translation had not affected the meaning or reading level of the text;
- No information had been omitted from or added to the translated text;
- The test items had not been made easier or more difficult;
- The instruments contained all of the correct items and response options, in the same order as in the international version;
- All national adaptations implemented in the instruments were documented in the NAF.

The verifiers used the editing functions of Microsoft® Word (“Track Changes” and “Insert Comments”) to document any errors or suggested changes directly in the submitted instruments. Verifiers were asked to provide suggestions that would improve the comparability of the instruments when appropriate, and to evaluate the overall quality, accuracy, and cultural relevance of the translation.

To help NRCs understand the comparability of the translated text with the international version, verifiers were asked to assign a “severity code” to any deviations. The severity code indicated how major or minor the deviation was. Severity codes ranged from 1 (*major change or error*) to 4 (*acceptable change*), as described below.

1. *Major change or error*: examples included the incorrect order of choices in a multiple-choice item; incorrect order of items; omission of a graphic, item, or answer option; incorrect translation resulting in the answer being suggested by the question; and an incorrect translation that changed the meaning or difficulty level of an item.
2. *Minor change or error*: examples included spelling errors that did not affect comprehension.
3. *Suggestion for alternative*: the translation was deemed adequate, but the verifier suggested a different wording.
4. *Acceptable change*: the change was deemed acceptable and appropriate, but was not necessarily documented in the NAF. An example of an acceptable adaptation is the case where a reference to winter was changed from January to July in the instruments for participating countries from the Southern Hemisphere.

Additionally, for countries that participated in CIVED, verifiers were required to compare the translation of items as administered in CIVED against the translation used for ICCS. Any discrepancies were documented in a special form.



The translation verification feedback was returned to the NRC of each participating country. The NRCs were responsible for reviewing the translation verifiers' suggestions and revising the instruments according to this feedback. NRCs were also asked to complete a translation verification summary form after the field trial verification, and to comment on verifier suggestions that they had decided not to implement.

#### *Results of the translation verification*

In general, the translation verifiers considered the national translations/adaptations to be well documented and of very high quality, showing a good balance between faithfulness and fluency. Translation verifiers of the main study instruments also noted the great care with which their verification feedback after the field trial was implemented.

Some typical language errors identified during translation verification included mistranslations, fluency issues ("free" vs. "word-for-word" translations, Anglicism), inconsistencies, omissions/additions, adaptations of names (fictional vs. real), style (gender agreement, formality), and grammar. Some verifiers noted the challenge of translating and adapting certain concepts related to civics and citizenship for the particular national context, especially in the case of non-existing institutions (e.g., school council, school governing board), as well as in the use of acronyms and abbreviations (adapted English name vs. translated name). Through the extensive documentation collected in the NAF, verifiers could provide meaningful feedback on any issues arising with the national adaptations, taking into account both the considerations reported by the national center and the recommendations given by the ISC during the adaptation negotiation.

#### **Layout verification**

Once adaptation and translation verification had been completed, national centers were asked to compile their final set of instruments in PDF format for each test language to be used in the main survey. These documents were uploaded to a secure server, along with an updated NAF (Version III), which contained any changes resulting from translation verification.

These files were accessed by staff at the ISC for layout verification. Two independent reviewers at the ISC reviewed each set of materials. All layout issues identified were documented in a worksheet added to the NAF. The layout issues in each set of instruments were grouped as to whether they were general layout issues relating to the set of instruments, or whether they related to a specific question or specific group of questions within an instrument. A wide range of layout issues was identified across countries. These included formatting issues (e.g., spacing, font size, margins, consistency across questions), incorrect order of questions, missing text, and the addition of questions not agreed upon from adaptation review.

National centers were provided with a summary of all layout issues. In cases where layout issues were considered minor, national centers were given feedback and were asked to make the appropriate changes to their materials without need for further verification. In cases where more substantial layout issues were identified, national centers were provided with detailed feedback concerning all issues and were asked to resubmit their materials for further layout verification.

After layout verification was complete and the ICCS instruments were finalized, a final version of the NAF (Version IV) was prepared and used by the IEA Data Processing and Research Center for data processing.



### *Quality control monitor review*

IEA hired international quality-control monitors (IQCMs) from each country to document the quality of the ICCS assessment administration, including the survey materials.<sup>1</sup> An important part of the IQCMs' responsibilities was a careful review of the instruments used during the main survey data collection. The IQCMs compared the final (printed) version of the questionnaires and test booklets against the translation verification record to ensure that the recommendations of the translation verifier were addressed appropriately.

### **Summary**

The survey instruments and verification procedures were developed through an extensive process of cooperation, independent review, and consensus. Detailed documents helped the national centers follow the internationally agreed procedures for preparing national instruments, and some additional quality-assurance measures were implemented to ensure international comparability. Reports from the verifiers indicated that the procedures for the translation and adaptation of the ICCS assessment and questionnaires were generally very well followed, and that the translated and adapted instruments were of high quality.

### *References*

ICCS International Study Center. (2008). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 1: School contact and material preparation*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement.

UNESCO Institute for Statistics. (1998). *Operational manual for ISCED-1997: International Standard Classification of Education*. Paris, France: Author.



---

<sup>1</sup> For more information about the ICCS quality-control procedures, please see Chapter 9.

# Sampling design and implementation

Olaf Zuehlke

## Introduction

In an international comparative survey such as ICCS, the selection of high-quality samples is critically important. Unless students and teachers are selected through the use of sound methods, the study data will lack accurate, precise, and internationally comparable estimates. ICCS followed all requirements for sampling quality specified in *Technical Standards for IEA Studies* (Martin, Rust, & Adams, 1999).

The international sample design used for ICCS was a stratified two-stage probability. During the first stage, schools were sampled with probability proportional to size. During the second stage, one intact class of target-grade students and a fixed number of target-grade teachers were randomly selected. This chapter provides a description of this sampling design, addressing in particular the following issues:

- The precise definition of the target populations of students and teachers;
- The definition of those parts of the population not covered by or excluded from ICCS;
- The international sample design;
- The intended and achieved sample sizes for students and teachers.

## Target-population definitions

When undertaking a quantitative study, it is important that researchers clearly define the target population they intend to study. Survey results from a representative sample allow one to make inferences about the group of units described by this definition. Because ICCS was designed as both a student survey and a teacher survey, two distinct target populations needed to be defined.

### *Student target population*

ICCS defined the target population of students as follows:

The student target population in ICCS consists of all students enrolled in the grade that represents eight years of schooling, counting from the first year of ISCED Level 1,<sup>1</sup> providing the mean age at the time of testing is at least 13.5 years. Students older than 17 years are not part of the target population.

For most countries, the target grade was the eighth grade, or its national equivalent. If the average age in Grade 8 was below 13.5 in a country, generally because students started formal schooling at age five, the target grade became Grade 9. To ensure international comparability, the ICCS national research coordinators (NRCs) had to specify their country's legal school entry age, the name of the target grade, and an estimate of the mean age of the students in that grade.

Students who were not covered by the definition above were regarded as “out of scope” (i.e., students in a different grade than the target grade). In the following sections, the term “students” is used to describe “students in the ICCS target population.”

In CIVED 1999, the target-population definition referred to students' age rather than their years of schooling. Therefore, for some countries participating in both surveys, different grades were assessed as a consequence of this change in the target-population definition. To obtain comparable data to those from the CIVED survey in 1999, four of these countries (Greece, Norway, Slovenia, and Sweden) surveyed students in Grade 9 in addition to those in Grade 8.



<sup>1</sup> ISCED stands for International Standard Classification of Education (UNESCO, 2006).

### *Teacher target population*

ICCS defined the target population of teachers as follows:

The teacher target population in ICCS consists of all teachers teaching regular school subjects to students of the target grade. Teachers are defined as school staff members who provide student instruction through the delivery of lessons to students. Teachers may work with students as a whole class in a classroom, in small groups in resource rooms, or one-to-one inside or outside of classrooms. Teachers who have joined a school after the beginning of the school year are not part of the target population.

This definition included all teachers teaching regular school subjects to students of the target grade (regardless of the subject or the number of hours taught) during the ICCS testing period.

School staff from the following categories were not part of the target population (i.e., were out of scope):

- Staff attending to the needs of target-grade students but not teaching any lessons (e.g., psychological counselors, chaplains, etc.);
- Assistant teachers and parent-helpers;
- Non-staff teachers teaching (non-compulsory) subjects not in the curriculum (e.g., cases where religion, although not a regular subject, was being taught by external persons).

In the following sections, the term “teachers” is used to describe “teachers of students in the target population.”

## **Coverage and exclusions**

### *Population coverage*

The ICCS international sampling team encouraged all ICCS countries to include in the study all students and teachers covered by the target population definition. However, countries could elect to remove larger groups of schools, students, and/or teachers from the target population for political, operational, or administrative reasons. This removal of schools is referred to as *reduced population coverage*.

The Slovak Republic chose this option. It restricted the study to students in schools with Slovak as the language of instruction; these constituted about 94 percent of the student population. Its national center withdrew Hungarian-language schools from the study because of the limited time available to prepare field trial instruments in Hungarian. Results from the Slovak Republic were annotated accordingly in all ICCS reports as “National Desired Population does not cover all of International Desired Population.”

### *Student exclusions*

In most ICCS countries, smaller groups of students had to be removed from the target population for practical reasons, such as difficult test conditions or increased survey costs. Such removals were regarded as exclusions. Some students were excluded because their entire school was excluded (*school-level exclusions*). Other students were excluded within sampled and participating schools (*within-sample exclusions*).

The overall exclusion rate consisted of the school-level exclusion rate (which was calculated on the basis of information provided by the NRCs) and the weighted within-sample exclusion rate (which was estimated on the basis of information collected in the sampled schools). Each country was required to keep the overall rate of excluded students below five percent of the target population. Only two participating ICCS countries exceeded this limit.



National centers were able to define those groups of schools that had to be excluded in their respective national contexts from the ICCS student survey. Within-sample exclusions could consist of students with physical or mental disabilities or students who could not speak the language of the test (typically, students with less than one year of instruction in the test language). Any other types of within-sample student exclusions were not permitted. Details about the exclusion categories for each country can be found in Appendix B of this report.

#### *Teacher exclusions*

Unlike the situation regarding the student survey, there was no intention to exclude teachers from the ICCS survey. If a teacher was part of the teacher target population, he or she was eligible to participate in the study. Therefore, no minimum exclusion rates for teachers were specified. However, teachers working at schools that were excluded did not have a chance to participate, and thus had to be regarded as excluded. Each country was asked to provide information about the proportion of teachers in excluded schools. Because statistics about teachers per grade are rarely available, some countries could not provide exact figures, but only rough estimates, or no estimates at all.

#### *Overview of exclusions*

Table 6.1 shows the population coverage and the exclusion rates for the student survey and the teacher survey for all ICCS countries.

### **Sample size requirements**

ICCS set some limits on intended sample sizes (the number or expected number of selected units) and achieved sample sizes (the number of units that actually participated in the study) for both the student and the teacher survey.

#### *Sample size in the student survey*

The overall goal of the student sample design was to achieve an *effective sample size* of at least 400 students for the main variables of interest. This meant that the complex sample design of ICCS needed to yield the same sampling precision as a hypothetical simple random sample of 400 students. Because students from the same schools tend to be more similar to one another than students from different schools, it was necessary to survey a far larger number of students than was needed to achieve this goal.

The civic-knowledge score and questionnaire scales reflecting civic-related perceptions were regarded as the main variables of interest. Given the international metric for these scales, the minimum requirements for sample precision were roughly equivalent to obtaining standard errors for civic knowledge scores that did not exceed 5.0 score points and for questionnaire scales that did not exceed 0.5 score points.

The ICCS sampling team asked, with respect to the ICCS student survey, each participating country to have a minimum intended school sample size of 150 selected schools. This meant selecting at least one intact class from each school. Once non-participation of schools and students had been taken into account, these requirements were expected to result in an achieved student sample size of roughly 3,000 tested students.

Countries with fewer than 150 eligible schools included all schools in the survey. In several countries, more than 150 schools were selected. Increases in sample size could be implemented for different reasons:

- As shown in previous student surveys, variation in student achievement across schools in a country can be large. This occurrence in the ICCS countries meant that the standards for sampling precision could only be met by increasing the school sample size;



Table 6.1: Population coverage and exclusion rates

Country	Student Survey				Teacher Survey
	Population coverage (%)	School-level exclusions (%)	Within-sample exclusions (%)	Overall exclusions (%)	Overall exclusions (NRC estimate) %
Austria	100	2.7	0.2	2.9	1.3
Belgium (Flemish)	100	2.7	0.4	3.1	8.6
Bulgaria	100	1.6	0.1	1.7	2.5
Chile	100	0.1	1.6	1.6	0.2
Chinese Taipei	100	0.4	1.5	1.9	0.9
Colombia	100	1.1	0.3	1.5	2.4
Cyprus	100	0.0	0.5	0.5	0.0
Czech Republic	100	4.6	0.1	4.7	4.6
Denmark	100	1.9	1.6	3.6	1.9
Dominican Republic	100	0.0	0.0	0.0	0.0
England	100	2.0	2.3	4.3	9.1
Estonia	100	3.8	0.0	3.8	2.5
Finland	100	2.7	1.1	3.8	2.5
Greece	100	0.6	1.4	2.0	n.a.
Guatemala	100	0.6	1.3	1.9	n.a.
Hong Kong SAR	100	1.2	0.0	1.2	6.5
Indonesia	100	0.9	0.2	1.1	1.6
Ireland	100	0.1	1.2	1.2	n.a.
Italy	100	0.1	4.4	4.5	0.2
Korea, Republic of	100	1.6	0.3	1.8	n.a.
Latvia	100	5.0	0.7	5.7	4.9
Liechtenstein	100	0.0	2.7	2.7	0.0
Lithuania	100	1.7	3.0	4.7	3.0
Luxembourg	100	1.1	0.1	1.2	9.6
Malta	100	1.3	2.4	3.7	1.0
Mexico	100	1.0	0.2	1.2	2.3
Netherlands	100	4.6	3.4	8.0	4.6
New Zealand	100	1.9	2.3	4.2	1.7
Norway	100	1.0	1.4	2.5	2.0
Paraguay	100	2.3	0.1	2.4	12.3
Poland	100	2.3	1.2	3.5	7.3
Russian Federation	100	2.9	1.9	4.8	n.a.
Slovak Republic	94.3	0.0	2.5	2.5	n.a.
Slovenia	100	1.8	3.0	4.7	1.9
Spain	100	0.4	2.2	2.6	0.8
Sweden	100	2.2	2.6	4.8	n.a.
Switzerland	100	0.8	1.2	2.0	n.a.
Thailand	100	2.7	0.3	3.0	0.7
<b>Additional grade sample</b>					
Greece	100	0.6	1.3		
Norway	100	1.0	0.9		
Slovenia	100	1.8	3.4		
Sweden	100	2.2	2.1		

**Note:** Because results are rounded to the nearest whole number, some total may appear inconsistent.



- The average class size in a country was so small that it was not possible to reach, through selection of only 150 schools, the student sample size requirement of 3,000 students;
- The NRC requested a sample-size increase in order to increase the amount of data available for analysis.

Because of non-participation, school closures, and inaccuracies in the school sampling frame, the achieved sample size of schools was smaller than the intended sample size in most of the countries.

In each sampled school, at least one *classroom* of the target grade was selected. In some countries, more than one classroom was selected. This was done because:

- The total number of schools in a country was so small that the student sample size requirements could not be met by selecting only one classroom per school;
- The NRC had asked to select two classes to allow for class-level variance analysis;
- Large sampling weight fluctuations would likely have otherwise occurred.

Each country was required to have an achieved *student* sample size of 3,000 tested students. Because of non-response, school closures, decreasing student populations, or other reasons, some countries did not meet this requirement. The ICCS sampling team did not regard this outcome as problematic as long as the country met the overall participation rate requirements.

#### *Sample size in the teacher survey*

The school size requirements for the ICCS teacher survey were the same as those for the student survey. Within each selected school, a minimum intended teacher sample size of 15 teachers was required. In schools with fewer than 15 teachers, all of the teachers were included in the survey. If the number of eligible teachers was higher than 15, but fewer than or equal to 20, all teachers were selected to prevent a situation where only a few teachers were not included in the survey. ICCS did not specify a minimum achieved teacher sample size.

Some NRCs requested all teachers in sampled schools who were teaching civic-related subjects to be part of the national teacher sample. Other countries expressed the wish to select all home-class teachers in a sampled school.

The IEA Data Processing and Research Center (DPC) developed and provided the participating countries with specialized software called *Windows® Within-School Sampling Software* (WinW3S) (IEA DPC, 2008). This gave countries the option of selecting defined groups of teachers with certainty. In those countries that did choose this option, the overall number of teachers to sample in schools was systematically increased in order to prevent the remaining groups of teachers from being under-represented in the sample.

#### *Overview of sample sizes*

Table 6.2 lists the intended and achieved school sample sizes, the achieved student sample sizes, and the achieved teacher sample sizes in the participating countries.

### **School sampling design**

The IEA DPC undertook the school sample selection for all of the ICCS countries. The DPC used as its general approach a stratified two-stage probability sampling design, in which the schools were selected systematically with probability proportional to size (PPS). The following subsections outline the school sample design for ICCS.





Table 6.2: School, student, and teacher sample sizes

Country	Originally Sampled Schools	Student Survey		Teacher Survey	
		Participating schools	Participating students	Participating schools	Participating teachers
Austria	150	135	3,385	75	999
Belgium (Flemish)	160	151	2,968	135	1,630
Bulgaria	175	158	3,257	158	1,850
Chile	180	177	5,192	177	1,756
Chinese Taipei	150	150	5,167	143	2,367
Colombia	200	196	6,204	188	2,010
Cyprus	68	68	3,194	66	906
Czech Republic	150	144	4,630	147	1,599
Denmark	240	193	4,508	113	928
Dominican Republic	150	145	4,589	145	778
England	160	124	2,916	118	1,505
Estonia	150	140	2,743	133	1,863
Finland	186	176	3,307	174	2,295
Greece	155	153	3,153	98	1,271
Guatemala	150	145	4,002	145	1,138
Hong Kong SAR	150	76	2,902	101	1,446
Indonesia	150	142	5,068	141	2,097
Ireland	165	144	3,355	137	1,861
Italy	172	172	3,366	168	3,023
Korea, Republic of	150	150	5,254	148	2,340
Latvia	160	150	2,761	146	2,077
Liechtenstein	9	9	357	9	115
Lithuania	200	199	3,902	199	2,774
Luxembourg	31	31	4,852	24	290
Malta	55	55	2,143	55	900
Mexico	220	215	6,576	202	1,844
Netherlands	150	67	1,964	22	236
New Zealand	175	146	3,979	115	1,347
Norway	150	129	3,013	73	492
Paraguay	150	149	3,399	139	1,176
Poland	150	150	3,249	150	2,081
Russian Federation	210	210	4,295	210	3,081
Slovak Republic	142	138	2,970	139	1,984
Slovenia	170	163	3,070	164	2,755
Spain	150	148	3,309	148	2,017
Sweden	175	166	3,464	156	1,942
Switzerland	187	156	2,924	144	1,571
Thailand	150	149	5263	149	1,766
<b>Additional grade sample</b>					
Greece	155	151	3,009		
Norway	150	129	2,926		
Slovenia	170	163	3,042		
Sweden	175	167	3,515		

### *Stratification of schools*

Prior to sampling, schools were *stratified*. Strata are groups of units (schools in the case of ICCS) that share some common characteristic (e.g., geographic region, urbanization level, source of financing). Generally, ICCS used stratification for the following reasons:

- To improve the efficiency of the sample design—the national centers were asked to provide stratification variables that were expected to be closely associated with students' learning-outcome variables;
- To apply different sample designs, such as disproportionate sample allocations, to specific groups of schools (e.g., states or provinces);
- To ensure adequate representation of specific groups of interest (domains) of the target population in the sample.

ICCS applied two different methods of stratification—one explicit, the other implicit.

- If explicit strata were used, the total sample of schools was apportioned to the explicit strata, and independent samples of schools were selected from each explicit stratum.
- Implicit strata were used to sort or arrange schools within explicit strata.

The combined use of implicit strata and systematic sampling is a way of ensuring a proportional sample allocation of schools across all implicit strata. Each country applied different stratification schemes after discussion with the IEA sampling team members. Appendix B of this report provides details about the stratification variables for each participant.

### *School sampling frame*

In order to prepare the selection of a sample of schools, the IEA sampling team asked national centers to provide a list of schools with students enrolled in the target grade. (A comprehensive national list of all eligible schools is called a *school sampling frame*.) The team carefully double-checked the ICCS school-sampling frames in order to ensure that they provided complete coverage of the target population and did not include incorrect entries, duplicate entries, or entries that referred to elements that were not part of the target population. The team then verified the plausibility of the information against official statistics.

For each eligible school in the sampling frame, the sampling team required the following information:

- A unique identifier, such as a national identification number;
- A measure of size (MOS) of the school, which was usually the number of students in the target grade or an adjacent grade;
- Values for each of the intended stratification variables.

### *School sample selection*

In order to select the school samples for the ICCS main survey, the sampling team used *stratified PPS* (probabilities proportional to size) *systematic sampling*. As noted earlier, this method is customary in most large-scale social surveys, and notably in most IEA surveys.

The process of selecting the school samples from each country started with sorting the school sampling frame. The team sorted it by explicit strata, then within each explicit stratum by implicit strata, and finally within each implicit stratum by MOS (alternately sorted in increasing and decreasing order).

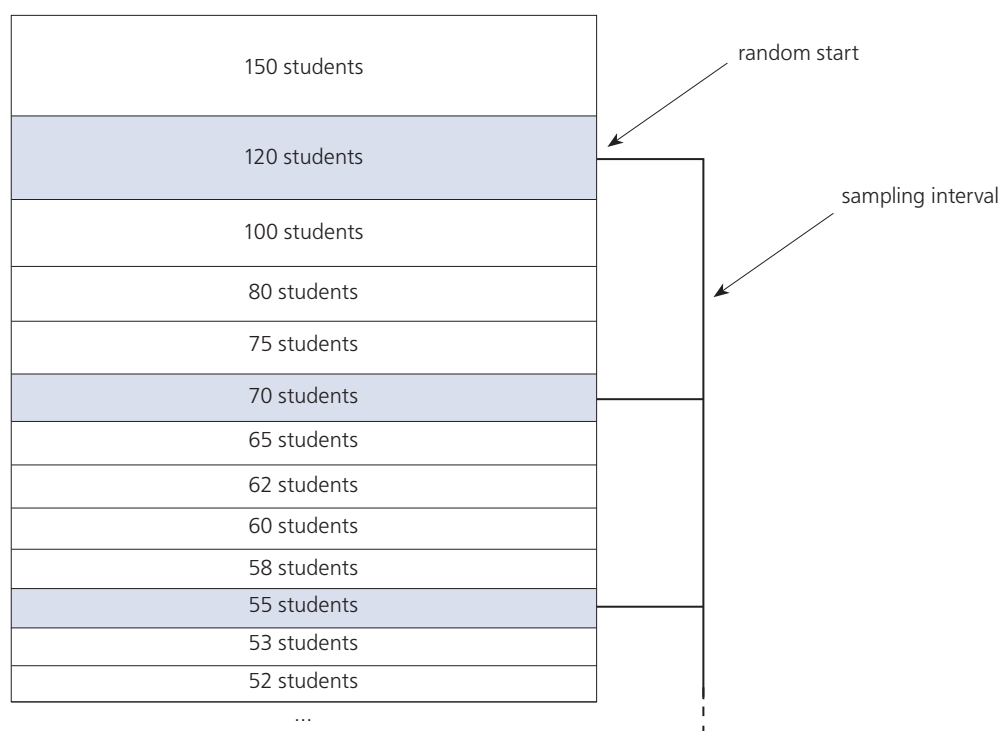


The team next selected a sample from the sorted school sampling frame by engaging in the following tasks:

- Calculating a sampling interval in each explicit stratum, a process that involved dividing the total MOS in the stratum by the number of units to be sampled from that stratum;
- Determining a random starting point in each explicit stratum, a step that decided the first sampled school in the explicit stratum;
- Selecting the units by adding the sampling interval to the point of the random start and then subsequently to each new value every time a school was selected. Whenever the cumulated MOS equaled or exceeded the corresponding value, the team selected the corresponding unit.

Figure 6.1 visualizes the process of systematic PPS sampling within an explicit stratum. In this diagram, the schools in the sampling frame are sorted in descending order by MOS. The height of the cells reflects the number of target-grade students in each school. A random start determines the second school in the list for selection, and a constant sampling interval determines the next two sampled schools. Sampled schools are displayed in blue.

*Figure 6.1: Systematic PPS sampling of schools*



The team occasionally deviated from this general procedure. If very small schools are selected with PPS, there is a risk of obtaining extremely large sampling weights for students from those schools. In order to prevent this, it is necessary to select small schools with equal selection probabilities. The ICCS team regarded a school as small if the number of students enrolled in the target grade was lower than the number enrolled in a class of average size in the school's explicit stratum. Conversely, technical problems arise whenever the MOS of a school is larger than the sampling interval. In this case, the sampling team set the MOS of the school to the sampling interval, thereby ensuring that the school would be selected with certainty but not more than once.



In order to reduce the considerable traveling costs for administering the study in the Russian Federation, the sampling team introduced an additional sampling step. This involved selecting a sample of 45 regions in a first stage, using PPS sampling. An enlarged sample of 210 schools was then selected from these regions in order to compensate for the increased sampling variance due to the additional sampling stage.

Most ICCS countries conducted an extensive field trial of the study instruments prior to the main data-collection phase. Had a school been selected for both the field trial and the main survey, this could have caused response contamination and a drop in the participation rate for the main survey. The schools, or the teachers within the schools, might then have been reluctant to participate in both the field trial and the main survey. Selecting the same school for both parts of the study was therefore avoided, whenever possible. For many countries, avoidance involved selecting the main survey sample and the field trial sample simultaneously.

The sampling team selected a sample of replacement schools at the same time that it selected the primary sample of schools. The team did this in order to maintain the sample size and reduce non-response bias in case of problems with school participation. Generally, two replacement schools with similar characteristics were assigned to each originally sampled school. The similarity was secured by selecting those two schools adjacent to the sampled school in the sorted sampling frame. The first replacement school was the one below the sampled school; the second replacement school was the one above. Schools that were part of the original sample could not be selected as replacement schools.

Because ICCS was conducted in the same year as the OECD Programme for International Student Assessment (PISA) 2009, several countries requested that the two studies not be undertaken at the same schools. The IEA DPC collaborated closely with the PISA sampling team to prevent school sample overlap whenever this was possible, but all the while the two teams worked to ensure randomness of selection and correct selection probabilities for both studies.

In all countries that decided to test an additional grade for estimating trends from CIVED 1999, the population of schools that had Grade 9 students was identical to the population of schools that had Grade 8 enrolments. This occurrence made it possible to survey the Grade 9 students in the same sample of schools selected for Grade 8. The sampling team expected that the MOS for Grade 9 for these schools would be approximately similar to the one for Grade 8 that had been used to select schools. The risk of discrepancies between the MOS for both grades and a certain loss in sample precision was viewed as being outweighed by benefits from the survey cost reductions that resulted from assessing both populations at the same schools.

### **Within-school sampling design**

Within-school sampling constituted the second stage of the ICCS sampling process. The NRCs or their appointed data managers carried out the selection of classes and teachers. The use of WinW3S software in each participating country ensured the random selection of classes and teachers within the sampled schools.

#### *Student sampling*

The sampling team used systematic random sampling to select one or more classes from each school that participated in ICCS. All participating schools were asked to list all their classes of the target grade and to provide this list to their ICCS national study center. Center staff then used WinW3S software to select the classes from these lists. Sampled classes could not be replaced or substituted. However, center staff could exclude a class from selection if it consisted solely of excluded students.



Systematic sampling was used for selecting classes from lists provided by the participating schools. This procedure was similar to the one used for systematic school sampling except that each class in a school had the same probability of being selected. Each student in a participating school had the same selection probability because all students within sampled classes were selected for participation in ICCS.

Whenever a class was smaller than half of the average class size, it was grouped with one or more other classes prior to sample selection to form a pseudo-class. This approach was used to avoid fluctuations in the total student sample size and to ensure efficient use of study resources.

### Teacher sampling

WinW3S employed systematic sampling with equal selection probabilities to select teachers from lists provided by the participating schools. In order to ensure a proportional allocation of teachers by gender, the implicit stratification was applied when using WinW3S to sample teachers.

As mentioned above, it was possible to select specific groups of teachers with certainty. The sampling team accounted for the higher selection probabilities of these teachers when conducting weight calculations.

### Summary

The ICCS student target population consisted of students enrolled in the grade that represented eight years of schooling, providing that the students' mean age at the time of testing was at least 13.5 years. The teacher target population consisted of teachers teaching regular school subjects to students of the target grade.

National centers were allowed to exclude groups of students from the study for practical reasons. However, each country was required to keep the overall rate of excluded students below five percent of the target population.

As a default, ICCS required a minimum sample size of 150 schools, in which one intact classroom and 15 teachers were selected for the study. The national samples were designed to yield a student sample size of roughly 3,000 tested students.

The international sample design was a stratified two-stage probability sampling design. Schools were selected with PPS; classes and teachers within sampled schools were selected with equal selection probabilities.

### References

- IEA Data Processing and Research Center. (2008). *Windows® within-school sampling software* (WinW3S) [computer software]. Hamburg, Germany: Author.
- Martin, M. O., Rust, K., & Adams, R. J. (1999). *Technical standards for IEA studies*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- UNESCO. (2006). *International Standard Classification of Education: ISCED 1997*. Montreal, Quebec, Canada: UNESCO Institute for Statistics.



# Sampling weights and participation rates

*Olaf Zuehlke and Caroline Vandenplas*

## Introduction

A major objective of ICCS was to obtain accurate, precise, and internationally comparable estimates of population characteristics. Several considerations had to be taken into account to achieve this goal.

This chapter begins with an outline of the definition of what constituted student or teacher participation and what constituted the requirement for within-school participation within each sampled school. Not every student or teacher who completed a survey instrument was automatically regarded as a participant in ICCS. Also, because the risk of bias greatly increases if only a minority of the sampled students or teachers in a school participate in the survey, data from affected schools were disregarded.

The next three sections of the chapter contain a description of the several sets of weights that were computed to ensure results based on ICCS data resembled those in the underlying target populations. As explained in Chapter 6, the complex sampling design of ICCS resulted in varying selection probabilities for the selected students and teachers. Furthermore, varying patterns of non-participation between strata had the potential to bias results. Both factors emphasized the need to use weighted data to achieve accurate estimates of population parameters. To this end, the IEA Data Processing and Research Center (DPC) calculated weights for all participating units in ICCS. All findings presented in ICCS reports are based on weighted data. Anyone conducting secondary analysis of the data in the ICCS database should follow this approach.

The final section of this chapter describes the participation rates at each sampling stage, the minimum acceptable participation requirements (unweighted and weighted) for students and teachers, and the categories of sample implementation quality that each country achieved. The ICCS research team regarded response rates as an important indicator of data quality. Although the team made considerable effort to ensure full participation, not all sampled units were included in the study. National samples were accordingly adjudicated with regard to sample participation requirements in the student and teacher surveys.

## Within-school participation requirements

### *Student survey participation requirements*

When the student response rate within a school is very low, the likelihood of biased results increases. There is evidence that low-performing students in particular tend to be more frequently absent from school than high-performing students. Therefore, ICCS defined a required minimum student participation rate within each school. This rate determined whether or not a school could be considered a “participant” in ICCS.

In most participating countries, only one class per school was selected for ICCS. In these countries, schools had to meet the following participation requirement:

- A sampled school was regarded as a “participating school” if, in its sampled class, at least 50 percent of its students participated in the student survey.
- If a school did not meet this requirement, it was regarded as a non-participating school in the student survey. The non-participation of this school had an effect on the school participation rate, but the students from this school were not included in the calculation of the overall student participation rate.



In a small number of countries, the selected school sample contained some schools where more than one classroom was selected. For these schools, the participation requirement was modified as follows:

- A sampled class was regarded as a “participating class” if at least 50 percent of its students participated.
- A sampled school was regarded as a “participating school” if all sampled classes participated.

In four ICCS countries (Cyprus, Liechtenstein, Luxembourg, and Malta), all of the schools in the population were selected for the study. Usually, more than one class per school was selected in these countries, and the primary sampling units were classes, rather than schools. The class participation requirement applied in these countries, too; however, if one or more classes did not participate in a school from one of these countries, the school was not automatically regarded as a non-participant.

Whenever there was an indication that the survey operation procedures in a school were not properly followed, the school was regarded as non-participant. For example, if a school had not listed all their eligible classes for class sample selection, the corresponding student data from that school were not included in the ICCS database.

#### *Teacher survey participation requirements*

Similar to the process used for the student survey, each school had to meet a minimum teacher participation requirement to be counted as participating.

- A school was regarded as a “participating school” in the teacher survey if at least 50 percent of its sampled teachers had participated.
- If a school did not meet this requirement, it was regarded as a non-participant with respect to the teacher survey.

If the survey operation procedures in a school were not followed properly, the school was regarded as non-participating. For example, if a school had not listed all of their eligible teachers for teacher sample selection, or if the teacher selection procedures had not been followed, that school’s respective teacher data were not included in the ICCS database.

### **Calculating student weights**

The ICCS student weight is a product of several weight components. Generally, it is possible to discriminate between two different types of weight components:

- *Base weights* reflect the selection probabilities of sampled units. At each level of sample selection, the base weight is the inverse of the selection probability of a sampled unit.
- *Non-response adjustments* aim to compensate the potential for bias due to non-participation of sampled units.

#### *School base weight (WGTFAC1)*

The first stage of sampling for ICCS involved selecting the schools in each country. The school base weight reflects the selection probabilities of this sampling step. When explicit stratification was used, the school samples were selected independently in each explicit stratum  $h$ , with  $h = 1, \dots, H$ . If no explicit strata were formed, the entire country was regarded as being one explicit stratum.

In most countries, ICCS drew a systematic sample of schools, with the selection probability of school  $i$  being proportional to its school size. Usually, the measure of school size  $M_{hi}$  was defined by the number of students in the ICCS target grade. If schools were small (smaller than





the average class size in the explicit stratum), the measure of size  $M_{bi}$  was defined as the average size of all small schools in that stratum.

The school base weight was defined as the inverse of the school's selection probability. For school  $i$  in stratum  $h$ , the school base weight was given by:

$$WGTFAC1_{hi} = \frac{M_b}{n_b^s \times M_{bi}}$$

where  $n_b^s$  is the number of sampled schools in stratum  $h$ ,  $M_b$  is the total number of students enrolled in the schools of explicit stratum  $h$ , and  $M_{bi}$  is the measure of size of the selected school  $i$ .

In the Russian Federation, the first sampling stage involved selection of regions. Therefore, each school weight was multiplied by a region weight component that reflected the probability of selecting that region.

#### *School non-response adjustment (WGTADJ1S)*

Given the fact that some schools refused to participate in ICCS or had to be removed from the international dataset, the school base weights had to be adjusted to account for the sample size loss. Adjustments were calculated within non-response groups defined by the explicit strata. Within each explicit stratum, a school non-response adjustment was calculated for each participating school  $i$  in stratum  $h$  as:

$$WGTADJ1S_{hi} = \frac{n_b^{s,e}}{n_b^{p-std}}$$

where  $n_b^{s,e}$  is the number of sampled eligible schools and  $n_b^{p-std}$  is the number of participating schools in the student survey in explicit stratum  $h$ .

The number  $n_b^{s,e}$  in this section is not necessarily equal to  $n_b^s$  in the preceding section, as  $n_b^{s,e}$  was restricted to schools deemed *eligible* in ICCS. Because there was a lapse of one or two years between the school sampling and the actual ICCS test, some selected schools were no longer eligible for participation in ICCS. This happened if a school had recently closed, did not have target grade students, or had enrolled only excluded students. In these cases, the ineligible school was not taken into account when calculating the non-response adjustment.

#### *Class base weight (WGTFAC2S)*

In each participating school, *Windows® Within-School Sampling Software* (WinW3S, IEA DPC, 2008) was used to randomly select one or more classes. More specifically, this process involved a systematic random method with equal selection probabilities for each class. In this sampling step, the class base weight is the inverse of the selection probability.

For each sampled class  $j$ , the class base weight was given by:

$$WGTFAC2S_{hij} = \frac{C_{bi}}{c_{bi}^s}$$

where  $C_{bi}$  is the total number of classes with eligible students enrolled in the target grade and  $c_{bi}^s$  is the number of sampled classes in school  $i$  in stratum  $h$ .





### *Class non-response adjustment (WGTADJ2S)*

In most countries, one class per school was selected for ICCS. Thus, non-response at the class level was equivalent to non-response at the school level, and any adjustments for non-response were conducted as described above. In a few countries, two classes were selected in some of the schools. If one of the two classes did not participate, the entire school was regarded as non-participating. As a consequence, the non-response adjustment was also performed at stratum level.

However, in situations where a census of schools was taken in a stratum, classes became the primary sampling units. In situations of class non-participation, a class weight adjustment was computed at the school level to correct for class non-response. The class weight adjustment for each participating class  $j$  was calculated as:

$$WGTADJ2S_{bij} = \frac{c_{bi}^s}{c_{bi}^p}$$

where  $c_{bi}^s$  is the total number of sampled classes and  $c_{bi}^p$  is the total number of participating classes in school  $i$  in explicit stratum  $b$ .

### *Student non-response adjustment (WGTADJ3S)*

Two different approaches to calculate student non-response adjustments were taken. The approach used depended on differences in non-responses between male and female students. For each country, the percentage of non-responding male students and the percentage of non-responding female students were compared. If the absolute difference in response rates between male and female teachers did not exceed three percent in a country's unweighted data, then, for all schools in this country, the adjustment for student non-response inside each class for each participating student  $k$  was calculated as follows:

$$WGTADJ3S_{bijk} = \frac{s_{bij}^e}{s_{bij}^p}$$

Here,  $s_{bij}^e$  is the number of eligible students and  $s_{bij}^p$  is the number of participating students in class  $j$  in school  $i$  in stratum  $b$ . In the context of student weight adjustment, students of the target population were regarded as eligible if they had not been excluded due to disabilities or language problems and if they had not left the sampled school after class sampling.

In Liechtenstein, the overall difference in response rates between male and female students was 4.5 percent with respect to the unweighted data. To take this difference in male and female survey participation into account, the sampling team performed, for all schools in this country, student weight adjustment within class-gender cells:

$$WGTADJ3S_{bijk} = \begin{cases} \frac{s_{bij}^{e-male}}{s_{bij}^{p-male}}, & \text{for participating male students} \\ \frac{s_{bij}^{e-female}}{s_{bij}^{p-female}}, & \text{for participating female students} \end{cases}$$

Here,  $s_{bij}^{e-male}$  and  $s_{bij}^{e-female}$  are the number of eligible males and females in the class, respectively, and  $s_{bij}^{p-male}$  and  $s_{bij}^{p-female}$  are the number of participating male and female students, respectively, in class  $j$  in school  $i$  in stratum  $b$ .

In order to allow for the calculation of weighted exclusion rates, excluded students within sampled classes received an adjustment of 1.



### Final student weight (TOTWGTs)

The final student weight of each student  $k$  in class  $j$  of school  $i$  in stratum  $b$  is the product of the five student-weight components:

$$TOTWGTs_{bjk} = WGTFAC1_{bi} \times WGTADJ1S_{bi} \times WGTFAC2S_{bij} \times WGTADJ2S_{bi} \times WGTADJ3S_{bjk}$$

Note that ICCS has no student base weight component such as  $WGTFAC3S$ . Because all students were selected for the study as soon as their classroom was selected, their within-class selection probability was 1, which means that the within-class student weight was 1 for all students in the ICCS study.

## Calculating teacher weights

### School base weight (WGTFAC1)

Because ICCS sampled the same schools for the student survey and the teacher survey, the school base weight of the teacher survey was identical to the school base weight of the student survey.

### School non-response adjustment (WGTADJ1T)

A school non-response adjustment for the teacher study was calculated in the same way as the student non-response adjustment. Because schools could be regarded as participating in the student survey but not in the teacher survey, and vice versa, the school non-participation adjustment potentially differed with respect to student data and to teacher data from the same school. To account for non-responding schools in the sample, a school weight adjustment for the teacher survey was calculated as follows for each school  $i$ :

$$WGTADJ1T_{bi} = \frac{n_b^{se}}{n_b^{p-tch}}$$

where  $n_b^{se}$  is again the number of sampled eligible schools and  $n_b^{p-tch}$  is the number of participating schools in the teacher survey in stratum  $b$ .

### Teacher base weight (WGTFAC2T)

In each school, teachers were randomly selected by the software WinW3S using a systematic random sampling method. However, in some countries, national centers chose to include all teachers of subjects related to civic and citizenship education or all home-room teachers in the teacher sample. In the following, those teachers that the country school coordinators identified for selection with certainty are referred to as *certainty teachers* and the remaining teachers (usually the majority) as *non-certainty teachers*.

The teacher base weight for each teacher  $l$  was calculated as:

$$WGTFAC2T_{bil} = \begin{cases} 1 & \text{for certainty teachers} \\ \frac{T_{bi} - T_{bi}^{cert}}{t_{bi}^s - T_{bi}^{cert}} & \text{for non-certainty teachers} \end{cases}$$

where  $T_{bi}$  is the total number of teachers,  $T_{bi}^{cert}$  is the number of certainty teachers, and  $t_{bi}^s$  is the number of sampled teachers (certainty or not) in school  $i$  in stratum  $b$ .



### Teacher non-response adjustment (WGTADJ2T)

The non-response adjustment was performed separately for certainty teachers and for sampled non-certainty teachers by computing the adjustment for each teacher  $l$  as:

$$WGTADJ2T_{hil} = \begin{cases} \frac{t_{hi}^{s,e-cert}}{t_{hi}^{p-cert}} & \text{for certainty teachers} \\ \frac{t_{hi}^{s,e-noncert}}{t_{hi}^{p-noncert}} & \text{for non-certainty teachers} \end{cases}$$

where  $t_{hi}^{s,e-cert}$  is the number of sampled eligible certainty teachers,  $t_{hi}^{p-cert}$  is the number of participating certainty teachers,  $t_{hi}^{s,e-noncert}$  is the number of sampled non-certainty teachers, and  $t_{hi}^{p-noncert}$  is the number of participating non-certainty teachers in school  $i$  in stratum  $b$ . In the context of teacher weight adjustment, teachers were regarded as eligible if they did not leave the school after teacher sampling.

If one of the adjustment cells (i.e., certainty teachers or non-certainty teachers) was empty in a school (e.g., if no certainty teachers participated in a school), the two adjustment cells within that school were combined and the adjustment was then calculated for all teachers at school level. If no certainty teachers participated, but some non-certainty teachers did, the adjustment for the participating non-certainty teachers was:

$$WGTADJ2T_{hil} = \frac{t_{hi}^{s,e-noncert} \times WGTFC2T_{hi} + t_{hi}^{e-cert}}{t_{hi}^p \times WGTFC2T_{hi}}$$

with  $t_{hi}^p$  being the number of participating teachers (all non-certainty),  $t_{hi}^{s,e-noncert}$  being the number of eligible sampled non-certainty teachers, and  $T_{hi}^{e-cert}$  being the number of eligible certainty teachers in school  $i$  in stratum  $b$ . In the standard case, where all sampled teachers within a school were eligible for ICCS, the formula was simplified as follows:

$$WGTADJ2T_{hil} = \frac{T_{hi}}{t_{hi}^p \times WGTFC2T_{hi}}$$

In situations where no non-certainty teachers participated, but some certainty teachers did, the above formulas were adapted accordingly.

As for students, in some instances a gender-specific adjustment for teachers was calculated. For each country, the percentage of non-responding male and non-responding female teachers was compared. If the absolute difference in response rates between male and female teachers did not exceed three percent in a country's unweighted data, then for all schools in that country the adjustment for teacher non-response was performed inside each group of certainty/non-certainty teachers, as described above. If the difference exceeded three percent, the teacher non-response adjustment for all schools in the country was calculated for each teacher  $l$  within the school-certainty gender-adjustment cells as follows:

$$WGTADJ2T_{hil} = \begin{cases} \frac{t_{hi}^{s,e-male}}{t_{hi}^{p-male}}, & \text{for participating male teachers} \\ \frac{t_{hi}^{s,e-female}}{t_{hi}^{p-female}}, & \text{for participating female teachers} \end{cases}$$

Here, within school  $i$  in explicit stratum  $b$ ,  $t_{hi}^{s,e-male}$  and  $t_{hi}^{s,e-female}$  are the numbers of eligible sampled male and female teachers, respectively, and  $t_{hi}^{p-male}$  and  $t_{hi}^{p-female}$  are the corresponding numbers of participants in each gender group. If one of these gender cells did not contain a valid



respondent, the gender cells were collapsed inside the groups of certainty teachers or non-certainty teachers, and then the adjustment was calculated in a similar way to that described above.

#### *Teacher multiplicity adjustment (WGTADJ3T)*

Some teachers in ICCS were teaching at the target grade in more than one school and therefore had a larger selection probability. In order to account for this, a teacher multiplicity adjustment was calculated as the inverse of the number of schools in which the teacher was teaching:<sup>1</sup>

$$WGTADJ3T_{hil} = \frac{1}{f_{hil}}$$

Here,  $f_{hil}$  is the number of schools where each teacher  $l$  in school  $i$  in stratum  $h$  was teaching.

#### *Final teacher weight (TOTWGTT)*

The final teacher weight for each teacher  $l$  of school  $i$  in stratum  $h$  is the product of the five teacher-weight components:

$$TOTWGTT_{hil} = WGTAC1_{hi} \times WGTADJ1T_{hi} \times WGTAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}$$

### **Calculating school weights**

ICCS was designed as a student and teacher survey, but not specifically as a school survey. Any statements about school level-variables have to be treated cautiously because they can be subject to large sampling errors. However, school weights were calculated for ICCS and included in the international database, in order to allow some weighted analyses of data from the school questionnaire.

#### *School base weight (WGTFAC1)*

This weight component is identical to the school base weight of the student survey and the teacher survey.

#### *School weight adjustment (WGTADJ1C)*

It is possible that some schools, for which their school principals or head teachers had not completed the school questionnaire, had participated in the student and/or the teacher survey. Consequently, there could be schools which were regarded as participants for the students and/or teacher survey but non-participants in the survey of school principals. In order to account for the non-responding school principals in the sample, a school weight adjustment component was calculated as follows for each participating school  $i$ :

$$WGTADJ1C_{hi} = \frac{n_b}{n_b^{p-sch}}$$

Here,  $n_b$  represents the number of sampled schools and  $n_b^{p-sch}$  represents the number of completed school questionnaires in stratum  $h$ .

#### *Final school weight*

The final school weight of each school  $i$  in stratum  $h$  is the product of the two weight components:

$$TOTWGTC_{hil} = WGTAC1_{hi} \times WGTADJ1C_{hi}$$



<sup>1</sup> The teacher questionnaire provided information about the number of schools a teacher was working in.

## Calculating participation rates

For ICCS, weighted and unweighted participation rates were calculated at student and teacher levels to facilitate the evaluation of data quality and the risk of potential biases due to non-response.

### Unweighted participation rates in the student survey

Let  $op$  denote the set of originally sampled eligible and participating schools,  $fp$  the full set of eligible participating schools, including replacement schools, and  $np$  the set of eligible but non-participating schools in the student survey. Let  $n^{op}$ ,  $n^{fp}$ , and  $n^{np}$  denote the numbers of schools in each of the respective sets. The unweighted school participation rate in the student survey *before* replacement can then be calculated as:

$$UPRS_{schools\_BR} = \frac{n^{op}}{n^{fp} + n^{np}}$$

The unweighted school participation rate in the student survey *after* replacement can be computed as:

$$UPRS_{schools\_AR} = \frac{n^{fp}}{n^{fp} + n^{np}}$$

The unweighted class participation rate  $UPRS_{classes}$  was 100 percent in almost all countries. In Luxembourg, one of the 283 sampled classes did not participate, which meant that the unweighted class participation rate was  $282/283 = 99.6$  percent.

Let  $sfp$  be the set of eligible and participating students in all participating schools, that is, in the schools that constitute  $fp$ , the full set of eligible participating schools. Let  $snp$  be the set of eligible but non-participating students in schools that constitute  $fp$ , and let  $s^{sfp}$  and  $s^{snp}$  be the number of students in the respective groups. The unweighted student response rate can then be computed as:

$$UPRS_{students} = \frac{s^{sfp}}{s^{sfp} + s^{snp}}$$

The unweighted overall participation rate in the student survey *before* replacement is:

$$UPRS_{overall\_BR} = UPRS_{schools\_BR} \times UPRS_{classes} \times UPRS_{students}$$

The unweighted overall participation rate in the student survey *after* replacement is:

$$UPRS_{overall\_AR} = UPRS_{schools\_AR} \times UPRS_{classes} \times UPRS_{students}$$

### Weighted participation rates in the student survey

The weighted school participation rate in the student survey *before* replacement was calculated as the ratio of summations of all participating students  $k$  in strata  $h$ , schools  $i$ , and classes  $j$ :

$$WPRS_{schools\_BR} = \frac{\sum_h \sum_{i \in op} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3 S_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGTADJ1S_{hi} \times WGT FAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}}$$

Here, the students in the numerator were computed as the sum over the originally-sampled participating schools only, whereas the students in the denominator were calculated as the total overall participating schools.



The weighted school participation rate in the student survey *after* replacement is:

$$WPRS_{schools\_AR} = \frac{\sum_b \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{bi} \times WGT FAC2S_{bij} \times WGTADJ2S_{bij} \times WGTADJ3S_{bijk}}{\sum_b \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{bi} \times WGTADJ1S_{bi} \times WGT FAC2S_{bij} \times WGTADJ2S_{bij} \times WGTADJ3S_{bijk}}$$

The weighted class participation rate is:

$$WPRS_{classes} = \frac{\sum_b \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{bi} \times WGT FAC2S_{bij} \times WGTADJ3S_{bijk}}{\sum_b \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{bi} \times WGT FAC2S_{bij} \times WGTADJ2S_{bij} \times WGTADJ3S_{bijk}}$$

The weighted student participation rate is:

$$WPRS_{students} = \frac{\sum_b \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{bi} \times WGT FAC2S_{bij} \times 1_{bijk}}{\sum_b \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{bi} \times WGT FAC2S_{bij} \times WGTADJ3S_{bijk}}$$

The weighted overall participation rate in the student survey *before* replacement is:

$$WPRS_{overall\_BR} = WPRS_{schools\_BR} \times WPRS_{classes} \times WPRS_{students}$$

The weighted overall participation rate in the student survey *after* replacement is:

$$WPRS_{overall\_AR} = WPRS_{schools\_AR} \times WPRS_{classes} \times WPRS_{students}$$

#### Overview of participation rates in the student survey

Table 7.1 displays the unweighted participation rates of all countries in the ICCS student survey. Table 7.2 displays the weighted participation rates of all countries in the ICCS student survey.

#### Unweighted participation rates in the teacher survey

Let  $op$ ,  $fp$ , and  $np$  be defined as above, such that the participation status now refers to the teacher survey instead of the student survey, and let  $n^{op}$ ,  $n^{fp}$ , and  $n^{np}$  be defined correspondingly. The unweighted school participation rate in the student survey *before* replacement can then be computed as:

$$UPRT_{schools\_BR} = \frac{n^{op}}{n^{fp} + n^{np}}$$

The unweighted school participation rate in the student survey *after* replacement can then be calculated as:

$$UPRT_{schools\_AR} = \frac{n^{fp}}{n^{fp} + n^{np}}$$

Let  $tfp$  be the set of eligible and participating teachers in schools that constitute  $fp$ ,  $tnp$  be the set of eligible but non-participating teachers in schools that constitute  $fp$ , and let  $t^{tfp}$  and  $t^{tnp}$  be the number of teachers in the respective groups. The unweighted teacher response rate can then be defined as:

$$UPRT_{teachers} = \frac{t^{tfp}}{t^{tfp} + t^{tnp}}$$



Table 7.1: Unweighted participation rates in the student survey

Country	School Participation Rate		Student Participation Rate (%)	Overall Participation Rate	
	Before replacement (%)	After replacement (%)		Before replacement (%)	After replacement (%)
Austria	82.0	90.0	92.6	75.9	83.3
Belgium (Flemish)	73.0	95.0	96.7	70.5	91.8
Bulgaria	99.4	100	95.3	94.7	95.3
Chile	97.8	99.4	96.3	94.1	95.8
Chinese Taipei	98.7	100	99.0	97.7	99.0
Colombia	93.9	99.5	95.5	89.7	95.0
Cyprus	100	100	93.4	93.4	93.4
Czech Republic	80.7	96.0	88.4	71.3	84.8
Denmark	53.1	84.6	91.9	48.8	77.8
Dominican Republic	99.3	99.3	95.2	94.5	94.5
England	51.3	78.5	93.6	48.0	73.5
Estonia	96.5	99.3	89.9	86.7	89.3
Finland	84.3	95.1	94.6	79.8	90.0
Greece	91.0	98.7	96.0	87.4	94.8
Guatemala	98.6	100	97.4	96.1	97.4
Hong Kong SAR	42.0	50.7	97.1	40.8	49.2
Indonesia	99.3	100	97.3	96.6	97.3
Ireland	82.3	87.8	91.5	75.4	80.4
Italy	93.0	100	96.6	89.8	96.6
Korea, Republic of	100	100	98.6	98.6	98.6
Latvia	88.1	93.8	90.2	79.5	84.5
Liechtenstein	100	100	97.8	97.8	97.8
Lithuania	99.0	99.5	93.2	92.3	92.8
Luxembourg*	100	100	97.3	97.0	97.0
Malta	100	100	93.4	93.4	93.4
Mexico	97.7	97.7	94.5	92.3	92.3
Netherlands	35.9	47.2	95.4	34.3	45.0
New Zealand	82.6	84.9	91.8	75.8	77.9
Norway	63.3	86.0	91.7	58.1	78.9
Paraguay	93.3	99.3	96.5	90.0	95.8
Poland	99.3	100	90.9	90.3	90.9
Russian Federation	100	100	96.6	96.6	96.6
Slovak Republic	87.9	97.9	96.4	84.7	94.3
Slovenia	91.8	95.9	93.6	85.9	89.7
Spain	97.3	98.7	91.6	89.2	90.4
Sweden	92.3	98.2	93.9	86.7	92.2
Switzerland	60.4	83.4	96.2	58.1	80.2
Thailand	73.8	100	98.2	72.5	98.2
<b>Additional grade sample</b>					
Greece	89.7	97.4	93.7	84.0	91.2
Norway	62.7	86.0	89.3	55.9	76.8
Slovenia	91.8	95.9	93.1	85.5	89.3
Sweden	92.9	98.8	92.8	85.2	91.7

**Note:** \* The unweighted class participation rate in Luxembourg is 99.6 percent.





Table 7.2: Weighted participation rates in the student survey

Country	School Participation Rate		Student Participation Rate (%)	Overall Participation Rate	
	Before replacement (%)	After replacement (%)		Before replacement (%)	After replacement (%)
Austria	82.0	90.1	92.4	75.8	83.2
Belgium (Flemish)	74.4	94.8	96.7	71.9	91.7
Bulgaria	99.1	100.0	95.4	94.5	95.4
Chile	98.3	99.4	96.2	94.6	95.7
Chinese Taipei	98.6	100	99.0	97.6	99.0
Colombia	93.2	99.5	95.3	88.8	94.8
Cyprus	100	100	93.4	93.4	93.4
Czech Republic	82.8	96.0	88.4	73.2	84.9
Denmark	53.1	84.6	91.7	48.7	77.6
Dominican Republic	99.4	99.4	95.6	95.1	95.1
England	51.6	78.5	93.8	48.4	73.6
Estonia	96.8	99.3	89.9	87.0	89.3
Finland	84.5	95.1	94.5	79.8	89.9
Greece	91.1	98.7	96.1	87.5	94.9
Guatemala	98.2	100	97.4	95.7	97.4
Hong Kong SAR	42.1	50.7	97.0	40.8	49.2
Indonesia	98.8	100	97.4	96.2	97.4
Ireland	81.8	87.4	91.6	74.9	80.1
Italy	93.2	100	96.6	90.0	96.6
Korea, Republic of	100	100	98.6	98.6	98.6
Latvia	85.8	93.4	90.9	78.0	84.9
Liechtenstein	100	100	97.8	97.8	97.8
Lithuania	99.4	99.9	94.1	93.5	94.0
Luxembourg*	100	100	97.2	96.5	96.5
Malta	100	100	93.9	93.9	93.9
Mexico	97.8	97.8	94.5	92.4	92.4
Netherlands	36.6	47.7	95.4	35.0	45.5
New Zealand	80.8	84.3	91.9	74.2	77.4
Norway	62.5	86.0	91.6	57.2	78.8
Paraguay	95.3	99.4	96.3	91.8	95.8
Poland	99.3	100	91.1	90.4	91.1
Russian Federation	100	100	96.8	96.8	96.8
Slovak Republic	87.1	97.8	96.3	83.9	94.1
Slovenia	92.5	95.9	93.9	86.9	90.1
Spain	97.1	98.7	91.9	89.2	90.7
Sweden	94.7	99.0	93.9	89.0	93.0
Switzerland	60.2	82.1	95.9	57.7	78.7
Thailand	75.2	100	98.1	73.8	98.1
Additional grade sample					
Greece	89.6	97.5	93.6	83.9	91.2
Norway	62.1	86.0	89.4	55.5	76.9
Slovenia	92.2	95.9	93.2	85.9	89.3
Sweden	95.3	99.4	92.9	88.6	92.4

**Note:** \* The weighted class participation rate in Luxembourg is 99.3 percent.



The unweighted overall participation rate in the teacher survey *before* replacement can then be computed as:

$$UPRT_{overall\_BR} = UPRT_{schools\_BR} \times UPRT_{teachers}$$

The unweighted overall participation rate in the teacher survey *after* replacement can then be calculated as:

$$UPRT_{overall\_AR} = UPRT_{schools\_AR} \times UPRT_{teachers}$$

#### Weighted participation rates in the teacher survey

The weighted school participation rate in the teacher survey *before* replacement was calculated as follows:

$$WPRT_{schools\_BR} = \frac{\sum_b \sum_{i \in fp} \sum_{l \in jfp} WGT FAC1_{bi} \times WGT FAC2T_{bil} \times WGT ADJ2T_{bil} \times WGT ADJ3T_{bil}}{\sum_b \sum_{i \in fp} \sum_{l \in jfp} WGT FAC1_{bi} \times WGT ADJ1T_{bi} \times WGT FAC2T_{bil} \times WGT ADJ2T_{bil} \times WGT ADJ3T_{bil}}$$

The weighted school participation rate in the teacher survey *after* replacement is:

$$WPRT_{schools\_AR} = \frac{\sum_b \sum_{i \in fp} \sum_{l \in jfp} WGT FAC1_{bi} \times WGT FAC2T_{bil} \times WGT ADJ2T_{bil} \times WGT ADJ3T_{bil}}{\sum_b \sum_{i \in fp} \sum_{l \in jfp} WGT FAC1_{bi} \times WGT ADJ1T_{bi} \times WGT FAC2T_{bil} \times WGT ADJ2T_{bil} \times WGT ADJ3T_{bil}}$$

The weighted teacher participation rate is:

$$WPRS_{teachers} = \frac{\sum_b \sum_{i \in fp} \sum_{l \in jfp} WGT FAC1_{bi} \times WGT FAC2T_{bil} \times WGT ADJ3T_{bil}}{\sum_b \sum_{i \in fp} \sum_{l \in jfp} WGT FAC1_{bi} \times WGT FAC2T_{bil} \times WGT ADJ2T_{bil} \times WGT ADJ3T_{bil}}$$

The weighted overall participation rate in the teacher survey *before* replacement is:

$$WPRT_{overall\_BR} = WPRT_{schools\_BR} \times WPRS_{teachers}$$

The weighted overall participation rate in the teacher survey *after* replacement is:

$$WPRT_{overall\_AR} = WPRT_{schools\_AR} \times WPRS_{teachers}$$

Table 7.3 displays the unweighted participation rates of all countries in the ICCS teacher survey, while Table 7.4 displays the weighted participation rates of all countries in the ICCS teacher survey.

#### ICCS standards for sampling participation

Despite each country's efforts to achieve participation rates as close to 100 percent as possible, higher levels of non-response were evident in a number of participating countries. As is customary in IEA studies, ICCS established guidelines for reporting data for countries with less than full participation. Three categories for sampling participation were defined, and these were applied separately to the student and the teacher survey.

Countries grouped in Category 1 met the ICCS sampling requirements. Countries in Category 2 met these requirements only after the inclusion of replacement schools. Countries in Category 3 failed to meet the ICCS sample participation requirements. During an ICCS sampling adjudication meeting in Hamburg (Germany) in December 2009, sampling referee Jean Dumais made binding decisions as to which country would be grouped in which category.



Table 7.3: Unweighted participation rates in the teacher survey

Country	School Participation Rate		Teacher Participation Rate (%)	Overall Participation Rate	
	Before replacement (%)	After replacement (%)		Before replacement (%)	After replacement (%)
Austria	44.7	50.0	73.6	32.9	36.8
Belgium (Flemish)	64.8	84.9	82.4	53.4	69.9
Bulgaria	99.4	100.0	98.9	98.3	98.9
Chile	97.8	99.4	97.2	95.0	96.7
Chinese Taipei	94.0	95.3	98.6	92.7	94.0
Colombia	89.8	95.4	91.0	81.7	86.8
Cyprus	97.1	97.1	91.0	88.3	88.3
Czech Republic	82.0	98.0	94.7	77.6	92.8
Denmark	26.3	49.6	82.1	21.6	40.7
Dominican Republic	99.3	99.3	94.5	93.9	93.9
England	48.7	74.7	89.6	43.7	66.9
Estonia	91.5	94.3	94.2	86.2	88.9
Finland	83.8	94.1	90.1	75.5	84.7
Greece	n.a.	n.a.	n.a.	n.a.	n.a.
Guatemala	98.6	100.0	99.1	97.8	99.1
Hong Kong SAR	50.0	67.3	94.9	47.4	63.9
Indonesia	98.6	99.3	91.6	90.3	90.9
Ireland	78.0	83.5	87.5	68.3	73.1
Italy	91.3	97.7	97.9	89.4	95.6
Korea, Republic of	98.7	98.7	99.7	98.4	98.4
Latvia	86.9	91.3	92.3	80.2	84.2
Liechtenstein	100.0	100.0	91.3	91.3	91.3
Lithuania	99.0	99.5	94.4	93.4	93.9
Luxembourg	77.4	77.4	81.9	63.4	63.4
Malta	100.0	100.0	98.5	98.5	98.5
Mexico	91.8	91.8	89.6	82.3	82.3
Netherlands	4.9	7.2	66.5	3.3	4.8
New Zealand	63.4	65.7	87.2	55.3	57.3
Norway	35.3	48.7	74.8	26.4	36.4
Paraguay	87.3	92.7	83.1	72.5	77.0
Poland	99.3	100.0	97.2	96.5	97.2
Russian Federation	100.0	100.0	99.7	99.7	99.7
Slovak Republic	88.7	98.6	99.2	88.0	97.8
Slovenia	92.4	96.5	91.6	84.6	88.4
Spain	97.3	98.7	96.2	93.7	94.9
Sweden	87.6	92.3	82.2	72.0	75.9
Switzerland	56.1	77.0	86.2	48.4	66.4
Thailand	73.8	100.0	99.9	73.7	99.9



Table 7.4: Weighted participation rates in the teacher survey

Country	School Participation Rate		Teacher Participation Rate (%)	Overall Participation Rate	
	Before replacement (%)	After replacement (%)		Before replacement (%)	After replacement (%)
Austria	44.5	49.2	73.8	32.8	36.3
Belgium (Flemish)	65.5	84.9	81.2	53.2	68.9
Bulgaria	98.9	100	99.2	98.2	99.2
Chile	89.7	99.5	97.7	96.4	97.2
Chinese Taipei	94.1	95.1	98.6	92.8	93.8
Colombia	87.8	95.6	92.3	81.1	88.2
Cyprus	97.1	97.1	91.0	88.3	88.3
Czech Republic	84.1	98.0	94.7	79.6	92.8
Denmark	24.8	49.6	83.8	20.8	41.5
Dominican Republic	98.9	98.9	95.4	94.3	94.3
England	49.7	74.7	89.3	44.4	66.7
Estonia	91.4	94.6	93.9	85.8	88.8
Finland	84.6	94.0	90.2	76.3	84.8
Greece	n.a.	n.a.	n.a.	n.a.	n.a.
Guatemala	97.1	100	99.0	96.1	99.0
Hong Kong SAR	49.7	67.2	95.8	47.6	64.3
Indonesia	98.7	99.3	89.8	88.7	89.2
Ireland	79.0	84.6	87.0	68.8	73.6
Italy	90.6	97.7	97.8	88.6	95.6
Korea, Republic of	98.7	98.7	99.7	98.5	98.5
Latvia	83.9	90.0	92.5	77.5	83.2
Liechtenstein	100	100	92.2	92.2	92.2
Lithuania	98.7	99.8	93.3	92.1	93.1
Luxembourg	77.4	77.4	79.9	61.8	61.8
Malta	100	100	98.9	98.9	98.9
Mexico	92.3	92.3	89.4	82.4	82.4
Netherlands	n.a.	n.a.	n.a.	n.a.	n.a.
New Zealand	63.0	65.5	87.7	55.2	57.4
Norway	37.4	48.6	72.9	27.3	35.4
Paraguay	87.1	93.2	85.3	74.3	79.5
Poland	99.5	100	96.2	95.8	96.2
Russian Federation	100	100	99.8	99.8	99.8
Slovak Republic	87.0	98.5	99.3	86.4	97.8
Slovenia	92.9	96.5	91.7	85.2	88.4
Spain	98.0	98.8	96.7	94.7	95.5
Sweden	89.3	92.5	82.7	73.9	76.4
Switzerland	56.4	75.3	85.2	48.0	64.2
Thailand	70.5	100	99.9	70.4	99.9



### *Student survey participation standards*

The categories for sampling participation in the ICCS student survey were defined according to the criteria presented in Table 7.5.

### *Teacher survey participation standards*

The sampling participation categories for the teacher survey were similar to those for the student survey. High response rates in the teacher survey were harder to achieve than in the student survey. However, there is no statistical justification for treating teacher data differently from student data with regard to an assessment of possible non-response bias, especially as teachers' motivation to participate in ICCS may have depended on the subjects they were teaching, or on their general attitude toward civic and citizenship education. Because non-response generally held a high potential for bias in both parts of the study, the participation requirements in the teacher survey were as strict as the ones in the student survey. The three categories for teacher sampling participation were defined according to the criteria set down in Table 7.6.<sup>2</sup>

### *Reporting data*

In those instances where a participating country could not be placed in participation Category 1, the ICCS research team considered it necessary to make readers of the international reports aware of the increased potential for bias. Please note that regardless of the participation category, all results were published, and no country was deleted from the international database or the international report for not having met the sample participation requirements. However, based on the sample participation categories, the survey results were reported in different ways:

- *Category 1:* Countries in this category appear in the tables and figures in the international reports without annotation.
- *Category 2:* Countries in this category are annotated in the tables and figures in the international reports.
- *Category 3:* Countries in this category appear in a separate section of the tables.

England failed to meet the requirements of the student survey, but only by a very close margin. This situation also applied to the Irish teacher survey participation. Because the data quality was not regarded as being significantly worse than for most other countries, the ICCS research team decided to include the data in the main part of the corresponding tables but to annotate these data by stating that the country “nearly satisfied guidelines for sample participation only after replacement schools were included.”

In the Netherlands, the teacher participation rate was extremely low, which made it impossible to generalize from sample data to population characteristics. Therefore, weights were not calculated, and the country was not included in the analysis of teacher data in the ICCS international reports. In Greece, unapproved teacher selection procedures were applied in the majority of schools, which made it impossible to calculate sampling weights. Therefore, it was not possible to report the country's teacher data together with the results from the other countries.

Table 7.7 lists the participation categories of each country for the student and the teacher surveys.

---

<sup>2</sup> Although the teacher survey data from the Czech Republic had satisfactory sample participation rates without the use of replacement schools, these data were erroneously annotated in the international reports as having met sample participation requirements only after the inclusion of replacement schools.



Table 7.5: Categories into which countries were placed with respect to student sampling participation rates

participation standards.

**Category 1: Satisfactory sampling participation rate without the use of replacement schools**

In order to be placed in this category, a country had to have:

- An unweighted school response rate without replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* an unweighted student response rate (after rounding) of at least 85 percent

*or*

- A weighted school response rate without replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* a weighted student response rate (after rounding) of at least 85 percent

*or*

- The product of the (unrounded) weighted school response rate without replacement and the (unrounded) weighted student response rate was at least 75 percent (after rounding to the nearest whole percentage point).

**Category 2: Satisfactory sampling participation rate only when replacement schools were included**

A country was placed in this category if:

- It failed to meet the requirements for Category 1 but had either an unweighted or a weighted school response rate without replacement of at least 50 percent (after rounding to the nearest whole percentage point)

*and had either*

- An unweighted school response rate with replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* an unweighted student response rate (after rounding) of at least 85 percent

*or*

- A weighted school response rate with replacement of at least 85 percent (after rounding to the nearest whole percent) *and* a weighted student response rate (after rounding) of at least 85 percent

*or*

- The product of the (unrounded) weighted school response rate with replacement and the (unrounded) weighted student response rate was at least 75 percent (after rounding to the nearest whole percentage point).

**Category 3: Unacceptable sampling response rate even when replacement schools were included**

Countries able to provide documentation showing that they complied with ICCS sampling procedures but did not meet the requirements for Category 1 or Category 2 were placed in Category 3.



Table 7.6: Categories into which countries were placed with respect to teacher sampling participation rates

**Category 1: Satisfactory sampling participation rate without the use of replacement schools**

In order to be placed in this category, a country had to have:

- An unweighted school response rate without replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* an unweighted teacher response rate (after rounding) of at least 85 percent

*or*

- A weighted school response rate without replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* a weighted teacher response rate (after rounding) of at least 85 percent

*or*

- The product of the (unrounded) weighted school response rate without replacement and the (unrounded) weighted teacher response rate was at least 75 percent (after rounding to the nearest whole percentage point).

**Category 2: Satisfactory sampling participation rate only when replacement schools were included**

A country was placed in Category 2 if:

- It failed to meet the requirements for Category 1 but had either an unweighted or a weighted school response rate without replacement of at least 50 percent (after rounding to the nearest whole percentage point)

*and had either*

- An unweighted school response rate with replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* an unweighted teacher response rate (after rounding) of at least 85 percent

*or*

- A weighted school response rate with replacement of at least 85 percent (after rounding to the nearest whole percentage point) *and* a weighted teacher response rate (after rounding) of at least 85 percent

*or*

- The product of the (unrounded) weighted school response rate with replacement and the (unrounded) weighted teacher response rate was at least 75 percent (after rounding to the nearest whole percentage point).

**Category 3: Unacceptable sampling response rate even when replacement schools were included**

Countries that could provide documentation showing that they complied with ICCS sampling procedures but did not meet the requirements for Category 1 or Category 2 were placed in Category 3.





Table 7.7: Participation by country in the student and teacher surveys

Country	Student Survey	Teacher Survey
Austria	1	3
Belgium (Flemish)	2	3
Bulgaria	1	1
Chile	1	1
Chinese Taipei	1	1
Colombia	1	1
Cyprus	1	1
Czech Republic	2	1
Denmark	2	3
Dominican Republic	1	1
England	2*	3
Estonia	1	1
Finland	1	1
Greece	1	-
Guatemala	1	1
Hong Kong SAR	3	3
Indonesia	1	1
Ireland	1	2*
Italy	1	1
Korea, Republic of	1	1
Latvia	1	1
Liechtenstein	1	1
Lithuania	1	1
Luxembourg *	1	3
Malta	1	1
Mexico	1	1
Netherlands	3	-
New Zealand	2	3
Norway	2	3
Paraguay	1	1
Poland	1	1
Russian Federation	1	1
Slovak Republic	1	1
Slovenia	1	1
Spain	1	1
Sweden	1	2
Switzerland	2	3
Thailand	2	2
<b>Additional grade sample</b>		
Greece	1	
Norway	2	
Slovenia	1	
Sweden	1	

**Note:** \*nearly satisfied guidelines for sample participation after replacement schools were included.



## Summary

When student or teacher response rates within a school are very low, the likelihood of biased results increases. Therefore, minimum student and teacher participation rates within each school were defined for ICCS.

Several sets of weights were computed for ICCS data. These weights reflect varying selection probabilities for the selected students and teachers as well as varying patterns of non-participation between strata. All findings presented in ICCS reports are based on weighted data, and any secondary analysis should be undertaken likewise.

Unweighted and weighted response rates were calculated for the student and the teacher surveys. In order to inform readers of the ICCS reports about the quality of sample implementation, national student and teacher samples were assigned to different categories that were determined according to the extent to which the samples had met ICCS sample

## Reference

IEA Data Processing and Research Center. (2008). *Windows® within-school sampling software* (WinW3S) [computer software]. Hamburg, Germany: Author.





# ICCS survey operations procedures

*Falk Brese and Michael Jung*

## Introduction

Successful administration of the ICCS assessment depended heavily on the contributions of the study's national research coordinators (NRCs) and national center staff. Administration of the assessment, along with the overall coordination and logistical aspects of ICCS in general, represented a significant challenge for each participating country.

The ICCS international project team<sup>1</sup> therefore developed internationally standardized survey operations procedures to assist the NRCs and to aid the synchronization of activities. The team designed these procedures to be flexible enough to simultaneously meet the needs of individual participants and the high quality of IEA survey standards. The team began by referring to the procedures developed for IEA's Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) and then tailoring these to suit the ICCS requirements.

All national centers received guidelines on survey operations procedures for each stage of the assessment, including contacting schools and sampling classes, preparing materials for data collection, administering the assessment, scoring the assessment, and creating data files. National centers also received material setting out procedures for quality control, and they were asked to complete online questionnaires that asked for feedback on survey activities.

## The role of the national research coordinators

One of the first steps that all countries or education systems<sup>2</sup> participating in ICCS took when establishing the study in their country was to appoint a NRC. The NRC acted as the contact person for all those involved in ICCS within the country. He or she also represented the country at the international level.

NRCs were in charge of the overall implementation of the study and were strongly involved in national decisions regarding ICCS. They also, where necessary, and with guidance from the international project staff and national experts (i.e., people with expertise in civic and citizenship education), implemented and adapted the internationally agreed-upon procedures for the national context.

## Documentation and software

The international project team sent the ICCS survey operations procedures to the NRCs in three units, each of which was accompanied by additional materials, including specialist manuals and software packages. All of this material was organized and distributed chronologically according to the different stages of the study.

The three units and their accompanying manuals and software packages comprised the following:

- *Unit 1: School Contact and Material Preparation* (ICCS International Study Center, 2008a).
- *Unit 2: Within-School Sampling and Test Administration* (ICCS International Study Center, 2008b).
- *Unit 3: Online Data Collection, Scoring, and Data Entry* (ICCS International Study Center, 2008c).

<sup>1</sup> This collaborative team was made up of staff from the ICCS International Study Center (ISC) at the Australian Council for Educational Research (ACER), the IEA Secretariat, the IEA Data Processing and Research Center (DPC, Hamburg), and Statistics Canada.

<sup>2</sup> The majority of the entities that participated in ICCS were countries. Some subunits of countries featuring a distinct education system also participated in ICCS. An example is Hong Kong, a Special Administrative Region of China. For reasons of simplicity, the text refers to both participating countries and education systems as "countries."



- *The School Sampling Manual* (ICCS International Study Center, 2008d): this defined the ICCS target populations and sampling goals and described the procedures to be used when carrying out sampling of schools.
- *The School Coordinator Manual* (ICCS International Study Center, 2008e): this described the role and responsibilities of the school coordinator. This person was the main contact person within each participating school. His or her responsibilities included assisting the national center in identifying classes, teachers, and students, supporting the administration of the test and questionnaires, and keeping test materials secure and confidential while they were in the school.
- *The Test Administrator Manual* (ICCS International Study Center, 2008f): this described the role and responsibilities of the test administrator. His or her work included distribution of the student test instruments according to the student tracking forms, supervising the testing sessions, ensuring the testing sessions took place within the specified times, and recording student participation.
- *The International and National Quality Control Monitor Manuals* (ICCS International Study Center, 2008g, 2008h): these provided quality control monitors (QCMs) with information about ICCS and described their role and responsibilities during the project. The manuals specified the timelines, actions, and procedures that the QCMs needed to follow in order to carry out the international and national quality assurance programs.
- *The Scoring Guides for Constructed-Response Items* (ICCS International Study Center, 2008i): these provided detailed, explicit guidelines on how to score each constructed-response item.
- *The Windows® Within-School Sampling Software* (WinW3S, IEA DPC, 2008a): this enabled the ICCS national centers to randomly select classes in each sampled school. The centers also used the software to track school, teacher, and student information, prepare the survey tracking forms, and assign test instruments to students. The software furthermore provided centers with the means to print labels for all the test booklets and questionnaires.
- *The Windows® Data-Entry Manager Software* (WinDEM, IEA DPC, 2008b): this provided a tool for entering, editing, and verifying the ICCS data. Along with the software, countries also received codebooks describing the properties and the layout of the variables to be entered from each ICCS assessment instrument.
- *The IEA SurveySystem*: this computer software enabled text passages on the paper questionnaires to be transferred to online questionnaires and these online versions to then be delivered to respondents.

In addition to its work preparing the software and manuals, the IEA DPC conducted a data-management seminar designed to train national center staff in the use of the WinDEM, WinW3S, and IEA SurveySystem software.

### Survey tracking forms

ICCS relied on six survey forms to sample classes, assign booklets and questionnaires, and track the participation status of students and teachers. These forms facilitated the data-collection and data-verification process. They also provided information on how to compute sampling weights and were used to evaluate the quality of the sampling process.

Most of these forms were created automatically by the WinW3S software. The forms were then completed by schools and returned to the national centers. The six tracking and listing forms used in ICCS were:

- *The school tracking form*: the IEA DPC sampling team sent this form to national centers and asked them to list the sampled schools and their replacements, provide the various school



identification codes (see section on “Linking” below), give a school measure of size (MOS), and provide the name of each school as well as school contact information.

- *Class listing form*: this form was created in WinW3S for each sampled school and then sent to the school coordinators for completion. The school coordinators listed the eligible target-grade classes in the participating schools and provided details about these classes, such as the number of students, “stream” or study program (if applicable), and exclusion status (in case complete classes comprised only students with physical or mental disabilities or non-native-language speakers).
- *Student listing form*: this form was also created in WinW3S and sent to each school’s school coordinator for him or her to complete prior to test administration. The school coordinators listed the eligible target-grade students in the sampled classes of the participating schools. They also provided details about the students, such as their names (if country regulations allowed national centers to give out names), birth month and year, gender, exclusion status (for students with physical or mental disabilities or non-native-language speakers), and the language of the assessment they would be using (in case the national center provided different language versions of the test booklets and questionnaires).
- *Student tracking form*: this form, also created in WinW3S, was sent to the schools with students’ test booklets and questionnaires for completion by the test administrators during test administration. The test administrators used this form to verify the assignment of assessment instruments to students and to indicate student participation and use of surplus instruments.
- *Teacher listing form*: this form was created in WinW3S and sent to the schools for completion by the school coordinator prior to test administration. The school coordinators listed the eligible target-grade teachers of the participating schools and provided details about these teachers, such as their names (if country regulations allowed for names to be sent to the national center), birth month and year, gender, and indication as to whether the teacher needed to be sampled with certainty. (The latter provided national centers with the option of increasing the teacher sample by selecting all target-grade teachers of subjects related to civic and citizenship education.)
- *Teacher tracking form*: this form was created in WinW3S and sent, with the teacher questionnaires, to the school coordinators. The school coordinators used this form to indicate the completion and return status of the teacher questionnaires.

## Contacting schools and sampling classes

Once NRCs had obtained a list of the schools sampled for ICCS (for more information on sampling procedures, refer to Chapter 6 of this report), it was important for the ongoing success of the study that the NRCs and the national centers established good working relationships with these schools. NRCs were responsible for contacting the schools and encouraging them to take part in the assessment, a process that often involved obtaining support from national or regional educational authorities, depending on the national context.

In cooperation with school principals, national centers identified and trained school coordinators for all participating schools. The school coordinator could be a teacher or guidance counselor in the school. In cases where the school coordinator also acted as the test administrator at the school, he or she was not allowed to be a teacher of the sampled class. In some cases, national centers appointed one of their own members to fill this role. Often this person was responsible for several schools in an area. Each school coordinator received a copy of the *ICCS School Coordinator Manual*. This described the coordinators’ responsibilities in detail and encouraged these individuals to contact the NRC if they had any questions.

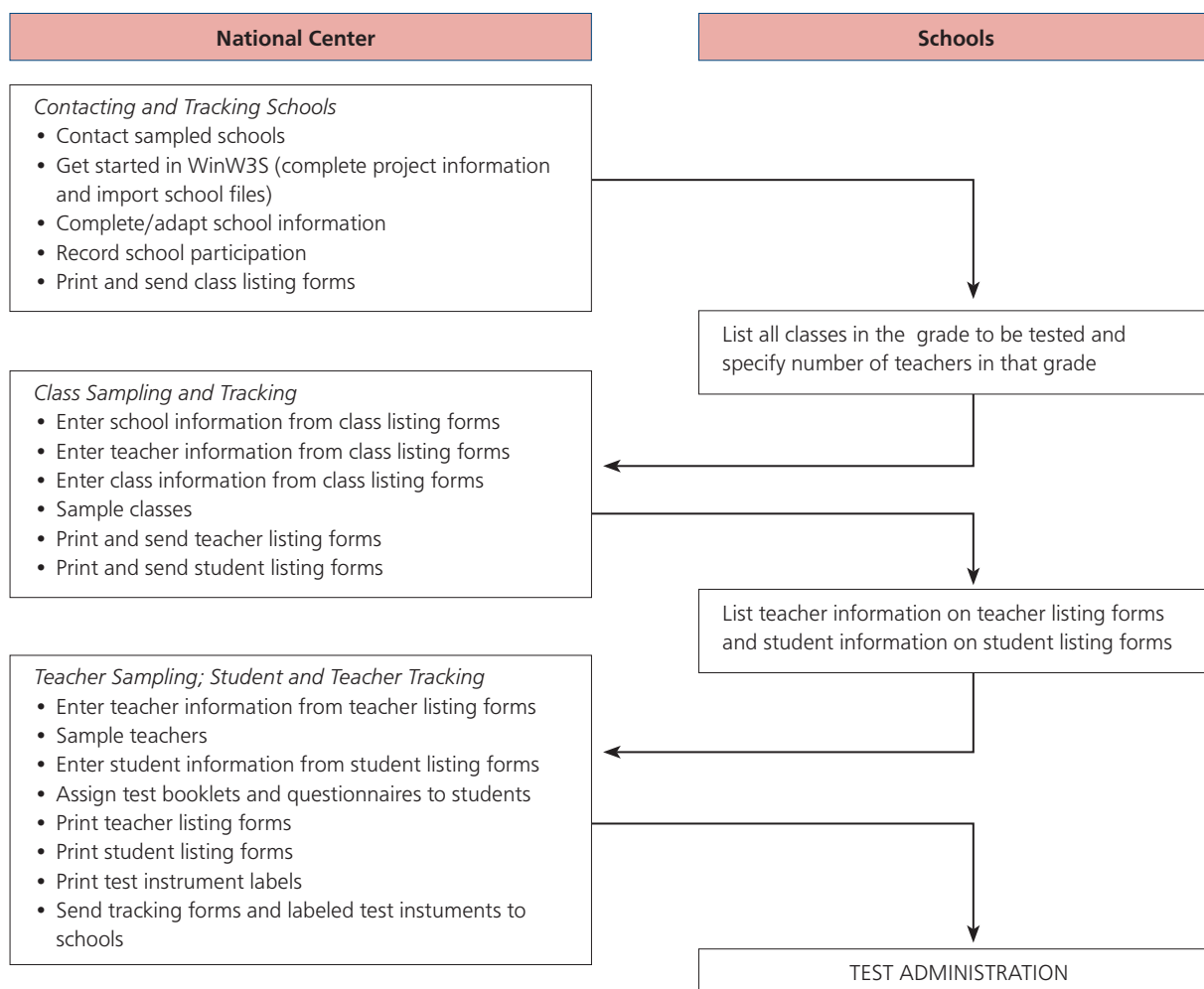


School coordinators were required to provide all specified information about their respective schools, coordinate the date, time, and place for testing, obtain parental permission (if necessary), liaise with the test administrator to coordinate the test session, distribute teacher and school questionnaires, and coordinate completion of the student tracking forms and teacher tracking forms. School coordinators also ensured that all assessment materials had been received, were kept secure at all times, and were returned to the national center after the test administration.

National centers sent a class listing form to each school coordinator and asked him or her to provide information on all the eligible target-grade classes in the school. Using this information, the national centers sampled classes within the schools. Intact classes were sampled in order to ensure that every student in the school was in only one class (course) and that no student was listed in more than one class. Ensuring that there was no overlap was a necessary requirement for obtaining a random sample of classes that was representative of all target-grade students at the school.

Figure 8.1 presents the major activities conducted by the national centers when working with schools to sample classes, to track schools, teachers, and students, and to prepare for test administration.

*Figure 8.1: Procedures for working with schools to prepare for test administration*





Although all students enrolled in the sampled classes were part of the target population, ICCS recognized that some student exclusions were necessary because of a physical or intellectual disability or in cases where there were non-native-language speakers. Accordingly, the sampling guidelines allowed the exclusion of students with any of several disabilities. (For more information on sampling procedures, see Chapter 6.) Countries were required to track and account for all excluded students and were cautioned that excluding more than five percent of students would require annotation of their results in the ICCS reports. It was important that the conditions under which countries excluded students was carefully documented, because the definition of disability could vary from country to country.

## Linking students to classes and schools and teachers to schools

The international project staff established a system to assign hierarchical identification codes (IDs). These uniquely identified and allowed tracking of the sampled schools, teachers, and classes. Table 8.1 represents the hierarchical identification system codes.

*Table 8.1: Hierarchical identification (ID) system*

Unit	ID Components	ID Structure	Numeric Example
School	School	CCCC	1001
Class	School + Class within School	CCCCKK	100101
Student	School + Class within School + Student within Class	CCCCKKSS	10010101
Teacher	School + Teacher within School	CCCCCTT	100101

Every sampled student was assigned an eight-digit identification number unique within each country. Each number consisted of the four-digit number identifying the school, followed by a two-digit number identifying the class within the school, and a two-digit number identifying the student within the class.

Each sampled target-grade teacher of the selected school (i.e., those teachers listed on the teacher tracking form) was assigned a teacher identification number consisting of the four-digit school number followed by a two-digit teacher number unique within the school.

## Preparing the test instruments for data collection

As outlined in Chapter 5, NRCs were required to document any national adaptations to the assessment instruments on the national adaptation forms (NAFs) and to submit these to the ISC for review and further discussion. The NAFs, provided in Microsoft Excel format, included all question-related texts (e.g., question stem, response options, answer categories, etc.) as well as variable names.

The ISC provided countries with all the necessary instrument production files, including fonts, style guides, graphic files, and explicit instructions on how to use the materials in order to produce good-quality test instruments. (The instructions were given in Unit 1 of the ICCS survey operations procedures.) The national centers managed the translation of the assessment instruments from English into the language(s) used in their countries and later submitted them for independent verification (see Chapter 5 of this report for details).

Following translation verification and revision of the instruments, national center staff assembled the final assessment instruments. The ISC undertook a final layout verification of the instruments, and national centers were asked to revise them, where necessary, prior to printing.



For countries administering the school and teacher questionnaires online, instrument preparation comprised an additional verification step. Countries were asked not to set up their online questionnaires until the paper-based instruments had been verified (as described in Chapter 5). Countries then used the IEA SurveySystem to set up the survey online. To ensure that data from both administration modes were comparable, the IEA DPC conducted a systematic check of the paper and online questionnaires. Apart from a few inevitable exceptions, which were necessary because of the different administration modes and which were set down for NRCs in “online adaptation notes,”<sup>3</sup> any deviations with regard to content and layout between paper and online instruments were reported back to the countries. In such cases, NRCs were requested to update their online instruments to match the paper instruments.

### Administering the ICCS assessment

Distribution of the printed materials to the schools required the national centers to engage in careful organization and planning. The labels and student tracking form produced by WinW3S allowed each sampled student to be assigned one test booklet. The test booklets were assigned in a completely balanced rotated design so that each test-item cluster within the booklets was assigned to an approximately equal number of students. Each student was also assigned a *student questionnaire*, labeled in a way that linked it to the corresponding test booklet. Depending on the country’s participation in a regional module, each student was also assigned a *regional instrument*. The materials were packaged separately for each sampled class. In addition, the *teacher questionnaires* were assigned and sent to each school for each teacher listed on the teacher tracking form. A *school questionnaire* was sent to the school principal.

For teachers and school principals who would be completing their questionnaires online, national centers prepared and sent cover letters that contained login information and instructions on how to complete the online questionnaire. National center staff sent the packaged materials to the school coordinators prior to the testing date and asked them to confirm the receipt of all instruments. School coordinators then distributed the *school questionnaire* and *teacher questionnaires* (or the cover letters for the online participants) while ensuring that the other instruments were kept in a secure room until the assessment date.

Having referred to the relevant procedures described in the *Test Administrator Manual*, national centers assigned a test administrator to each sampled class. This person’s role was to administer the test along with the *student questionnaires* and a regional instrument (where applicable). This person was chosen and trained by the national center, although, in some cases, the school coordinator also filled the test administrator role. The test administrator was responsible for distributing materials to the appropriate students, leading students through the assessment, and accurately timing the sessions. After students had completed the test, the test administrators also administered the *student questionnaire* as well as a *regional instrument* (where applicable).

The administration of the ICCS assessment consisted of either two or three parts. The first part concerned the achievement booklets. This was followed by the completion of the *student questionnaire*. If the country participated in a regional module, a regional instrument was administered afterwards. The time allotted for each part was standardized across countries.

To complete each part of the achievement test, target-grade students were allowed 45 minutes. If a student had completed the assessment before the allotted time was over, he or she could review his or her answers or read quietly but was not allowed to leave the testing room. In order to complete the *student questionnaire*, students were given at least 40 minutes, and were



<sup>3</sup> For example, in online questionnaires, respondents were redirected automatically by way of filtering rules. In contrast, the paper questionnaires required respondents to turn pages manually. Hence, instructions differed depending on the administration mode.

allowed to continue if they needed extra time. In countries participating in a regional module, students were given additional time to complete the *regional instrument*. Test administrators were required to document the starting and ending time of each part of the assessment administration on the test administration form. Table 8.2 details the time allotted to the different parts of the student assessment.

*Table 8.2: Time allowed for administering the ICCS student instruments*

Instrument	Length
Student achievement booklet	45 minutes exactly
Student questionnaire	40+ minutes
Administering the European module instrument (where applicable)	12 minutes exactly (test) 17 +minutes (questionnaire) 29 +minutes (total)
Administering the Latin American module instrument (where applicable)	15 minutes exactly (test) 15+ minutes (questionnaire) 30+ minutes (total)
Administering the Asian module instrument (where applicable)	20+ minutes

The test administrator used the student tracking form to distribute the booklets to the correct students and to document student participation. The school coordinator used the information on the participation status form to calculate the participation rate. If this was below 90 percent in any class, the school coordinator then had to hold a makeup session for the absent students before returning all of the testing materials to the national center.

The national centers entered the information recorded on the student and teacher tracking forms into WinW3S software.

## Quality control

Considerable effort was invested in developing standardized materials and procedures to ensure maximum comparability of the data collected in each country. In order to further ensure the quality of the ICCS data, an international quality control program was developed to document data collection activities around the world. The NRCs were required to nominate an international quality control monitor (IQCM) for their country. This person was then hired and trained by the IEA Secretariat. The role and responsibilities of the IQCMs were described in the *International Quality Control Monitor Manual* and included collecting and submitting a number of ICCS materials from the national centers, observing test sessions in 15 of the sampled schools in the particular country, interviewing school coordinators and/or the test administrators, and checking final assessment instruments.

The international project team also asked countries to conduct their own quality control procedures in 10 percent of the sampled schools. To assist them, the international team also provided countries with a *National Quality Control Monitor Manual*, modified to suit the national system and used for training the observers. Chapter 9 provides details on the quality-control procedures for ICCS.



## Scoring the ICCS assessment

The success of assessments containing constructed-response questions depends on the degree to which student responses are scored reliably. Six of the ICCS assessment items were constructed-response items, and it was critical to the quality of the ICCS results that they were scored in a reliable manner. This was accomplished by providing national centers with explicit scoring guides, extensive training of scoring staff, and continuous monitoring of the quality of the work during scoring procedures.

International scoring training was conducted, during which national center staff members were trained to score the constructed-response items in the ICCS assessment. Scoring training was run both before the field trial and before the main survey. The training that took place prior to the field trial provided the participants with their first opportunity to give extensive feedback on the scoring guides. The training that they received was based on a set of pilot responses collected in three English-speaking countries. The training conducted prior to the main survey enabled national center staff to give additional feedback on the scoring guides, with that feedback based on their experiences of scoring the field-trial items.

The *ICCS Main Survey Scoring Guide for Open-Ended Response Items*, the development of which is described in detail in Chapter 2, was reviewed during the scoring training preceding the main survey. The scorer training employed a sample set of student responses collected during the field trial in English-speaking ICCS countries that had already been scored.

The responses applied during scorer training were a mixture of those that clearly represented the scoring categories and those that were relatively difficult to score because they were partially ambiguous, unusually expressed, or on the “borderlines” of scoring categories. The scores that national center staff gave to these practice papers were shared with the group, with discussion focusing on discrepancies in particular. The scoring guides and practice responses were refined following the scoring training to clarify areas of uncertainty identified during the scorer training.

Once training had been completed, the ISC provided national centers with a final set of scored sample responses as well as the final version of the scoring guide. National centers used this information to train their scoring staff on how to apply the scoring guides for constructed-response items. In some cases, national centers created their own example papers and practice papers from the student responses collected in their country.

To prepare for this task, the ISC provided national centers not only with suggestions on how to organize staff but also with materials, procedures, and details on the scoring process. The ISC encouraged the national centers to hire scorers who were attentive to detail and familiar with education and who, to the greatest extent possible, had a background in civic and citizenship education. The ISC also provided guidelines on how to train scorers to accurately and reliably score the constructed-response achievement items.

## Documenting scoring reliability

In order to demonstrate the quality of the ICCS data, it was important to document the reliability of the scoring process within countries. Scoring reliability within each country required two different scorers to independently score a random sample of 300 responses for each constructed-response item. This number corresponded to 100 responses from each of the seven test booklets. The WinW3S software included a facility that allowed scorers to obtain a random sample of test booklets designated to be scored twice.

The degree of agreement between the scores, as assigned by the two scorers, provided a measure of the reliability of the scoring process. Items with relatively low inter-rater reliability



within a given country were not used in the estimation of student achievement for that country. Chapter 11 outlines the adjudication process relating to inter-rater reliability.

The ISC recommended that national centers integrate the reliability scoring with the normal scoring activity so that scorers would not be influenced by the knowledge of the context in which they were scoring (reliability or normal scoring). Scorers completed their scoring using reliability scoring sheets; the two reliability scorers were unaware of each other's scores.

### *Creating the ICCS data files*

To facilitate data entry and data verification, the IEA DPC created and distributed a software package called WinDEM (*Windows® Data Entry Manager*, IEA DPC, 2008b). The *ICCS Survey Operations Manual, Unit 3* accompanying the software provided information on installing and using this program. Because the program worked in conjunction with the WinW3S software, national center staff did not have to re-enter tracking information that had been recorded in WinW3S. (WinDEM is primarily designed for entering data from test booklets and questionnaires.) The software also offered data- and file-management capabilities, a convenient checking and editing mechanism, interactive error detection, and reporting and quality-control procedures.

The IEA DPC provided national center staff with sessions on how to use the WinW3S, IEA SurveySystem, and WinDEM software. These sessions, conducted at various stages of ICCS, also covered operational procedures on data management. In total, training included an extensive four-day seminar before the field trial and another one before the main survey data collection.

One of the most important benefits for ICCS of using WinDEM was that it incorporated the international codebooks describing all variables and their characteristics, thus ensuring that the data files produced fulfilled the ICCS rules and standards for data entry. There was one codebook for each of the questionnaires, one for the test booklets, and one for the reliability scoring sheets. The codebooks were used to inform the creation of files for entering ICCS data. However, the codebooks had to match the national instruments exactly so that the answers of the respondents could be entered properly. Therefore, any adaptations done to the international instruments also required adaptations of the international codebooks.

The adapted national codebooks were then used to create the ICCS data files within each participating country. Data from the questionnaires, achievement booklets, and reliability scoring sheets were recorded into the following WinDEM data files:

- *The school questionnaire data file;*
- *The teacher questionnaire data file;*
- *The student achievement data file* with responses from the test booklets;
- *The student questionnaire data file;* and
- *The reliability scoring file* with the codes from the constructed-response reliability scoring sheets.

For those countries participating in one of the regional modules, data from the regional instruments were recorded as follows:

- *The European module data file:* this contained responses from the European regional test and questionnaire;
- *The Latin American module data file:* this contained responses from the Latin American regional test and questionnaire;
- *The Asian module data file:* this contained responses from the Asian regional questionnaire.



Quality control throughout the data-entry process was essential for the maintenance of an accurate database. National centers were therefore responsible for performing periodic reliability checks during the data entry and for applying, prior to submission of the data files to the IEA DPC, the series of data-verification checks built into the WinDEM software package.

During this process, national centers required their data-entry staff to double-enter at least 30 units of each instrument type to ensure the reliability of the data-entry process. An error rate of 1.0 percent or less was acceptable for the questionnaire files. An error rate of 0.1 percent or less was required for the student achievement files and the reliability scoring files. If the required agreement was not reached, key punchers had to be retrained.

The data-verification module of WinDEM was also able to identify a range of problems, such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. The WinDEM software also allowed verification of the integrity of the linkage between the students, teachers, and schools entered into the WinDEM data files and tracking of information for those specified in WinW3S.

Once all data files had passed the WinDEM quality-control checks, the national centers submitted them to the IEA DPC along with data documentation for further checking and processing. Chapter 10 describes in detail the procedures that the IEA DPC used to process data.

### Online data collection for school principal and teacher questionnaires

ICCS offered participating countries the option of completing the school and teacher questionnaires online instead of using paper-based questionnaires. To ensure the comparability of the data from the online mode, only certain countries could use this mode during the main survey data collection. These countries were those that had trialed online data collection during the ICCS field trial and that had then received approval from the IEA DPC to go ahead with the online collection. Six countries administered the school and teacher questionnaires online during the ICCS main survey. They were Belgium (Flemish), Lithuania, Luxembourg, the Netherlands,<sup>4</sup> Slovenia, and Sweden.

The IEA SurveySystem developed by the IEA DPC was used to prepare and administer the online questionnaires. The IEA SurveySystem is a hierarchical model of a survey that stores and manages all questionnaire-related information, including text passages, translations and adaptations, verification rules, variable names, and data-management information. It allowed metadata to be consolidated into a single set of files that the ICCS national and international centers could then easily exchange through the internet. This feature ensured a consistent way of managing the localized online versions of the questionnaires.

To serve the different possible usage scenarios, the IEA DPC developed three distinct components of the IEA SurveySystem:

- The *Designer*: this was used to create, delete, disable, and edit survey components (e.g., questions and categories) and their properties. It allowed for translation of all text passages in the existing national paper questionnaires and additional system texts, and it included a complete web server able to verify and test-drive the survey exactly as if it were being conducted under live conditions. The Designer also supported the export of codebooks to IEA's generic data-entry software WinDEM to allow for isomorphic data entry of online and paper questionnaires.



---

<sup>4</sup> The Netherlands administered the teacher questionnaire in online mode only. Due to extremely low participation rates, it was decided not to report the teacher data from this country.



- The *Web* component: this compiled application provided respondents with questionnaires in HTML format that they could then complete from within standard internet browsers.
- The *Monitor* component: this allowed national centers to audit participation in real-time. It also allowed the centers to follow up schools that sent in incomplete questionnaires or had not returned questionnaires, and to conduct this process in a similar way to that used during administration of the paper questionnaires. The live systems were hosted on dedicated high-performance servers rented from a reliable and experienced solution provider in Germany.

To correctly sequence work steps and to ensure comparability of data, national centers were requested to finalize first paper versions of the teacher and school principal questionnaires in terms of translation and layout verification, even if the expectation was that all or nearly all of the data would be collected online. Preparation of the online questionnaires was based on the final paper versions, and this stage was followed by a final optical and content verification.

The electronic versions of the ICCS teacher and school principal questionnaires could only be completed via the internet. Respondents were not allowed to use other delivery options, such as sending PDF documents via email or printing out the online questionnaires and mailing them to the national center.

To limit the administrative burden and necessary communication with schools, national centers made the initial decision on whether to assign the online or the paper questionnaire as a default to respondents. This decision was based on the centers' and the schools' prior experience of participation in similar surveys and during the ICCS field trial. Usually, every respondent in a particular school was assigned the same mode, either online or paper. However, national centers were requested to take into account the mode that a specific school or a particular individual preferred. National centers had to ensure that every respondent assigned to the online mode by default had the option to request and complete a paper questionnaire, regardless of the reasons for not being willing or being unable to answer online.

To ensure confidentiality, every respondent received individual login information. The national centers sent this information, along with general information on how to access the online questionnaire, to respondents in the form of "cover letters." In line with the procedures used during distribution of the paper questionnaires, the school coordinator delivered this information to the designated individuals.

During the administration period, respondents could log in and out as many times as needed and resume answering the questionnaire at the question they had last responded to in their previous session. Answers were automatically saved whenever respondents moved to another question, and respondents could change any answer at any time before completing the questionnaire. During administration, the national center was available for support; the center, in turn, could contact the IEA DPC if unable to solve a problem locally.

Because the national centers were able to monitor the responses to the online questionnaires in real-time, they could send reminders to those schools where people had not responded in the expected period of time. Typically, in these cases, the centers asked the school coordinators to follow up with those individuals who had not responded.

Although countries using the online mode in ICCS faced parallel workload and complexity before and during the data collection, they had the benefit of a reduction in workload afterwards. Because answers to online questionnaires were already in electronic format, and responses were stored on servers maintained by the IEA DPC, there was no need for separate data entry.

Table 8.3 shows the (weighted) percentages of school and teacher questionnaires that were completed online.





Table 8.3: Weighted percentages of online mode administration for school and teacher questionnaires

Country	School Questionnaire	Teacher Questionnaire
Belgium (Flemish)	95.3 (2.1)	97.3 (1.2)
Lithuania	99.5 (0.4)	95.7 (1.7)
Luxembourg	100.0 (0.0)	100.0 (0.0)
Slovenia	13.7 (3.6)	27.3 (3.4)
Sweden	100.0 (0.0)	99.2 (0.6)

#### *Online data collection for survey activities questionnaire*

In order to obtain feedback about survey operations from NRCs, the international project team set up a *survey activities questionnaire* online. The questionnaire was prepared and administered using the IEA SurveySystem developed by the IEA DPC and hosted on its server. Because, unlike the other ICCS questionnaires, the survey activities questionnaire did not require national adaptations and was completed in English, it was well suited for an online data collection.

Administering this questionnaire via the internet offered many benefits for a large-scale assessment such as ICCS. Online data collection saves money and time in terms of printing and distributing the materials. Online administration also facilitates data entry, cleaning, and analysis, and responses can be directly stored in an MS SQL server.

The purpose of the survey activities questionnaire was to gather opinions and information about the strengths and weaknesses of the ICCS assessment materials (e.g., test instruments, manuals, scoring guides, and software) as well as countries' experiences with the ICCS survey operations procedures. NRCs were asked to complete these questionnaires with the assistance of their data managers and the rest of the national center staff. The information was used to evaluate survey operations, and it is now used to improve the quality of survey activities and materials used in the 2009 and future ICCS cycles.

The IEA DPC sent the NRCs individual login information for accessing the online questionnaires; internet links pointed to the location of each one. Before submitting the responses to the IEA DPC, NRCs could go back and change their answers if necessary.

#### **ICCS field trial**

The ICCS field trial was a smaller administration of the ICCS assessment. On average, approximately 600 students were tested in each participating country.

The field trial was crucial to the development of the ICCS assessment instruments, the achievement tests in particular. Items were tried out in the field trial in order to investigate the psychometric characteristics of the achievement items and to allow well-informed decision-making about further use. Except for the inclusion of items from the CIVED assessment of 1999, the field trial involved five newly developed item clusters (with 15 to 18 items in each new cluster).

The field trial also served the purpose of testing the ICCS survey operations procedures in order to avoid any possible problems during the ICCS data collection. An essential step towards achieving this goal was to conduct a full-scale field trial of all instruments and operational procedures under conditions approximating, as closely as possible, those of the main survey data collection. This process also allowed the NRCs and their staff to acquaint themselves



with the activities, refine their national operations, and provide feedback that could be used to improve the data-collection procedures. The field trial resulted in some small modifications to survey operations procedures and contributed significantly to the successful execution of ICCS. In almost all participating countries, the international field trial was conducted from October to December 2007.

## Summary

Considerable efforts were made to ensure high standards of quality in the survey procedures for the ICCS data collection. The national research coordinators (NRCs) played a key role in implementing the data collection in each participating country. All followed the internationally agreed-upon survey operations procedures.

The ISC provided NRCs with a comprehensive set of manuals containing detailed guidelines for the preparation of the study, its administration, scoring of open-ended questions, and data processing. In addition, tailored software packages were made available to national centers for the sampling and tracking of classes and teachers within schools, for data capture, and for the optional online administration of the teacher and school questionnaires.

The international ICCS field trial in 2008 was crucial for testing survey operations procedures in participating countries and contributed to their successful implementation in the final data collection.

## References

- ICCS International Study Center. (2008a). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 1, main survey: School contact and material preparation*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008b). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 2, main survey: Within-school sampling and test administration*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008c). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 3, main survey: Online data collection, scoring, and data entry*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008d). *IEA International Civic and Citizenship Education Study (ICCS) 2009 school sampling manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008e). *IEA International Civic and Citizenship Education Study (ICCS) 2009 school coordinator manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center (2008f). *IEA International Civic and Citizenship Education Study (ICCS) 2009 test administrator manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008g). *IEA International Civic and Citizenship Education Study (ICCS) 2009 international quality control monitor manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008h). *IEA International Civic and Citizenship Education Study (ICCS) 2009 national quality control monitor manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).



ICCS International Study Center (2008i). *IEA International Civic and Citizenship Education Study (ICCS) 2009 scoring guides for constructed-response items*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

IEA Data Processing and Research Center (IEA DPC). (2008a). *Windows® within-school sampling software* (WinW3S) [computer software]. Hamburg, Germany: Author.

IEA Data Processing and Research Center (IEA DPC). (2008b). *Windows® data-entry manager software* (WinDEM) [computer software]. Hamburg, Germany: Author.



# Quality assurance in the ICCS data collection

Barbara Malak and Alana Yu

## Introduction

The ICCS international research team expended considerable effort on developing standardized materials and operational procedures for ICCS (see Chapter 8 for more information) so that the data collected in each country would be comparable to the greatest possible extent. Quality control encompassed internal mechanisms built into each stage of the data-collection process to ensure the quality of the ICCS data. Quality control also encompassed external reviews administered by trained *quality control monitors* (QCMs), individuals who were separate from the staff being evaluated. Quality control of the data collection was an integral part of ICCS at both the national and international levels.

Two independent quality-assurance programs were implemented in each participating country: an international quality-assurance program conducted by the IEA Secretariat, and a national quality-assurance program managed by each national center (both programs employed similar quality-control measures). The main purpose of these programs was to provide documented evidence that the countries followed standard international data-collection procedures at all times. This chapter describes the implementation and outcomes of these programs.

The IEA Secretariat appointed *international quality control monitors* (IQCMs) for each participating country. The NRCs in each country nominated suitable people for this role and submitted their recommendation to the Secretariat. The IQCMs observed the data-collection session in a random sample of 15 schools per country, and interviewed the people responsible for this activity. Altogether, 98 IQCMs and their assistants observed 535 testing sessions in 37 of the 38 countries that participated in ICCS.<sup>1</sup> They conducted their observations and interviews according to a defined protocol, and documented their observations and the interview responses in a standard form.

Another item of documentation was the *survey activities questionnaire* for the NRCs. They were asked to relate their experiences during implementation of the ICCS survey procedures and assessment materials. They also provided information about their national quality-assurance programs.

## Quality control observations of the ICCS data collection

### *International quality-assurance program*

The program was carried out by the IEA Secretariat. A core element of this program was the appointment of one or more *International Quality Control Monitors* (IQCMs) in each country, each of whom had been nominated by the respective national centers. The IQCM had to be someone external to the national center and familiar with the school environment (e.g., a school inspector, ministry official, retired school teacher), fluent in both English and the language(s) spoken in the schools to be visited, and likely to be acceptable as an observer at the selected schools. Where necessary, the IQCMs could recruit one assistant or more in order to efficiently cover the territory and testing timetable.

All monitors participated in a training seminar conducted by the IEA Secretariat to prepare them for completing the tasks associated with the international quality-assurance program.



---

<sup>1</sup> IQCM reports were not received from the Dominican Republic.

They received an overview of the study, which included information about instrument preparation and requirements for implementation, as well as a number of documents to support their work. This material comprised the following documents:

- *ICCS 2009 International Quality Control Monitor Manual* (ICCS International Study Center, 2008a);
- The classroom observation record (on which the IQCMS registered their observations);
- Relevant sections of the *ICCS 2009 Survey Operations Procedures Units*, 1–3 (ICCS International Study Center, 2008b, c, d);
- The *ICCS 2009 School Coordinator Manual* (ICCS International Study Center, 2008e);
- The *ICCS 2009 Test Administrator Manual* (ICCS International Study Center, 2008f); and
- A manual for the administration of a regional module, where applicable (see ICCS International Study Center, 2008g, h, i).

The IQCMs received an honorarium for their work from the Secretariat, which also reimbursed IQCMs for travel expenses associated with international training and school and national center visits.

The main responsibilities for each IQCM involved consulting with the NRC to gather required information and documentation, observing and reporting on the selected survey administration sessions, and reviewing use of the translation verification results (see Chapter 5 for information on translation verification). The IQCMs were required to:

- Become thoroughly familiar with the ICCS testing procedures;
- Gather from the NRC a complete set of national survey instruments and manuals, the translation verification record, and tracking forms for the selected schools;
- Select, in consultation with the NRC and according to specified guidelines, the schools where the data collection sessions would be observed;
- Contact the school coordinator and test administrator of each selected school to organize the monitoring visit and arrange the interview;
- Observe the selected survey administration sessions for their level of adherence to the administration guidelines, in each case documenting each session's activities on a classroom observation record;
- Verify the completeness and accuracy of the lists of participating classes, teachers, and students for each visited school;
- Interview the school coordinator and test administrator, and record their responses on the classroom observation record;
- Review the national research instruments and use made of the translation verification results, and document whether the verifiers' comments were addressed appropriately; and
- Submit all collected national materials and completed observation records to the IEA secretariat.

The IEA Secretariat received documentation of the international quality assurance program from 37 of the 38 participating ICCS countries. The IQCM from the Dominican Republic did not submit the required materials. The data from the classroom observation record presented in this section thus pertains to 37 countries.



### *Classroom observation record*

The classroom observation record was used to register the monitors' observations of the ICCS testing session and interview with the school coordinator and/or test administrator. The record was organized into four sections (listed below) in order to facilitate accurate recording of the test administration's major activities:

- Section A: Preliminary activities of the test administrator;
- Section B: Survey administration activities during the testing session;
- Section C: Summary observations;
- Section D: Interview with the school coordinator and/or test administrator.

### **Preliminary activities of the test administrator**

IQCMs registered, in the first section of the classroom observation record, their observations of the condition of the survey materials, the test administrator's level of preparation, and the suitability of the testing room. Table 9.1 provides a summary of the entered information.

In general, this information confirmed the very good quality of the preparations for testing. The IQCMs were able to provide reasonable explanations for the very few cases where they observed deviations from the correct preliminary procedures. For instance, they generally attributed discrepancies between the student identification information on the instruments and the student tracking form to new students having joined the class after the tracking form was completed or to typographical errors in the list that were then corrected.

Although the IQCMs judged most test administrators as being familiar with the test administration script, they reported some cases where delay in receiving the manuals affected administrators' level of preparedness. The test administrators who did not have a watch with a second hand typically used a cell phone or classroom clock to accurately control the time of the survey administration. In general, the monitors observed no procedural deviations in the test preparations that they deemed severe enough to jeopardize the integrity of the test administration.

*Table 9.1: Percentages of IQCM responses for preliminary activities of the test administrator*

Question	Yes (%)	No (%)	Not Answered (%)
Did the test administrator verify adequate supplies of the test booklets and questionnaires prior to the students' arrival?	94	5	1
Does the student identification information on the test booklets and student questionnaires correspond with the student tracking form?	97	2	1
Did the test administrator familiarize himself or herself with the test administration script prior to the testing?	90	8	1
Was there adequate seating space for the students to work without distractions?	95	5	0
Was there adequate room for the test administrator to move around during the session to ensure that students were following directions correctly?	96	4	1
Did the test administrator have a watch with a second hand (or stopwatch) for accurately timing the sessions?	92	7	1

**Note:** Percentages given in tables are rounded.

### Survey administration activities during the testing session

Section B of the classroom observation record addressed the key activities that took place during the actual survey administration of the student achievement test, questionnaire, and regional module (if applicable). In this section, IQCMs recorded the amount of time the test administrators spent preparing the students for the sessions, the accuracy and quality of the test instruction, and the actual time spent completing the ICCS survey instruments.

In rare instances, the IQCMs observed and carefully documented significant deviations from the administration procedures. In a few cases, they noted that the survey instruments were given in a different order than prescribed. In one country, because the instruments were not administered during all monitored sessions as separately timed sessions, the IQCMs were unable to observe some aspects of the instruction and timing of the sessions.

Table 9.2 presents the results for the administration of the achievement test. One of the most important parts of the standard procedures for the assessment administration was adherence to the survey administration script. In nearly all IQCM observations, the test administrators followed the script exactly when preparing the students, distributing the materials, and giving directions and examples. Generally, when changes to the survey administration script did occur, the IQCMs considered them minor and characterized them most frequently as additions (such as emphasizing important points, demonstrating how to mark an answer on the chalkboard, or encouraging students to do their best). In one country, the IQCM observed that, on most occasions, the test administrator provided oral instructions in a language different from than that of the administration script, either in addition to or instead of reading the script directly.

In about 20 percent of sessions, the total testing time was not exactly equal to the time allowed. This was typically because students completed the test before the allotted time had elapsed, a reason that was reflected in the mean testing time of 44 minutes (one minute under the time allocated). In almost all of the sessions, students complied well with the instruction to stop work, the test administrator made sure their booklets were closed, and the instruments were secured or attended in the room during the break.

Table 9.3 presents the results of the student questionnaire administration. In similar vein to the achievement test, the IQCMs typically considered the few changes to the administration script that occurred as minor (such as emphasizing the confidentiality of answers or summarizing the instructions). About 26 percent of the test administrators reported that they did not ask students if they had completed the questionnaire after 40 minutes had passed. However, in nearly all of these cases, this was because students had already finished the questionnaire (or, in a few instances, because students were asked at the 35-minute mark). As instructed in the *Test Administrator Manual*, students received extra time to complete the questionnaire when necessary; this occurred in 37 percent of reported observations. The extra time ranged in duration from 1 to 25 minutes, with a mean time of 6 minutes.

Table 9.4 provides observations concerning the administration of the three regional instruments for ICCS that were administered as part of the following regional modules:

- Asian module (5 participating countries);
- European module (24 participating countries);
- Latin American module (6 participating countries).

Data for this section came from all countries that participated in a regional module. The only exception, as noted earlier, was the Dominican Republic. The regional instruments for Europe and Latin America included a test and questionnaire whereas the Asian regional instrument consisted of a questionnaire only. Therefore, observations on the administration of the regional test were available only for the countries participating in the European and Latin American modules.





Table 9.2: Percentages of IQCM responses for administration of the ICCS achievement test

Question	Yes (%)	No (%)		Not Answered (%)	
Did the test administrator follow the instruction exactly in each of the following tasks?		Minor changes	Major changes		
• Preparing the students	85	10	4		1
• Distributing the materials	91	6	2		1
• Giving directions	81	13	4		2
• Giving examples	81	12	4		3
If the test administrator made changes to the instruction, how would you describe them?				Not answered	Not applicable
• Additions	17	7		4	71
• Revisions	11	13		6	71
• Deletions	10	13		6	71
Did the test administrator distribute the test booklets according to the booklet assignments on the student tracking form?	97	1			2
Did the test administrator record attendance correctly on the student tracking form?	94	2			4
Did the total testing time equal the time allowed?	78	20			3
Did the test administrator announce, "You have 10 minutes left" prior to the end of the test session?	83	13			4
Were there any other "time remaining" announcements made during testing?	23	74			3
When the test administrator ended the testing session, how well did the students comply with the instruction to stop work, i.e., close their booklets and put their pens down?					5
• Very well; all students stopped work	82				
• Well; almost all students stopped work	11				
• Fairly well; some students did not stop		2			
• Not well at all; many students did not stop		0			
At the end of the testing session, did the test administrator make sure all students had closed their booklets?	90	5			5
Were the booklets left unattended or unsecured during the break?	3	94			3

As in previous sessions, the IQCMs judged the test administrators to have followed the administration script exactly in nearly all cases. Changes were generally considered minor and were most commonly characterized as deletions. When asked to explain the deletions, IQCMs said that some test administrators referred to similar instructions or to examples given earlier during administration of the international student test or questionnaire, and they asked students to read through the skipped portions silently.

In many cases, the time required to fill out the regional test was shorter than the time expected, so time-remaining announcements were not required. In about 28 percent of observations, students were granted additional time to complete the regional questionnaire. IQCMs indicated that in about 13 percent of the observations, the test administrators did not make sure that students closed their booklets at the end of the regional testing session. In some sessions, this was because students were asked to proceed directly with the regional questionnaire; in others, students were observed to have left their booklets open at the page displaying the "stop" sign. Note that the slightly higher percentages of IQCMs not responding to questions occurred because some schools administered the regional instrument on a separate date, which meant that the IQCMs could not observe these sessions.



Table 9.3: Percentages of IQCM responses for administration of the ICCS student questionnaire

Question	Yes (%)	No (%)		Not Answered (%)	
		Minor changes	Major changes	Not answered	Not applicable
Did the test administrator follow the instruction exactly in each of the following tasks?					
• Preparing the students	84	6	4		6
• Distributing the materials	89	4	2		5
• Giving directions	77	13	4		6
• Giving examples	81	6	5		7
If the test administrator made changes to the instruction, how would you describe them?					
• Additions	12	8		4	76
• Revisions	8	11		6	76
• Deletions	9	9		7	76
Were students asked after 40 minutes if they had all completed the questionnaire?	69	26		5	
Was additional time allowed?	37	58		5	
Were the questionnaires collected and secured after the questionnaire session?	92	3		6	

### Summary observations

The IQCMs provided, in Section C of the classroom observation record, their general impressions of how the testing session was conducted, how well the test administrator monitored students, and any unusual circumstances that arose during the session (e.g., students refusing to participate, defective instruments, cheating). The results presented in Tables 9.5 and 9.6 show that, for most testing sessions, the IQCMs observed no major problems.

Table 9.5 reports the IQCMs' general observations of student behavior and the quality of the administration session. In nearly all instances, the IQCMs considered the students orderly and cooperative. Ninety-one percent of the IQCMs described the overall quality of the observed sessions as "excellent," "very good," or "good."

Table 9.6 presents various other observations made by the IQCMs. Occasionally, they noted that while space constraints in the survey rooms prevented the test administrators from walking around the class, they were still able to monitor students from the front of the room. In almost all cases, the IQCMs considered that the test administrators had addressed students' questions appropriately. In 15 percent of cases, IQCMs reported evidence of students attempting to cheat. However, in many of these instances, the IQCMs characterized the situation as "communicating" rather than cheating, explaining that students seemed curious about how their classmates had responded to items in the questionnaire. They also said that test administrators intervened when necessary. Because the ICCS test design involved seven different achievement booklets, students were unlikely to have had the same booklet as their neighbors. In the few sessions where defective instruments were detected, the test administrator almost always replaced the booklet appropriately.

There were very few reports of students refusing to take the survey, and when this did occur, the students were typically observed to have ended the survey early (due, for instance, to a prior appointment, illness, or presumed lack of interest in continuing the assessment). More commonly (in 18% of observations), IQCMs reported that students briefly left the room during the session. On nearly all of these occasions, the test administrators responded appropriately by collecting the booklet; in some sessions, the booklet was left closed on the student's desk until the student returned to class.



Table 9.4: Percentages of IQCM responses for administration of the ICCS regional module

Question	Yes (%)	No (%)		Not Answered (%)	
Did the test administrator follow the instruction exactly in each of the following tasks?		<i>Minor changes</i>	<i>Major changes</i>		
• <i>Preparing the students</i>	85	4	4	7	
• <i>Distributing the materials</i>	88	3	2	7	
• <i>Giving directions</i>	76	11	6	7	
• <i>Giving examples</i>	77	8	8	8	
If the test administrator made changes to the instruction, how would you describe them?				<i>Not answered</i>	<i>Not applicable</i>
• <i>Additions</i>	8	12		4	76
• <i>Revisions</i>	8	12		4	76
• <i>Deletions</i>	12	8		4	76
Did the test administrator record attendance correctly on the student tracking form?	89	3		8	
Did the total testing time equal the time allowed?*	58	36		7	
Did the test administrator announce, "You have one minute left" prior to the end of the test session?*	52	42		6	
Were there any other "time remaining" announcements made during testing?*	9	85		6	
When the test administrator ended the test session, how well did the students comply with the instruction to stop work, i.e., close their booklets and put their pens down?*				7	
• <i>Very well; all students stopped work</i>	84				
• <i>Well; almost all students stopped work</i>	7				
• <i>Fairly well; some students did not stop</i>		1			
• <i>Not well at all; many students did not stop</i>		0			
At the end of the testing session, did the test administrator make sure all students had closed their booklets?*	79	13		8	
Did the test administrator accurately read the script and give directions for the questionnaire?	69	<i>Minor changes</i>	<i>Major changes</i>		
		11	8	12	
If there were changes, how would you describe them?				<i>Not answered</i>	<i>Not applicable</i>
• <i>Not answered</i>					
• <i>Additions</i>	7	10		2	81
• <i>Deletions</i>	9	6		4	81
Was additional time allowed?	28	63		9	
Were the questionnaires collected and secured after the questionnaire session?	90	1		10	
At the end of the session, prior to dismissing the students, did the test administrator thank the students for participating in the study?+	85	7		8	

**Notes:**

\* Among the countries that participated in the European or Latin American module.

+ Among all ICCS participating countries.



Table 9.5: Percentages of IQCM responses for observations of student behavior

Question	Extremely (%)	Moderately (%)	Somewhat (%)	Hardly (%)	Not Answered (%)
To what extent would you describe the students as orderly and cooperative?	56	34	9	1	0

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered (%)
In general, how would you describe the overall quality of the survey administration session?	41	36	14	7	1	1

#### *Interview with the school coordinator and/or test administrator*

The purpose of the IQCMs' interviews with the school coordinators and/or test administrators was to solicit these individuals' evaluations of the ICCS survey administration, gather suggestions for improvement, and obtain additional background information on survey-related activities. The latter included the shipment of assessment materials, arrangements for test administration, how readily NRCs responded to queries, and organization of classes in the school. The IQCMs recorded the results of these interviews on Section D of the classroom observation record. Tables 9.7 and 9.8 present summaries of this information.

Overall, the school coordinators and test administrators expressed a favorable impression of the ICCS survey, reporting that it went well with few problems. Their reports also suggest that school staff members held mostly positive attitudes toward the survey. These results are shown in Table 9.7.

Table 9.8 presents the other major outcomes of the IQCM interviews. The table shows that instances of the school coordinator and/or test administrator not receiving the correct shipment of materials were rare. Not receiving the material in time to check it for possible defects (the time available ranged from weeks to days) was also rare. The items most frequently reported as missing were the manuals and envelopes/boxes for returning the materials after the assessment; however, these occurrences did not exceed 12 percent of observations. (As a point of comparison, booklets and questionnaires were missing in about three to four percent of cases.) The survey materials were almost always stored in a secure location at the school (such as a locked office or safe), or kept with the test administrator.

In order to better estimate the time required for completing the teacher questionnaire, school coordinators were asked if the anticipated time of 30 minutes was sufficient. Some of the school coordinators were unable to collect the completed questionnaires before the test administration. About 8 percent of the coordinators reported that the questionnaire required more time to complete; 14 percent reported that it required less time. School coordinators were also asked about the quality of the *School Coordinator Manual*. Only seven percent of them said that the manual "needed improvement;" in these cases, the most frequent suggestions were for more clarity and less repetition.

About 55 percent of the school coordinators noted that students received some kind of special instruction, motivational talk, or incentive to participate. This effort typically consisted of a talk by the school principal or class teacher to inform students about the study and to motivate them to do their best. Some schools sent a letter to the students' parents. In a few instances, schools gave students small gifts during the break or after the assessment.



Table 9.6: Percentages of IQCM responses for general observations

Question	Yes (%)	No (%)	Not Answered (%)	
During the administration sessions, did the test administrator walk around the room to be sure students were working on the correct section of the instrument and/or behaving properly?	93	7	0	
Did the test administrator address students' questions appropriately? <sup>1</sup>	98	2	0	
Did you see any evidence of students attempting to cheat (e.g., by copying from a neighbor)?	15	85	0	
Were any defective test booklets and/or questionnaires detected and replaced before the testing session began?	4	93	2	
Were any defective test booklets and/or questionnaires detected and replaced after the testing began?	2	94	3	
If any defective test booklets and/or questionnaires were replaced, did the test administrator replace them appropriately?	4	1	Not answered 0	Not applicable 96
Were any late students admitted to the survey administration room? • No, there were no late students • No, they were not admitted • Yes, but before the testing began • Yes, after the testing began	8 6	85 2		
Did any students refuse to take the survey either prior to or during the survey administration?	2	96	1	
If a student refused, did the test administrator accurately follow the instructions for excusing the student (collect the instrument and record the incident on the student tracking form?)	1	0	Not answered 1	Not applicable 98
Did any students leave the room for an "emergency" during the session?	18	81	2	
If a student left the room for an emergency during the session, did the test administrator address the situation appropriately (collect the test booklet or questionnaire and, if the student was readmitted, return the instrument)?	13	5	Not answered 0	Not applicable 82

**Note:**

<sup>1</sup> Test administrators were instructed not to answer any questions about the content of the survey questions. They were permitted, however, to answer questions about what was generally required from respondents and how they should record their answers.

In order to validate the within-school sampling procedures, IQCMs asked the school coordinators about the selection of classes and students in the schools. Almost 90 percent of the coordinators reported having achieved accurate class and student sampling information. Obtaining accurate teacher sampling information proved to be a more difficult task. (In one country, for example, confidentiality rules forbade the use of teachers' names on the list.) However, about 71 percent of cases managed to achieve accurate information.

In about 14 percent of observations, school coordinators anticipated that a test make-up session would be required, and most said they would conduct it themselves. Nearly all of the school coordinators said they would be willing to repeat their roles in another international assessment, with many remarking that the experience was enjoyable and positive overall.



Table 9.7: Percentages of IQCM responses from interviews with the school coordinator and/or test administrator: overall impressions

Question	Very Well, No Problems (%)	Satisfactorily, Few Problems (%)	Unsatisfactorily, Many Problems (%)	Not Answered (%)
Overall, how would you say the ICCS survey administration went?	76	21	2	1
Question	Positive (%)	Neutral (%)	Negative (%)	Not Answered (%)
Overall, how would you rate the attitude of the other school staff members toward the ICCS survey administration?	65	31	3	1

### Survey activities questionnaire

The IQCMs used the survey activities questionnaire to gather information from the NRCs about whether the implementation of survey procedures accorded with the standards outlined in the *Survey Operations Procedures* manual. The IQCMs also used the questionnaire to solicit feedback from the NRCs on the strengths and weaknesses of the approaches and materials used. The topics covered included sampling, contacting schools, recruiting school coordinators, translating and preparing the survey instruments, administering the assessment, implementing the national quality assurance program, scoring open-ended response items, and entering and submitting data. Another major purpose of the questionnaire was to gather information that could be used to improve the quality of future IEA surveys.

Data were collected online from each NRC personally, with the data manager and/or other national center staff assisting this process where necessary. The results of the questionnaire, presented in this section, reflect the quality of the ICCS procedures and materials in all 38 participating countries.

#### Sampling

The first part of the survey activities questionnaire collected information on the sampling procedures and manuals. Table 9.9, which provides a summary of this information, indicates that the sampling process worked well overall. In most countries, the NRCs reported no difficulties adapting the international sampling design to national specifications or compiling a sampling frame (a list of eligible schools). Among those national centers that experienced some level of difficulty, all felt well supported by the IEA Data Processing and Research Center's (DPC) sampling team. Nearly all countries indicated that the *Sampling Manual* sufficiently described the relevant processes and procedures, and that they were able to use the *Windows® Within-School Sampling Software* (WinW3S) provided by the IEA DPC (2008a) to select classes and teachers. In 18 countries, data protection laws required that numbers rather than names be used on student and/or teacher lists.

Six NRCs reported encountering organizational constraints that required deviations from the standard ICCS within-school sampling design. In two countries, school and/or teacher sampling were deemed unnecessary because the ICCS sample included the entire target population. Other deviations included administering the assessment to Grade 9 students at the beginning of the school year (instead of at the end of the Grade 8 school year) and some schools selecting two classes instead of one. In each of these cases, the deviation was documented and a sampling expert was consulted to ensure that the altered design met all sampling requirements.



*Table 9.8: Percentages of IQCM responses from interviews with the school coordinator and/or test administrator: receipt of materials and test administration*

Question	Yes (%)	No (%)		Not Answered (%)	
Prior to the administration day, did you have time to check the shipment of materials from your ICCS national research coordinator?	84	6		10	
Did you receive the correct shipment of the following items?					
• <i>School Coordinator Manual</i>	85	10		4	
• <i>Test Administrator Manual</i>	84	9		7	
• <i>Student tracking forms</i>	91	2		7	
• <i>Test booklets</i>	89	4		7	
• <i>Student questionnaires</i>	90	3		7	
• <i>Regional module instruments<sup>1</sup></i>	89	3		7	
• <i>Teacher questionnaires</i>	93	2		4	
• <i>School questionnaires</i>	93	2		4	
• <i>Test administration forms</i>	89	3		8	
• <i>Teacher tracking forms</i>	87	6		8	
Envelopes/boxes addressed to the national center for the purpose of returning the materials after the assessment	83	12		5	
Was the national research coordinator responsive to your questions or concerns?	84	4		12	
It was expected that the teacher questionnaire would require about 30 minutes to complete. Was that estimate correct?	58	More time 8	Less time 14	19	
Were you satisfied with the accommodation (testing room) you were able to arrange for the survey administration?	96	3		1	
Do you anticipate that makeup sessions will be required at your school?	14	77		8	
If you anticipate that a makeup session will be required, do you intend to conduct one?	11	3		Not answered 1	Not applicable 86
Did the students receive any special instructions, motivational talks, or incentives to prepare them for the assessment?	55	44		1	
Overall, do you feel the <i>ICCS School Coordinator Manual</i> worked well?	80	7		12	
Is this a complete list of the classes in this grade in this school?	89	8		3	
To the best of your knowledge, are there any students in this grade level who are not in any of these classes?	9	88		2	
To the best of your knowledge, are there any students in this grade level in more than one of these classes?	3	94		2	
Is this a complete list of the teachers teaching this grade in this school?	71	22		7	
If there was another international assessment, would you be willing to serve as a school coordinator?	87	10		3	

**Note:**

<sup>1</sup> Among countries that participated in a regional module.





Table 9.9: Numbers of NRC responses to the survey activities questionnaire: sampling

Question	Yes		No	Not Answered
	Very difficult	Somewhat difficult		
Was it difficult to:				
• adapt the international sampling design to your national specifications?	0	7	31	0
• compile a list of all eligible schools?	0	3	34	1
Did the sampling manual sufficiently describe the following procedures?				
• Defining and identifying the target population of the survey	38		0	0
• Creating a sampling frame	38		0	0
• Selecting a sample at the IEA DPC	37		1	0
Were there any conditions or organizational constraints that required deviations from the standard within-school sampling design?	6		32	0
Did you use numbers instead of names to identify students and/or teachers on the forms and labels due to data protection/confidentiality laws or rules in your country?	18		20	0
Did you use the WinW3S software to sample classes and teachers?	37		1	0

#### Contacting schools and recruiting school coordinators

Table 9.10 provides the NRCs' responses to questions about school participation and the school coordinators. Generally, countries used a variety of different materials to request school participation in the main survey, including sample letters provided in the *Survey Operations Procedures* manual and letters from relevant education ministries. A number of NRCs reported difficulty convincing the selected schools to participate. Common reasons cited were logistical issues (timing, availability of students and staff), concerns about overburdening students and teachers, and the sensitive nature of civic and citizenship education in some contexts.

Nearly all school coordinators for ICCS, typically chosen from among the school principals and head teachers, received written materials (such as manuals and letters) designed to instruct them in their roles. If necessary, the coordinators could also contact the national centers by email to ask questions and clarify instructions. In 17 countries, national center staff provided formal training sessions.

Table 9.10: Numbers of NRC responses to the survey activities questionnaire: contacting schools and recruiting school coordinators

Question	Yes	No	Not Answered
What materials did you use as the basis for requesting school participation in the main survey? (Select one)			0
• Example letters provided in Appendix A of the Survey Operations Procedures, Unit 1	14		
• Letters based on other national projects	15		
• Other	9		
Did you have any difficulties in convincing schools to participate?	22	16	0
How did you train the school coordinators? (Select all that apply)			0
• Formal training sessions	17		
• Through telephone, email, or video-link	21		
• Written instructions	36		
• Other	3		



### *Adapting and translating the ICCS assessment materials*

In this section of the *survey activities questionnaire*, NRCs provided information about the process of adapting and translating the ICCS achievement test, background questionnaires, and regional modules into national languages and contexts. Most NRCs identified national center staff as the individuals responsible for adapting and/or translating the survey instruments; in just over a third of all participating countries, the national centers also consulted, and then often worked with, outside specialists to conduct the translation. When NRCs were asked about their experiences with the external-adaptation negotiation and translation-verification processes organized by the ISC and the IEA Secretariat (respectively), a small number of them said that they did experience difficulties, mainly with respect to tight timeframes and reaching agreement on the adaptation of certain specialist concepts. Table 9.11 presents a summary of the information collected in this section of the questionnaire.

*Table 9.11: Numbers of NRC responses to the survey activities questionnaire: adapting and translating the ICCS assessment materials*

Question	Adaptation	Translation	Not Applicable
Who adapted and/or translated the international version of the student test? (Select all that apply)			0
• Own staff	30	19	
• Outside translator(s)	2	16	
• Outside reviewer(s)	5	4	
• Combination of the above	9	11	
Who adapted and/or translated the international version of the student, teacher, and principal questionnaires? (Select all that apply)			0
• Own staff	31	18	
• Outside translator(s)	2	14	
• Outside reviewer(s)	4	4	
• Combination of the above	10	12	
Who adapted and/or translated the international version of the regional module instruments? (Select all that apply)			3
• Own staff	27	17	
• Outside translator(s)	3	12	
• Outside reviewer(s)	1	3	
• Combination of the above	8	10	
Question	Yes	No	Not Answered
Did you have major problems regarding the process of external adaptation verification and/or translation verification of the instruments?			
• Adaptation verification	6	31	1
• Translation verification	6	30	2

### **Assembling and printing the ICCS assessment materials**

Table 9.12 shows that, as with the first two rounds of external verification, few countries experienced difficulties when completing the third step of the verification process—layout verification (this was overseen by the ISC). Tight timelines posed the biggest challenge in two countries. In general, NRCs reported no difficulties assembling the various survey instruments, although they detected, in a number of cases, printing errors, such as poor print quality, missing pages, wrong page order, and unintentional upside-down pages. In most cases, these errors were minor and were resolved in a straightforward manner. All cases where errors could not be corrected were carefully documented.



Table 9.12: Numbers of NRC responses to the survey activities questionnaire: assembling and printing the ICCS assessment materials

Question	Yes	No	Not Answered
Did you have major problems regarding the process of external layout verification of the instruments?	4	34	0
Did you detect any of the following errors during the printing process: poor print quality, missing pages, wrong page order, upside-down pages?	19	19	0
Did you discover any potential breaches of security?	2	36	0

### Administering the booklets and questionnaires

This section of the *Survey Activities Questionnaire* addressed some important aspects of the data-collection session; Table 9.13 presents selected results. All countries had the option of administering their school and teacher questionnaires online, although only six used this delivery method. The majority of countries reported that at least some of their test administrators were drawn from national center staff, but it was also common practice to recruit test administrators from among the school coordinators or teachers of the sampled schools (but not teachers of the sampled students).

Table 9.13: Numbers of NRC responses to the survey activities questionnaire: administering the booklets and questionnaires

Question	Yes	No	Not Answered
Did you use the online data collection mode for administering the school and/or teacher questionnaires?	6	30	2
Who were the test administrators for the main survey? (Select all that apply)			0
National center staff	22		
• Regional or district government staff	4		
• External contractor staff	11		
• Teachers from other schools	3		
• Teachers from the sampled schools but not of the sampled students	18		
• Teachers of the sampled students	2		
• School coordinators	16		
• Other	4		
How did you train the test administrators? (Select all that apply)			1
Formal training sessions	27		
• Through telephone, email, or video-link	13		
• Written instructions	32		
• Other	2		
For the administration of (paper-based) instruments, did you experience any difficulties reaching a high participation of the following groups?			
• Students	9	29	0
• Teachers	18	20	0
• School principals	9	29	0



As was the case with school coordinator training, most countries used more than one approach to train their test administrators. The main reliance was on written instructions (32 countries) supplemented with formal training sessions (27 countries) or email/telephone correspondence (13 countries). A number of NRCs reported difficulties reaching high participation rates for students (9 countries), teachers (18 countries), and high school principals (9 countries).

Some schools had problems with truancy, and logistical issues meant that it was not always possible to schedule a make-up session for absent students. NRCs reported lack of time, interest, and/or willingness to answer questions about political topics as the reasons that teachers and principals most frequently gave for non-response.

#### *National quality control monitoring*

Each national center received materials to help them conduct their own national quality assurance programs. The *ICCS 2009 Survey Operations Procedures, Unit 2* provided guidelines for selecting the national quality control monitors (NQCMs), as well as basic information about the monitors' duties. The *ICCS 2009 National Quality Control Monitor Manual* (ICCS ISC, 2008j) was developed to assist training. The ISC encouraged national centers to amend the manual and the provided classroom observation record so that these reflected any matters of special importance in their respective countries. The NQCMs' responsibilities were similar to those of the IQCMs, in that they involved visiting selected schools to observe and document the data-collection session.

According to the information given by the NRCs in the survey activities questionnaire, a total of 191 NQCMs and their assistants visited 536 schools in the participating countries. The NQCMs confirmed the good quality of the surveying process overall. Like the IQCMs, they detected several problems, such as defective survey materials, test administrators making errors, and unmotivated students. However, the issues that they reported were generally minor and were resolved promptly and appropriately (e.g., calling schools to verify the correct shipment of instruments, rescheduling the survey due to a teachers' strike, and shifting the European module to another day to cope with test fatigue among students). One school prematurely opened a box of questionnaires, which meant that this school's data were not included in the database. As evident in Table 9.14, only one country (Slovenia) did not use the provided support materials because it did not implement a national quality assurance program. (Slovenian national center staff conducted the test administration.)

*Table 9.14: Numbers of NRC responses to the survey activities questionnaire: national quality control monitoring*

Question	Yes	No	Not Answered
Did you use the national quality control monitor templates (manual and classroom observation record) provided by the international study center?	37	1	0

#### *Scoring open-ended response items and coding occupation data*

NRCs were also asked to comment on the persons responsible for scoring open-ended response items, the procedures used to verify scoring reliability, and the coding of occupation data (see Table 9.15). Between them, the participating countries used 349 scorers from a variety of backgrounds. The scorers typically included national center staff, teachers, and university students. Six NRCs said they experienced difficulties with implementing the reliability (independent double scoring) procedure in their countries. However, these difficulties generally occurred early in the process and were resolved after the scorers experienced additional scoring practice.



The majority of countries used the coding module available from the *Windows® Data-Entry Manager Software* (WinDEM, IEA DPC, 2008b) provided by the IEA DPC to enter and code occupation data. Alternative methods included coding occupations directly in the instruments (for subsequent scanning or data entry using WinDEM) or on an Excel spreadsheet.

#### *Entering and submitting data*

When asked about the persons responsible for entering questionnaire data, the NRCs said that they most frequently used national center staff, followed by a combination of staff from the national center and an external data-entry company (see Table 9.16). In some cases, university students or other external assistants did the data-entry work. Most countries used the WinDEM software to enter data manually. The few that did not typically used scanning procedures or relied on an external company to carry out the data entry according to a format that could later be imported into WinDEM.

*Table 9.15: Numbers of NRC responses to the survey activities questionnaire: scoring open-ended response items and coding occupation data*

Question	Yes	No	Not Answered
Who primarily scored your open-ended response items? (Select one)			0
• National center staff	7		
• Teachers/professional educators	7		
• University students	6		
• Combination of the above	14		
• Other	4		
Did you have any difficulties with the procedures for reliability scoring?	6	32	0
Did you use the coding module available from the WinDEM software for			
• entering occupations?	29	9	0
• coding occupations?	23	15	0

*Table 9.16: Numbers of NRC responses to the survey activities questionnaire: entering and submitting data*

Question	Yes	No	Not Answered
Who entered the data from the questionnaires into computer files? (Select one)			0
• Own staff	12		
• External data entry company	8		
• Combination of the above	10		
• Other	8		
Did you use the WinDEM software to manually enter your data?	32	6	0



### *Other experience*

The last section of the survey activities questionnaire provided an opportunity for NRCs to give feedback on the quality of the ICCS sampling, operational, school coordinator, test administrator, and scoring manuals. The NRCs were generally very positive about the manuals, and nearly all of them described the manuals as “very” or “somewhat” helpful for carrying out the survey; their responses are summarized in Table 9.17. When NRCs were asked for suggestions for improvement, some asked that the information in the manuals be made more concise so that the relevant instructions could be more easily located.

*Table 9.17: Numbers of NRC responses to the survey activities questionnaire: other experience*

Question	Very Helpful	Somewhat Helpful	A Little Helpful	Not Helpful	Not Answered
Did you find the manuals helpful for carrying out the ICCS 2009 main survey?					
• <i>Sampling Manual</i>	30	7	1	0	0
• <i>Survey Operations Procedures, Unit 1</i>	32	5	1	0	0
• <i>Survey Operations Procedures, Unit 2</i>	31	6	1	0	0
• <i>Survey Operations Procedures, Unit 3</i>	27	9	1	0	1
• <i>School Coordinator Manual</i>	24	12	1	1	0
• <i>Test Administrator Manual</i>	28	9	1	0	0
• <i>Scoring Guides</i>	29	7	1	0	1

### **Summary**

The ICCS quality assurance programs conducted at both the international and national levels provided crucial documentation of participating countries’ adherence to the standardized data-collection procedures. The classroom observation record formed an important part of this documentation. The registered observations of the IQCMs indicated that, in nearly all cases, the prescribed survey administration procedures, including the preparatory activities, delivery of the administration script, timing of the testing session, and handling of the instruments were very closely followed.

The survey activities questionnaire documented a number of major aspects related to the ICCS survey administration and provided information useful for improving the quality of future IEA surveys. The results of this questionnaire show that NRCs generally felt well supported by the study consortium and operational manuals, and that they were able to comply well with the international procedures related to sampling, external verification, and reliability scoring.

These monitoring results provide evidence of the high quality of the data-collection sessions. The observed good behavior of the participating students along with the strong evaluations from national and international QCMs, school coordinators, and test administrators, as well as the positive attitudes of other school staff members toward the survey administration also attest to the success of the planning and implementation of ICCS.



## References

- ICCS International Study Center. (2008a). *IEA International Civic and Citizenship Education Study (ICCS) 2009 international quality control monitor manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008b). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 1: School contact and material preparation*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008c). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 2: Within-school sampling and test administration*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008d). *IEA International Civic and Citizenship Education Study (ICCS) 2009 survey operations procedures, Unit 3: Online data collection, scoring, and data entry*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008e). *IEA International Civic and Citizenship Education Study (ICCS) 2009 school coordinator manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008f). *IEA International Civic and Citizenship Education Study (ICCS) 2009 test administrator manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008g). *IEA International Civic and Citizenship Education Study (ICCS) 2009 test administration instructions for the Asian regional module*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008h). *IEA International Civic and Citizenship Education Study (ICCS) 2009 test administration instructions for the European module instrument*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008i). *IEA International Civic and Citizenship Education Study (ICCS) 2009 test administration instructions for the Latin American module instrument*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- ICCS International Study Center. (2008j). *IEA International Civic and Citizenship Education Study (ICCS) 2009 national quality control monitor manual*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- IEA Data Processing and Research Center. (2008a). *Windows® within-school sampling software (WinW3S)* [computer software]. Hamburg, Germany: Author.
- IEA Data Processing and Research Center (IEA DPC). (2008b). *Windows® data-entry manager software (WinDEM)* [computer software]. Hamburg, Germany: Author.



# Data management and creation of the ICCS international database

*Michael Jung and Falk Brese*

## Introduction

This chapter describes the procedures for checking ICCS data and database creation that were implemented by IEA's Data Processing and Research Center (IEA DPC), the ICCS International Study Center (ISC) at the Australian Council for Educational Research (ACER), and the national centers of the participating countries. The main purposes of these procedures were to ensure the following:

- All information in the database conformed to the internationally defined data structure;
- The content of all codebooks and documentation appropriately reflected national adaptations to questionnaires; and
- All variables used for international comparisons were comparable across countries.

All institutions involved in this process applied control measures throughout it in order to assure the quality and accuracy of the ICCS data.

## Confirming the integrity of the ICCS international database

This process required close cooperation between the international and national institutions involved in ICCS. Quality assurance comprised several steps. During the first step, staff at the IEA DPC involved in this process checked the data files provided by each country. They applied standard cleaning rules to verify the accuracy and consistency of the data, and documented any deviations from the international file structure. The IEA DPC sent any queries to national centers, the staff of which modified their data files where necessary. After all modifications had been applied, staff at the IEA DPC re-checked all datasets. This process of editing the data, checking the reports, and implementing corrections was repeated as many times as necessary to ensure that all data were consistent within and comparable across countries.

After the national files had been checked, the IEA DPC provided national centers with national univariate and reliability statistics as well as data almanacs containing international univariate statistics and national item statistics. This material enabled national center staff to compare their national data against the international results.

This step was one of the most important data-quality procedures used because it ensured the international comparability of the data. For example, a particular statistic that might have seemed plausible within a national context could have appeared as an outlier when the national results were compared against the international results. Staff at the IEA DPC reviewed all such instances and, when necessary, addressed it either by recoding the corresponding variables or removing them from the international database.

Once the national databases had been verified and formatted according to the international file format, national center staff sent their data to the ISC, which then produced and subsequently reviewed the basic item statistics. At the same time, the IEA DPC produced data files containing information on the participation of schools and students in each country's sample. Staff at the IEA DPC then used this information, together with data provided by the NRC survey tracking forms<sup>1</sup> and the software designed to standardize operations and tasks, to calculate sampling weights, population coverage, and school, teacher, and student participation rates.<sup>2</sup>



<sup>1</sup> Survey tracking forms were used to record the sampling of schools, classes, teachers, and students (also see Chapter 8).

<sup>2</sup> Chapter 7 of this report provides details about the ICCS 2009 weighting procedures.



After the item-statistics review had been completed and the IEA DPC had finalized the computation of sampling weights, the ICCS ISC calculated the civic knowledge scores as well as questionnaire indices for each participating student. (The scaling methods and procedures are described in Chapters 11 and 12 of this report.) On completing their verification of the sampling weights and scale scores, the ISC sent this material to the IEA DPC for inclusion in the international database and for distribution to the national centers.

### Data checks at the IEA Data Processing and Research Center

As described in Chapter 8 of this report, national center staff members in each participating country were responsible for entering their national ICCS data into the appropriate data files and submitting these files to the IEA DPC. Staff at the IEA DPC then subjected these files to a comprehensive process of checking and editing. To facilitate the data cleaning process, the IEA DPC asked the national centers to provide them with detailed documentation of their data together with their national data files. The data documentation included copies of all original survey tracking forms and the national versions of test booklets and questionnaires, as well as information from the survey activities questionnaire (see details in Chapter 8). National centers also submitted their final national adaptation forms (NAFs) in order to provide and confirm complete documentation on all national adaptations.

### Data-cleaning quality control

Because ICCS 2009 was a large and highly complex study with very high standards for data quality, maintaining these standards required an extensive set of interrelated data checking and data-cleaning procedures. To ensure that all procedures were conducted in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the data quality control included the following steps:

- *Thorough testing of all data-cleaning programs:* before applying the programs to real datasets, the IEA DPC applied them to simulation datasets containing all possible problems and inconsistencies.
- *Registering all incoming data and documents in a specific database:* the IEA DPC recorded the date of arrival as well as specific issues requiring attention.
- *Carrying out data-cleaning according to strict rules:* deviations from the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal.
- *Documenting all systematic data recodings that applied to all countries:* the IEA DPC recorded these in the ICCS general cleaning documentation for the main survey (Brese, Jung, & Schulz, 2010).
- *Logging, in a recoding file, every “manual” correction to a country’s data files:* logging these changes, which occurred only occasionally, allowed IEA DPC staff to undo changes or to redo the whole manual cleaning process at any later stage of the data-cleaning process.
- *Repeating, on completion of data-cleaning for a country, all cleaning steps from the beginning:* this step allowed the IEA DPC to detect any problems that might have been inadvertently introduced during the data-cleaning process.
- *Working closely with national centers and at different steps of the cleaning process:* the IEA DPC provided national centers with the processed data files and accompanying documentation and statistics so that center staff could thoroughly review and correct any identified inconsistencies.





The IEA DPC compared national adaptations recorded in the documentation for the national datasets against the structure of the submitted national data files. IEA DPC staff then recorded any identified deviations from the international data structure in the national adaptation database and in the *ICCS User Guide* (Brese, Mirazchiyski, Schulz, & Zuehlke, 2011). Whenever possible, the IEA DPC recoded national deviations to ensure consistency with the international data structure. However, if international comparability could not be guaranteed, the IEA DPC removed the corresponding data from the international database.

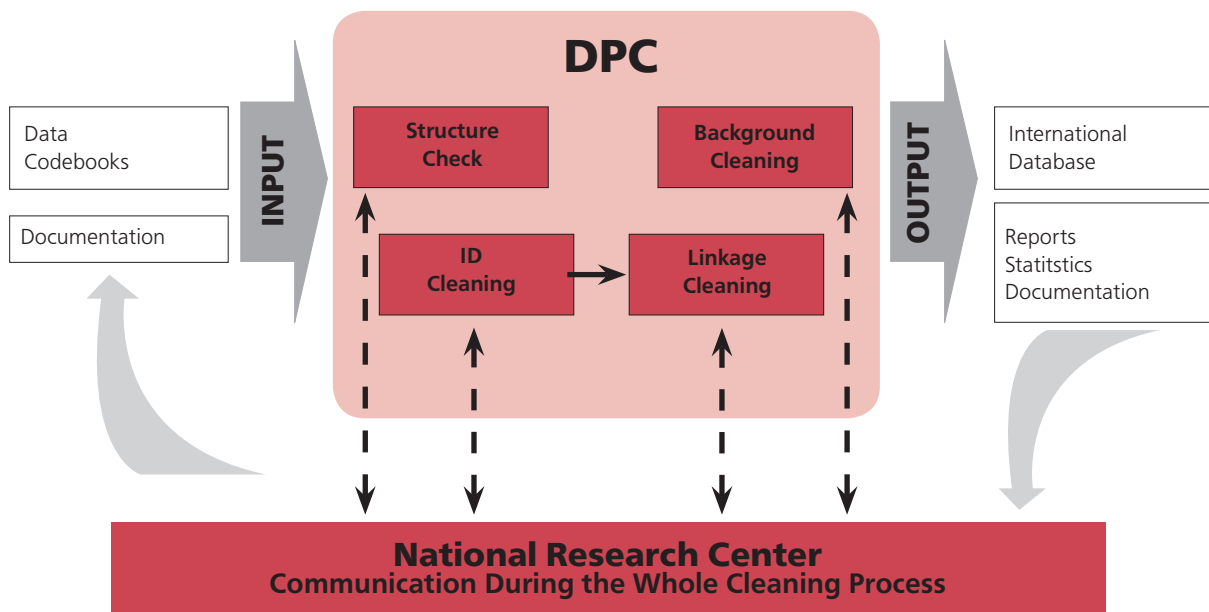
## Preparing national data files

The main objective of the data-cleaning process was to ensure that the data adhered to international formats, that school, teacher, and student information could be linked across different survey files, and that the data reflected the information collected within each country in an accurate and consistent manner.

The program-based data cleaning consisted of the following steps (shown in Figure 10.1 and explained in the following subsections):

- Documentation and structure check;
- Identification variable (ID) cleaning;
- Linkage cleaning;
- Resolving inconsistencies in questionnaire data.

Figure 10.1: Overview of data processing at the IEA DPC



### *Documentation and structure check*

For each country, data cleaning began with an exploratory review of its data-file structures and its data documentation (i.e., national adaptation forms, student tracking forms, and teacher tracking forms). National centers sent their national datasets and all required documentation to the IEA DPC. The fact that most centers sent this documentation greatly facilitated the data-checking process.

The IEA DPC began data cleaning by combining the tracking information and sampling information captured in the WinW3S database with the WinDEM data files containing the corresponding survey instrument data. During this step, IEA DPC staff also merged the data from the principal and teacher questionnaires that some countries had completed online (see Chapter 8 for more information).

The first checks implemented at the IEA DPC identified differences between the international file structure and the national file structures. Some countries made adaptations (such as adding national variables or omitting or modifying international variables) to their questionnaires. The extent and nature of such changes differed across countries: some countries administered the questionnaires without any modifications (apart from translations and necessary adaptations relating to culture or involving language-specific terms), whereas other countries inserted items within existing international variables or added national variables. To keep track of adaptations, the IEA DPC asked the national centers to complete national adaptation forms while they were adapting the international codebooks. Where necessary, the IEA DPC modified the structure of the national data files to ensure that the resulting data remained comparable across countries.

As part of this standardization process, the IEA DPC also rearranged the file structure from a booklet-oriented model designed to facilitate data entry to an item-oriented layout more suited to data analysis. This rearrangement was feasible because a direct correspondence between the data-collection instruments and the data files was no longer necessary. The IEA DPC discarded, at this time, variables created purely for verification purposes during data entry, and made provision for adding new variables necessary for analysis and reporting (these included reporting variables, derived variables, sampling weights, and scale scores).

Once IEA DPC staff had ensured that each data file matched the international format, as specified in the international codebooks, they defined a series of standard data-cleaning rules for further processing of the national data files. Processing at this stage employed software developed by IEA DPC staff. This software contained a facility able to identify and correct inconsistencies in the data. Each problem found at this stage was identified by a unique problem number, described, and recorded in a database. The action taken by the cleaning program or by IEA DPC staff with respect to each problem was also recorded.

The IEA DPC reported problems that could not be rectified automatically to the responsible NRC. National center staff then checked the original data-collection instruments and tracking forms to trace the source of these errors. Wherever possible, staff at the IEA DPC suggested a remedy and asked the national centers to either accept it or propose an alternative. If a national center could not solve problems through verification of the instruments or forms, the IEA DPC applied a general cleaning rule to the files to rectify this error. After all of the automatic updates had been applied, IEA DPC staff used SAS recoding scripts to directly apply any remaining corrections to the data files.



### *Identification variable (ID) cleaning*

Each record in a data file needs to have a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. If two records in an ICCS database shared the same ID number and contained exactly the same data, the IEA DPC deleted one of the records and kept the other one in the database. If both records contained different data and IEA DPC staff found it impossible to identify which record contained the “true data,” they removed both records from the database. The IEA DPC tried to keep such losses to a minimum; actual deletions were rare.

Although the ID cleaning covered all data from all instruments, it focused mainly on the student questionnaire file, which contained most of the critical ID variables. In addition to checking the unique student ID number, IEA DPC staff also needed to check variables pertaining to student participation and exclusion status, as well as students’ dates of birth and dates of testing in order to calculate student age at the time of testing. The student tracking forms provided an important tool in relation to resolving anomalies in the database. The IEA DPC conducted all cleaning procedures in close cooperation with the national centers. After national center staff had cleaned the identification variables, they passed the clean databases with information about student participation and exclusion on to the IEA DPC sampling unit, which used this information to calculate students’ participation rates, exclusion rates, and student sampling weights (see Chapter 7 for details).

### *Linkage check*

Because, in ICCS, data about students, their schools, and teachers appeared in a number of different files, it was crucial to correctly link these records so as to provide meaningful data for analysis and reporting. Linkage was implemented through a hierarchical ID numbering system that included a school, class, and student component,<sup>3</sup> cross-checked against the survey tracking forms. Student ID values in the achievement file and student questionnaire file had to be matched correctly, as did those in the reliability scoring file and the achievement file. In addition, it was important to ensure that teacher and student records linked to their corresponding schools.

### *Resolving inconsistencies in questionnaire data*

The amount of inconsistent and implausible responses in questionnaire data files varied considerably among countries. However, none of the national datasets was completely free of inconsistent responses. The IEA DPC determined the treatment of inconsistent responses on a question-by-question basis, using all available documentation to make an informed decision. IEA DPC staff also checked all questionnaire data for consistency across the responses given. For example, Question 20 in the school questionnaire asked for the total school enrolment (number of students) in all grades, while Question 21 asked for the enrolment in the target grade only. Clearly, the number given as a response to Question 21 could not possibly exceed the number provided by school principals in Question 20. The IEA DPC flagged inconsistencies of this kind and then asked the national centers to review these issues. IEA DPC staff recoded as “invalid” those cases that could not be corrected or where the data provided made no sense.

---

<sup>3</sup> The ID number of a higher level is included in the ID number of a lower sampling level. The class ID includes the school ID, and the student ID includes the class ID (e.g., student 10120523 may be described as student 23 of class 05 in school 1012).



Filter questions, which appeared in some questionnaires, directed respondents to a particular subquestion or further section of the questionnaire. The IEA DPC applied the following cleaning rules to these filter questions and the corresponding questions that followed:

- If the answer to the filter question was “no” or “not applicable,” the IEA DPC recoded any responses to the dependent questions as “logically not applicable”;
- If the response to the filter question was omitted, the IEA DPC either recoded the answers to the dependent questions or omitted them.

The IEA DPC applied a modified rule to a section in the teacher questionnaire that was designed to gather information about the teaching of subjects related to civic and citizenship education. Only teachers who responded that they were teaching a subject in this learning area were supposed to complete this section. The IEA DPC removed data in this part of the questionnaire from the database if the teacher concerned had reported that he or she did not teach a subject related to civic and citizenship education.

The IEA DPC also applied what are known as *split variable checks* to questions where the answer was coded into several variables. For example, Question 11b in the student questionnaire asked students to provide information about all people living at home with them most or all of the time. Student responses were captured in a set of nine variables, each one coded as “Yes” if the corresponding option was checked and “No” if the option was left unchecked. Occasionally, students checked the “Yes” boxes but left the “No” boxes unchecked. Because, in these cases, it was clear that the unchecked boxes actually meant “No,” the IEA DPC recoded these responses accordingly, provided that the students had given affirmative responses in the other categories.

## National cleaning documentation

The IEA DPC sent the NRCs a detailed report of all problems that were identified in their data and the steps taken to correct them. IEA DPC staff also recorded and sent a list of all deviations from the international data collection instruments and the international file structure.

The IEA DPC furthermore provided each national center with revised data files. These included all agreed-upon edits, updates, and structural modifications, as well as a list of a range of new variables that could be used for analytic purposes. For example, the student files included nationally standardized scores for civic knowledge, which meant that the national centers could conduct national analyses before the international database became available.

### Handling of missing data

Two types of entries were possible during the ICCS data capture: valid data values, and missing data values. Missing data can be assigned a value of omitted, invalid, or not administered during data entry. The IEA DPC applied additional missing codes to the data to facilitate further analyses. This process led to five different types of missing data being distinguished in the international database:

- *Omitted*: the respondent had a chance to answer the question but did not do so; the corresponding question or item was thus left blank.
- *Not administered*: the respondent was not administered the actual item or question and therefore could not read and answer the question.
- *Invalid*: this code was used in both the questionnaire and the achievement files for responses that were not interpretable (e.g., when respondents ticked more than one box in a multiple-choice question).
- *Logically not applicable*: the respondent answered a preceding filter question in a way that made the following dependent questions not applicable to him or her.



- *Not reached (used only in the achievement files)*: this code indicated those items that students did not reach because of a lack of time.<sup>4</sup>

### *Data products*

Data products sent to national centers by the IEA DPC and the ISC included both item statistics and data files.

### **Item statistics**

ISC staff produced and then sent a set of item statistics to each national center for review. Each set contained weighted summary statistics for the participating country on each variable included in the survey instruments. The ISC also used these datasets during their data reviews. In addition, IEA DPC staff produced a set of preliminary scoring reliability statistics for each national dataset. This contained summary statistics at the item level on the percent of agreement between scorers.

### **Versions of the national data files**

Building the international database was an iterative process. On completion of each major data-processing step, the IEA DPC sent a new version of data files to the national centers so that staff could review their data and run their own separate checks to validate the new data-file versions. This process meant that national centers received several versions of their data, and their data only, before the international database was published. IEA DPC staff sent the first of these versions as soon as they considered the data “clean” with respect to identification codes and linkage issues.

These first sets of files contained nationally standardized achievement scores that the IEA DPC calculated by means of Rasch scaling. They were also accompanied by documentation that included a list of the cleaning checks and corrections made to the data so that the national centers could review the cleaning process. The IEA DPC sent national centers another version of the data files together with data almanacs once the weights and international achievement scores were available. This step did not take place until all tables and figures contained in the ICCS international reports had been verified and final updates to the data files implemented. This approach enabled national center staff to replicate the results presented in the international reports.

## **The ICCS international database**

The ICCS international database incorporated all national data files from participating countries. The data processing that the IEA DPC conducted ensured that:

- Information coded in each variable was internationally comparable;
- National adaptations were reflected appropriately in all variables;
- Questions that were not internationally comparable had been removed from the database;
- All entries in the database could be linked to the appropriate respondent—student, teacher, or principal;
- Only those records adjudicated as participating remained in the international database files; and
- Sampling weights and student achievement scores were available for international comparisons.

More information about the ICCS international database is provided in the *ICCS User Guide for the International Database* (Brese et al., 2011).

<sup>4</sup> “Not reached” codes were derived as follows: an item received this coding if the student concerned did not respond to any of the items following it (i.e., did not continue on to the end of the test) and/or if he or she did not respond to the item preceding it.



## Summary

To achieve a high standard of data quality, ICCS implemented a series of data-management procedures that included checks to ensure the consistency of national database structures, provide proper documentation of all national adaptations, and safeguard the comparability of international variables across national datasets. Staff at the IEA DPC reviewed all national databases in cooperation with national centers after completing a series of thorough checking procedures and before creating the final ICCS database. Final data products included item statistics, preliminary versions of the national data files, the international database accompanied by a user guide and supplementary information.

## References

Brese, F., Jung, M., Mirazchiyski, P., Schulz, W., & Zuehlke, O. (2011). *ICCS 2009 user guide for the international database*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Brese, F., Jung, M., & Schulz, W. (2010). *ICCS main survey general cleaning documentation*. Hamburg, Germany: International Association for the Evaluation of Educational Achievement (IEA) Data Processing and Research Center.



# Scaling procedures for ICCS test items

Wolfram Schulz and Julian Fraillon

## Introduction

This chapter describes the procedures used to analyze and scale the ICCS international and regional test items that were administered to measure students' civic knowledge. The chapter covers these topics:

- The scaling model used to analyze and scale the test items;
- Test coverage and item dimensionality;
- Assessment of item fit;
- Assessment of scorer reliabilities for open-ended items;
- Differential item functioning by gender;
- Review of cross-national measurement equivalence;
- International item adjudication;
- International item calibration and test reliability;
- International ability estimates (plausible values and weighted likelihood estimates);
- Estimation of changes in civic content knowledge between 1999 and 2009; and
- Regional test items for the European and Latin American modules.

The development of the ICCS test items was described in Chapter 2 and was guided by the ICCS assessment framework (see Schulz, Fraillon, Ainley, Losito, & Kerr, 2008).

## The scaling model

Item response theory (IRT) scaling methodology was used to scale the test items.

Use of the one-parameter (Rasch) model (Rasch, 1960) for dichotomous items means that the probability of selecting Category 1 instead of 0 is modeled as

$$P_i(\theta) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)},$$

where  $P_i(\theta)$  is the probability for person  $n$  to score 1 on item  $i$ ,  $\theta_n$  is the estimated ability of person  $n$ , and  $\delta_i$  is the estimated location of item  $i$  on this dimension. For each item, item responses are modeled as a function of the latent trait  $\theta_n$ .

In the case of items with more than two ( $k$ ) categories (as, for example, with Likert-type items), this model can be generalized to the partial credit model (Masters & Wright, 1997), which takes the form of

$$P_{x_i}(\theta) = \frac{\exp\left(\sum_{k=0}^x (\theta_n - \delta_i + \tau_{ij})\right)}{\sum_{b=0}^{m_i} \exp\left(\sum_{k=0}^b (\theta_n - \delta_i + \tau_{ij})\right)} \quad x_i = 0, 1, \dots, m_i.$$

Here,  $P_{x_i}(\theta)$  denotes the probability of person  $n$  scoring  $x$  on item  $i$ , and  $\theta_n$  denotes the person's ability. The item parameter  $\delta_i$  gives the location of the item on the latent continuum;  $\tau_{ij}$  denotes an additional step parameter.

ACER Conquest, Version 2.0 software (Wu, Adams, Wilson, & Haldane, 2007) was used to scale the ICCS test data.





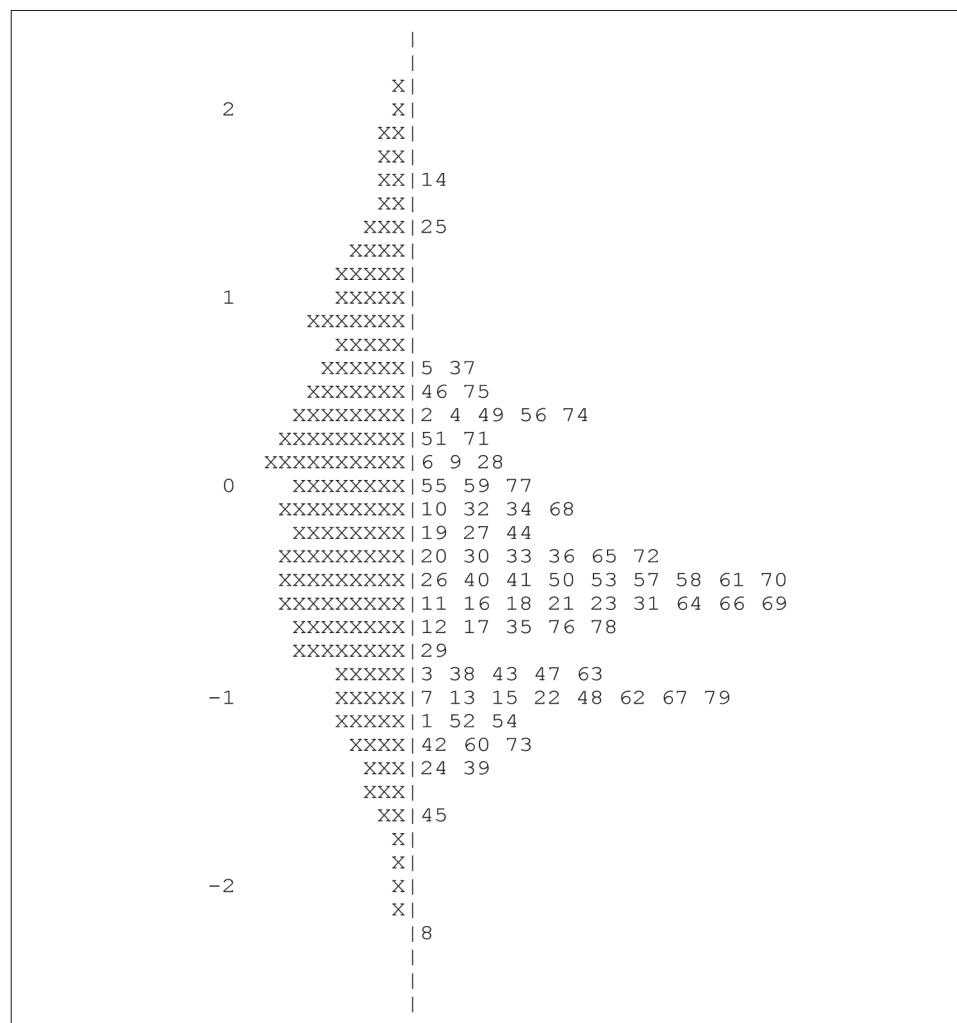
## Test coverage and item dimensionality

When measuring cognitive abilities, it is important to use test items that cover the different levels of achievement found in the target population. Figure 11.1 shows the distribution of cognitive abilities among ICCS students (for the representative sample used for the final calibration) and the location of items (at  $rp = 0.5$ ).

The range of item difficulties generally matched the abilities found in the student population, but the average item difficulties were somewhat lower than the average student abilities. Overall, the test items were better at targeting students in the lower than in the higher civic knowledge ranges. However, the nature of this targeting varied across countries according to the distribution of student achievement within each country.

Multidimensional item response models were used to assess the dimensionality of items. Two of the possible item dimensions that were explored are based on the structure of the cognitive domains described in the ICCS assessment framework (Schulz et al., 2008).

Figure 11.1: Mapping of student abilities and item difficulties



Note: Calibration results for the international calibration sample (adjudicated items only).



Dimensionality was explored in terms of the ICCS assessment framework content dimensions (*civic society and systems* versus *others*) and the ICCS assessment framework cognitive dimensions (*knowing* versus *reasoning and analyzing*). Multidimensional IRT models using ACER ConQuest typically showed latent correlations over 0.90, thus indicating high similarity between the item subgroups. Given these results, a decision was made not to include reports relating to the civic knowledge subscales in the ICCS 2009 international reports.

### Assessment of item fit

Goodness of fit for individual items can be determined by calculating a mean square statistic (Wright & Masters, 1982). Reviewing this residual-based item fit gives an indication of the extent to which each item fits the item response model. However, there are no clear rules for acceptable item fit, and some statisticians recommend that analysts and researchers interpret residual-based statistics with caution (see, for example, Rost & von Davier, 1994). It was consequently decided to use a range of item statistics to assess item fit.

ACER Conquest, Version 2.0 software (Wu et al., 2007) was used to estimate the item parameters and to conduct the analysis of item fit.

Table 11.1 shows the (corrected) item-total correlations of correct responses (or partial credit items) and the weighted item fit statistics. Only two items (CI2MOM1 and CI2HRM2) had item-total correlations below 0.2 (indicating low discrimination), and only CI2HRM2 showed relatively poor residual-based item fit (weighted mean square fit statistic of 1.23).

Conquest was also used to generate item characteristic curves (ICCs). These provide a graphical representation of item fit across the range of student abilities for each item, including dichotomous and partial credit items.

Figure 11.2 shows the ICC for item CI2MOM1. Although the discrimination is not entirely satisfactory and although Category 2 attracted responses from students with higher ability, it still shows that the students with higher levels of knowledge were those most likely to give the correct response (Option 4) and that the students with lower levels were those more likely to choose the incorrect option (2). This outcome led to the decision to retain this item for scaling.

Figure 11.3 shows the ICC for CI2HRM2, another item with relatively poor item discrimination. In this case, the curves for the correct response (2) and the curves for the distractor that attracted the most student responses (Option 4) run almost parallel. It was accordingly decided to exclude this item from the final scaling.

The functioning of the partial credit scoring guides was further analyzed through a review of the proportion of responses in each response category and the correct ordering of mean abilities of students across response categories. This analysis confirmed that the scaling properties of all six partial credit items could be satisfactorily included in the scaling of student test data.



Table 11.1: Item total-score correlations and weighted item fit for international calibration sample

Item	Item-Score Correlation	Weighted Item Fit	Item	Item-Score Correlation	Weighted Item Fit
CI2COM1	0.39	0.94	CI2PFM2	0.39	0.98
CI2MOM1	0.19	1.12	CI2PCM1	0.32	1.03
CI2MLM1	0.40	0.95	CI2PCM2	0.38	0.96
CI2MLM2	0.28	1.05	CI2VOM1	0.37	0.97
CI2PDO1	0.43	1.02	CI2VOM2	0.33	1.02
CI2RDM2	0.38	0.98	CI2VOM3	0.40	0.92
CI2SHM1	0.44	0.91	CI2DLM1	0.36	0.98
CI2SHM2	0.37	0.91	CI2HRM1	0.39	0.96
CI2TGM1	0.33	1.02	CI2JOM1	0.31	1.02
CI2TGM2	0.27	1.07	CI2WFO2	0.37	1.20
CI2BPM1	0.33	1.02	CI2PGM1	0.37	0.98
CI2BPM2	0.36	0.99	CI2PGM2	0.34	1.01
CI2GFM1	0.31	1.02	CI2ECM1	0.42	0.93
CI2BIO1	0.36	1.04	CI2ECM2	0.27	1.07
CI2GLM1	0.35	0.99	CI2CEM1	0.37	0.97
CI2GLM2	0.33	1.01	CI2CEM2	0.20	1.13
CI2FDM1	0.39	0.97	CI2WFO1	0.40	1.05
CI2FSM1	0.41	0.96	CI2ORM1	0.43	0.94
CI2SCM1	0.41	0.95	CI2RCM1	0.36	0.99
CI2SCM2	0.28	1.07	CI2PJM1	0.20	1.13
CI2ASM1	0.35	1.00	CI2PJM2	0.41	0.93
CI2ASM2	0.40	0.95	CI2REM2	0.29	1.05
CI2CNM1	0.38	0.98	CI2REM3	0.39	0.96
CI2CNM2	0.46	0.89	CI101M1	0.40	0.96
CI2ETO1	0.36	1.04	CI109M1	0.38	0.98
CI2ETM2	0.26	1.08	CI108M1	0.26	1.08
CI2BCM1	0.39	0.98	CI128M1	0.40	0.96
CI2PRM1	0.33	1.02	CI137M1	0.36	0.98
CI2CCM1	0.36	0.99	CI110M1	0.42	0.95
CI2CCM2	0.34	1.01	CI113M1	0.41	0.96
CI2SRM1	0.36	0.99	CI104M1	0.37	0.99
CI2SRM2	0.40	0.97	CI115M1	0.37	0.99
CI2SRM3	0.30	1.05	CI119M1	0.44	0.93
CI2OMM1	0.29	1.06	CI120M1	0.30	1.02
CI2OMM2	0.49	0.89	CI121M1	0.25	1.09
CI2OMM3	0.35	1.01	CI127M1	0.23	1.10
CI2HRM2	0.09	1.23	CI132M1	0.32	1.03
CI2RRO1	0.43	1.03	CI129M1	0.33	1.02
CI2DCM1	0.44	0.92	CI130M1	0.48	0.89
CI2PFM1	0.31	1.01	CI106M1	0.46	0.90

**Note:** International calibration sample. Item-total correlations below 0.2 and item fit values below 0.8 and above 1.2 are shaded in grey.

Figure 11.2: Item characteristic curve by category for Item CI2MOM1

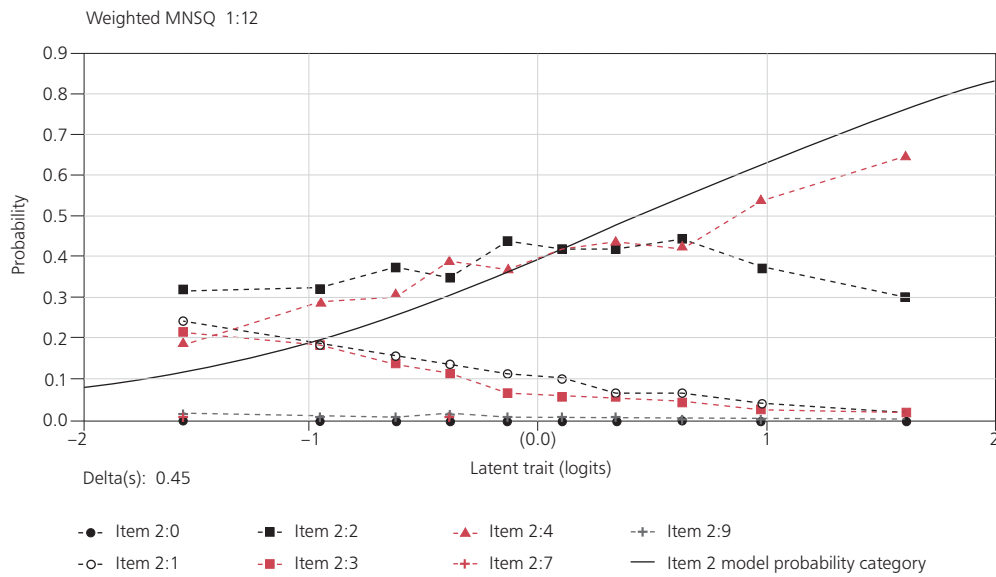
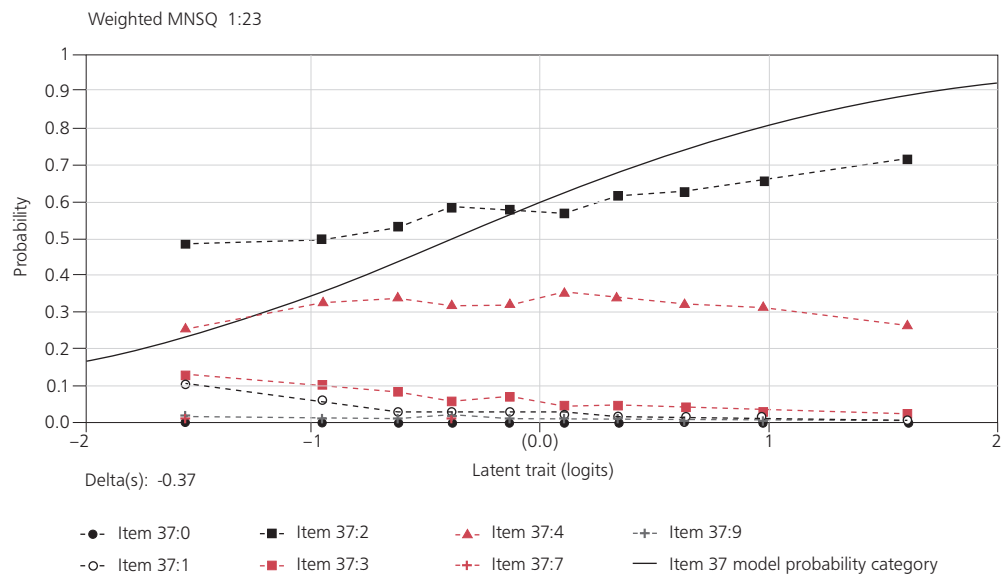


Figure 11.3: Item characteristic curve by category for Item CI2HRM2



## Assessment of scorer reliabilities

The open-ended items in the ICCS cognitive test were scored according to the scoring guides that were refined as an outcome of experiences in the international field trial of test items. Within countries, for each of the seven booklets, subsamples of about 100 student records were scored twice by different scorers. This double-scoring procedure provided an assessment of scorer reliabilities. Table 11.2 shows the percentages of scorer agreement, which ranged from 49 to 100 percent. On average, scorer agreement for the six items was between 85 and 89 percent.

As has been the practice in other IEA studies, data from items scored with a minimum of 70 percent scorer agreement were retained for scaling and inclusion in the international database. This adjudication was made for each open-response item scored in each country.

## Differential item functioning by gender

Further exploration of the quality of the items was conducted through an assessment of differential item functioning (DIF) by gender. DIF occurs when groups of students with the same degree of ability have different probabilities of responding correctly to an item. For example, if boys have a higher probability than girls with the same degree of ability of correctly answering an item, the item shows gender DIF. This situation is a violation of the model, which assumes that the probability is a function of ability only and not of any group membership.

Estimates of gender DIF were derived by including interaction terms in the item response model. Gender DIF for dichotomous items could then be estimated as:

$$P_i(\theta) = \frac{\exp(\theta_n - (\delta_i - \eta_g - \lambda_{ig}))}{1 + \exp(\theta_n - (\delta_i - \eta_g - \lambda_{ig}))}$$

For the purpose of measuring parameter equivalence across the two gender groups  $g$ , an additional parameter for gender effects  $\lambda_{ig}$  is added to the scaling model, where  $\theta_n$  is the estimated ability of person  $n$  and  $\delta_i$  is the estimated location of item  $i$ . However, to obtain proper estimates, there is also a need to include the overall gender effect ( $\eta_g$ ) in the model.<sup>1</sup> Both item-by-gender interaction estimates  $\lambda_{ig}$  and overall gender effects ( $\eta_g$ ) were constrained to have a sum of 0.

Gender DIF estimates for a partial credit model for items with more than two categories (here, constructed items) could then be modeled as:

$$P_{x_i}(\theta) = \frac{\exp \sum_{k=0}^m (\theta_n - (\delta_i - \eta_g - \lambda_{ig} + \tau_{ij}))}{\sum_{h=0}^m \exp \sum_{j=0}^h (\theta_n - (\delta_i - \eta_g - \lambda_{ig} + \tau_{ij}))} \quad x_i = 0, 1, 2, \dots, m_i$$

Here,  $\theta_n$  denotes the person's ability,  $\delta_i$  gives the item location parameter on the latent continuum,  $\tau_{ij}$  is the step parameter,  $\lambda_{ig}$  is the item-by-gender interaction effect, and  $\eta_g$  is the overall gender effect.

Table 11.3 shows the gender DIF estimates for those items retained for scaling. As is apparent in the table, only a few items—five multiple-choice and one open-ended—showed some (limited) form of DIF (estimates larger than 0.3 logits). In general, because the gender DIF for ICCS test items was viewed as not posing a serious problem, it was decided not to exclude from scaling any items on the basis of gender DIF.



<sup>1</sup> The minus sign ensures that higher values of the gender effect parameters indicate higher levels of item endorsement in the gender group with a higher value (here, females).

Table 11.2: Percentages of scorer agreement for open-ended ICCS test items

Country	CI2PDO1	CI2BIO1	CI2ETO1	CI2RRO1	CI2WFO2	CI2WFO1
Austria	100	100	100	100	100	100
Belgium (Flemish)	100	100	100	100	100	100
Bulgaria	70	77	74	74	76	75
Chile	81	86	88	86	84	81
Chinese Taipei	92	92	95	96	97	93
Colombia	81	85	83	80	83	85
Cyprus	100	100	100	100	100	100
Czech Republic	94	98	91	87	84	89
Denmark	81	84	87	87	96	87
Dominican Republic	73	<b>68</b>	<b>67</b>	<b>68</b>	<b>64</b>	<b>61</b>
England	88	93	87	84	85	85
Estonia	<b>65</b>	72	76	71	76	74
Finland	81	91	91	87	84	86
Greece	97	98	97	95	99	97
Guatemala	91	96	97	96	94	92
Hong Kong SAR	70	71	86	<b>68</b>	<b>69</b>	<b>69</b>
Indonesia	84	95	80	87	89	85
Ireland	98	96	92	99	93	97
Italy	83	83	86	81	85	87
Korea, Republic of	93	96	97	91	99	91
Latvia	<b>52</b>	<b>68</b>	<b>60</b>	<b>49</b>	<b>61</b>	<b>64</b>
Liechtenstein	89	90	96	93	90	91
Lithuania	98	97	98	96	99	98
Luxembourg	<b>53</b>	79	76	<b>69</b>	72	70
Malta	70	80	<b>65</b>	<b>68</b>	<b>59</b>	<b>64</b>
Mexico	100	100	100	100	100	100
Netherlands	93	90	86	86	89	90
New Zealand	82	91	94	92	87	92
Norway	83	92	89	84	88	84
Paraguay	95	98	96	93	96	97
Poland	81	76	79	79	76	75
Russian Federation	98	96	97	98	98	99
Slovak Republic	82	93	86	85	91	87
Slovenia	91	96	97	94	94	93
Spain	84	89	91	86	86	93
Sweden	72	80	84	<b>69</b>	82	80
Switzerland	84	92	94	88	87	90
Thailand	99	99	99	98	99	100
Average	85	89	88	86	87	87

**Note:** Based on national subsamples. Percentages below 70 percent in **bold**.



Table 11.3: Gender DIF estimates for retained test items

Item	Gender DIF Estimate	Item	Gender DIF Estimate	Item	Gender DIF Estimate
CI2COM1	0.29	CI2PRM1	-0.07	CI2CEM2	-0.05
CI2MOM1	0.21	CI2CCM1	-0.13	CI2WFO1	0.01
CI2MLM1	0.17	CI2CCM2	-0.02	CI2ORM1	-0.03
CI2MLM2	-0.15	CI2SRM1	0.10	CI2RCM1	-0.05
CI2PDO1	0.17	CI2SRM2	-0.08	CI2PJM1	0.12
CI2RDM2	-0.13	CI2SRM3	0.05	CI2PJM2	0.15
CI2SHM1	-0.12	CI2OMM1	0.07	CI2REM2	0.06
CI2SHM2	0.26	CI2OMM2	0.00	CI2REM3	0.24
CI2TGM1	-0.05	CI2OMM3	0.24	CI101M1	-0.18
CI2TGM2	-0.56	CI2RRO1	0.20	CI109M1	0.05
CI2BPM1	-0.07	CI2DCM1	0.21	CI108M1	-0.20
CI2BPM2	0.20	CI2PFM1	0.05	CI128M1	0.07
CI2GFM1	0.14	CI2PFM2	0.08	CI137M1	0.12
CI2BIO1	0.30	CI2PCM1	-0.02	CI110M1	-0.19
CI2GLM1	-0.11	CI2PCM2	0.19	CI113M1	-0.05
CI2GLM2	0.13	CI2VOM1	-0.02	CI104M1	-0.27
CI2FDM1	0.09	CI2VOM2	0.04	CI115M1	-0.19
CI2FSM1	0.03	CI2VOM3	0.17	CI119M1	-0.13
CI2SCM1	0.00	CI2DLM1	-0.14	CI120M1	0.28
CI2SCM2	-0.14	CI2HRM1	0.04	CI121M1	-0.32
CI2ASM1	-0.32	CI2JOM1	-0.01	CI127M1	-0.48
CI2ASM2	-0.18	CI2WFO2	0.09	CI132M1	0.02
CI2CNM1	-0.03	CI2PGM1	0.12	CI129M1	-0.32
CI2CNM2	0.25	CI2PGM2	0.04	CI130M1	-0.11
CI2ETO1	0.04	CI2ECM1	0.16	CI106M1	0.02
CI2ETM2	-0.24	CI2ECM2	-0.04		
CI2BCM1	-0.01	CI2CEM1	-0.13		

**Note:** International calibration sample: gender DIF estimates below -0.3 and above 0.3 shaded in grey. Negative values show differential item functioning in favour of females; positive values show DIF in favour of males.

## Cross-national measurement equivalence

With any test used to assess student achievement cross-nationally, it is important that the test items function similarly across those countries. Items show *item-by-country interaction* when students from different countries but with the same ability vary in their probability of answering these questions. Test items with considerable item-by-country interaction are not suitable for the scaling of cognitive test items in international surveys.

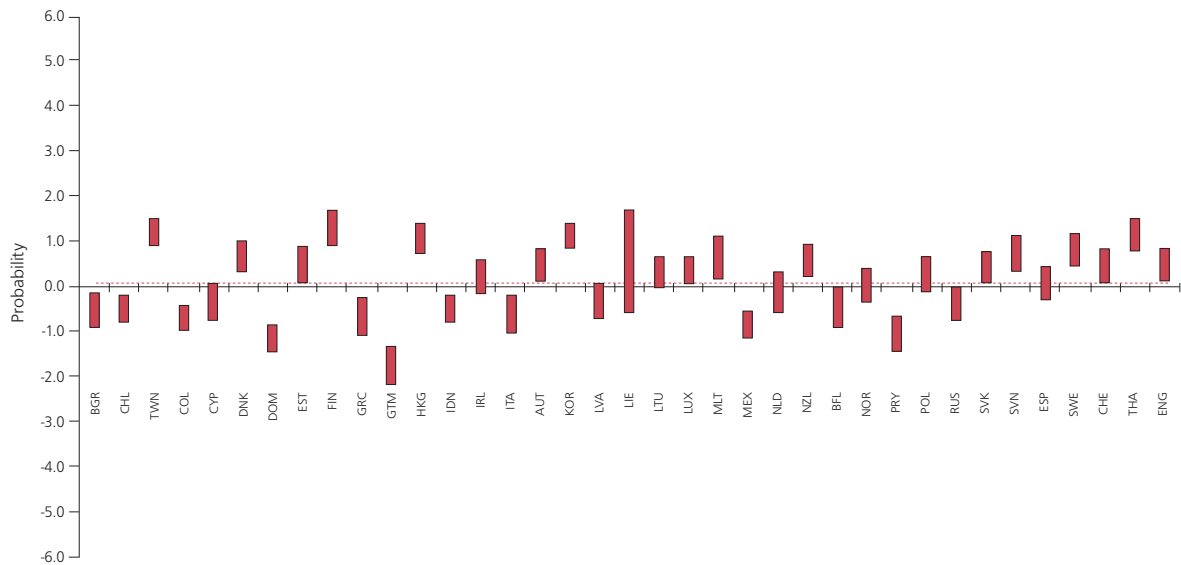
For the main survey analyses of ICCS test items, national calibrations were compared with international item parameters in order to assess the occurrence of item-by-country interaction. Confidence intervals were computed for each national item parameter. Computation was based on the respective standard errors, and the confidence intervals were then adjusted for possible design effects and for multiple comparisons.





Figure 11.4 shows the item-by-country interaction graph for item CI2HRM2, which was not retained for international scaling. The figure shows clear and considerable variation in the item difficulties across countries. Similar graphs produced for each test item were used in the test-item adjudication process at the international and national levels, while information about occurrence of cross-national DIF was used to identify items for post-verification checks after completion of the main data collection.

Figure 11.4: Example of item-by-country interaction graph for Item CI2HRM2



Although the ICCS test items showed generally only limited item-by-country interactions, there were some national item difficulties that deviated quite considerably from the international item difficulty. In these cases, these items were omitted from scaling for those national samples where larger deviations were observed.

Item-by-country interaction was also examined for the open-ended items. With these items, item-by-country interaction can be evidence of differences in the relative harshness of markers across countries. Comparison of the relative difficulties of open-ended items with multiple-choice items across all countries made evident that students in the Dominican Republic and Indonesia appeared to find it easier to answer the open-ended items correctly than did students in the other countries. This situation suggested problems with how the scoring procedures were conducted. All open-ended items for these two countries were subsequently removed from scaling and the international database.

### Missing data issues

There were three possible types of missing responses in the ICCS test. These were omitted items (coded as 9), not-administered items (coded as 8), and invalid responses (coded as 7). The omitted response category was used when a student provided no response at all to an item administered to him or her. Not-administered items were those that, although in the whole item pool, were not in a booklet administered to a student either deliberately (when there were alternative or rotated test booklets) or, in rare cases, in error. Invalid responses occurred when, for example, students ticked more than one of the possible answers to a multiple-choice item.



Table 11.4 shows the percentages of omitted and invalid responses for the international calibration sample. There were considerably more omissions for open-ended items than for multiple-choice items. The percentages of invalid responses were generally low.

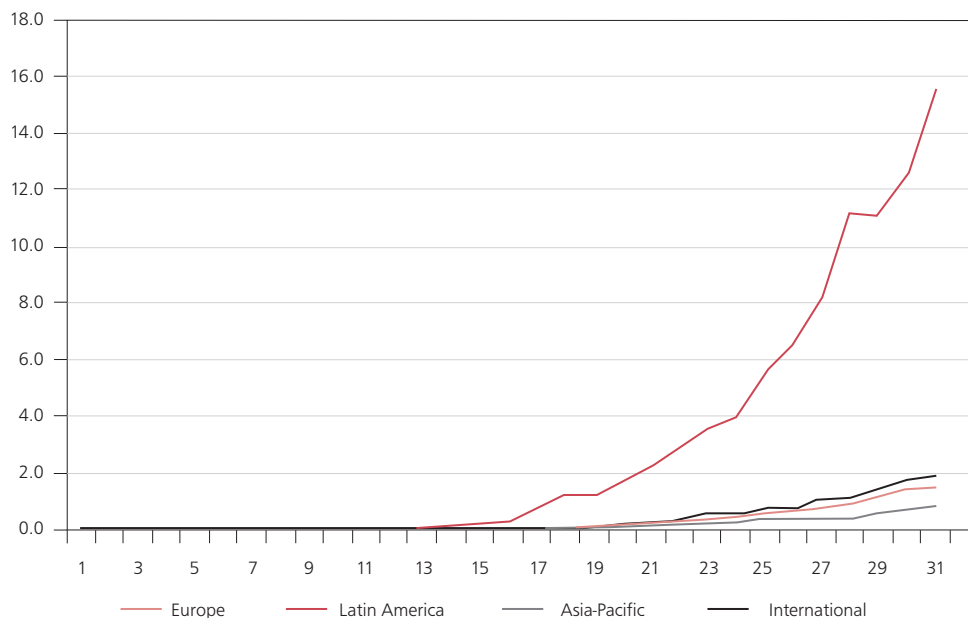
*Table 11.4: Percentages of omitted and invalid responses for test items*

Item	Percentage Omitted	Percentage Invalid	Item	Percentage Omitted	Percentage Invalid
CI2COM1	0.4	0.2	CI2PFM2	0.7	0.3
CI2MOM1	0.8	0.2	CI2PCM1	0.5	0.5
CI2MLM1	0.5	0.2	CI2PCM2	0.5	0.3
CI2MLM2	0.8	0.2	CI2VOM1	1.0	0.2
CI2PDO1	16.2	0.0	CI2VOM2	0.7	0.3
CI2RDM2	1.1	0.2	CI2VOM3	0.6	0.3
CI2SHM1	0.6	0.3	CI2DLM1	1.6	0.2
CI2SHM2	0.6	0.3	CI2HRM1	0.6	0.3
CI2TGM1	1.0	0.3	CI2JOM1	0.6	0.3
CI2TGM2	0.6	0.5	CI2WFO2	13.0	0.0
CI2BPM1	0.5	0.4	CI2PGM1	1.2	0.3
CI2BPM2	0.7	0.2	CI2PGM2	1.6	0.4
CI2GFM1	0.6	0.2	CI2ECM1	0.3	0.3
CI2BIO1	14.1	0.0	CI2ECM2	0.6	0.4
CI2GLM1	0.8	0.5	CI2CEM1	0.6	0.2
CI2GLM2	0.9	0.4	CI2CEM2	0.8	0.4
CI2FDM1	1.1	0.2	CI2WFO1	9.9	0.0
CI2FSM1	0.8	0.5	CI2ORM1	1.1	0.3
CI2SCM1	1.2	0.3	CI2RCM1	1.0	0.2
CI2SCM2	0.8	0.3	CI2PJM1	0.7	0.3
CI2ASM1	0.6	0.5	CI2PJM2	0.8	0.2
CI2ASM2	0.7	0.3	CI2REM2	0.9	0.4
CI2CNM1	0.5	0.7	CI2REM3	0.7	0.9
CI2CNM2	0.4	0.3	CI101M1	0.7	0.2
CI2ETO1	16.5	0.0	CI109M1	0.8	0.4
CI2ETM2	1.0	0.4	CI108M1	1.0	0.4
CI2BCM1	0.6	0.5	CI128M1	1.3	0.3
CI2PRM1	0.5	0.4	CI137M1	0.6	0.5
CI2CCM1	0.7	0.3	CI110M1	1.2	0.2
CI2CCM2	0.8	0.4	CI113M1	0.6	0.2
CI2SRM1	0.8	0.4	CI104M1	0.6	0.3
CI2SRM2	0.6	0.3	CI115M1	1.1	0.3
CI2SRM3	0.7	0.4	CI119M1	1.0	0.3
CI2OMM1	1.0	0.3	CI120M1	0.6	0.6
CI2OMM2	0.6	0.3	CI121M1	1.0	0.3
CI2OMM3	0.7	0.5	CI127M1	1.5	0.2
CI2HRM2	0.9	0.6	CI132M1	0.7	0.5
CI2RRO1	16.0	0.0	CI129M1	1.3	0.2
CI2DCM1	0.6	0.5	CI130M1	1.0	0.2
CI2PFM1	0.5	0.4	CI106M1	0.7	0.3

A separate missing category called “not reached” (coded as 6) was created for analysis purposes. An item was coded as not reached if the student concerned did not respond to any of the items following it (i.e., did not continue on to the end of the test) and/or if he or she did not respond to the item preceding it. The extent of occurrence of Code 6 items provided information about the appropriateness of the test’s length as well as the appropriateness of its difficulty.

Figure 11.5 shows the percentages of not-reached response by item position in Test Booklet 1 for regional groups of countries. As can be seen, the occurrence of not-reached responses was far higher in the Latin American countries than in the other groupings of countries, where nearly all students had no problem with test length. In the Latin American countries, about 15 to 16 percent of students, on average, did not reach the last item in Test Booklet 1. Regional patterns in relation to the other booklets were similar. However, note that there was some variation within the country groups. In Latin America, for example, the national percentages of not reached for the last booklet item ranged from 9 to 24 percent.

Figure 11.5: Percentages of not-reached responses for groups of countries for Test Booklet 1



## International item adjudication

Adjudication of test items was carried out first at the international level for the ICCS calibration sample and then separately for each national subsample.

At the *international level*, item characteristics were assessed for the calibration sample. Here, the review encompassed item-fit statistics, item-score correlations, item characteristic curves, general measurement equivalence across countries (item-by-country interaction), and gender DIF. For open-ended items, account scorer reliabilities and the correct ordering of average ability estimates per category were also taken into account. Only one of the 80 test items (CI2HRM2) had inadequate scaling properties. It was removed from the international scaling of civic knowledge.

At the *national level*, test items were reviewed by comparing national item-fit statistics with international item-fit statistics. Test items for individual countries that showed large item-by-country interactions were flagged, and open-ended national items for which scorer agreement



fell below 70 percent were removed. All open-ended items for two countries were omitted because it was evident that the students in them found it easier than their international counterparts to answer these items correctly.

National centers were provided with item statistics (see example in Table 11.5) and requested to review flagged test items. These items included cases of unusual item-total correlation (e.g., negative correlations between correct response and overall score) and those showing large differences between national and international item difficulties. They also included open-ended items where the category-total correlations were disordered. In some cases, national centers informed the international study center (ISC) of translation problems that had not been detected during verification. In these cases, the items were categorized as “not administered” in the international database and were excluded from scaling of the corresponding national data.

Working independently from those conducting the national item reviews, members of the ISC flagged national items that showed irregular scaling properties (item misfit or large item-by-country interactions) and conducted post-verifications of item translation. In a number of cases, they identified additional national items that needed to be set to “not administered” in the international database and then excluded from scaling of the corresponding national data.

There were instances of items being correctly translated but showing item-by-country interaction estimates larger than 1.3 logits (a measurement akin to about two standard deviations of the overall distribution of item difficulties in the test). In all cases, national items were removed from scaling of the national data but included in the international database. Table 11.6 lists the items that were excluded from scaling across the various national samples because of translation/printing errors or large item-by-country interactions.

### **International item calibration and test reliability**

Item parameters were obtained from calibration samples consisting of randomly selected subsamples from each country. The calibration of student item parameters involved randomly selecting subsamples of 500 students from each national sample. This process ensured that each country that had met sample participation requirements was equally represented in the sample. The random selection was based on the final student weights, and the final calibration sample included data from 18,000 students.

Missing student responses that were likely to be due to problems with test length (“not reached items”) were omitted from the calibration of item parameters, but were treated as “incorrect” during scaling of the student responses. The not-reached items were defined as all consecutive missing values that occurred from the end of the test back. However, the first missing value of each of these not-reached series was coded as “missing.”

Data from countries that did not meet the sampling requirements after inclusion of replacement schools (Category 3) were not included in the calibration of item parameters. Table 11.7 shows the final item parameters used to scale the ICCS test data that were based on the international calibration sample. The table also shows the standard errors for these parameters.

In order to account for possible positioning effects caused by the allocation of items to different booklets during scaling, a facet model that included a booklet effect, as estimated via the software package ACER ConQuest, was used. The booklet effects that emerged were generally rather small, with Booklet 5 being about 0.03 logits easier and Booklet 7 about 0.03 logits more difficult than the average booklet difficulty. The inclusion of the booklet facet in the scaling did not change the estimated item parameters but ensured that differences in booklet differences did not affect the scaling of student abilities. Table 11.8 shows the booklet parameters used for the final scaling of test data.



Table 11.5: Example of item statistics provided to national centers

ICCS 2009 MS - Item Statistics for Country: CNT Target Grade																												
How to read the item names:																												
Other abbreviations:																												
Subject: C = Civic and Citizenship Education																												
Internationally administered item																												
Item developed for: 1 = ICSS 1999, 2 = ICSS 2009																												
Item unit: Unique identifier of unit																												
Item type: M = Multiple-choice item, C = Constructed item, O = Open-ended item																												
Number of item within unit																												
0.1,2,3,4,5 % of students choosing different answer options																												
NR (for MC) % of students not reaching the item																												
INV (for CR) % of students omitting the item																												
OM (for CR) % of students who gave invalid answer																												
Point bi-serial correlations																												
Internal corr. for students choosing different answer options																												
International																												
Rasch results																												
EST Rasch item difficulty																												
ERR Rasch standard error of item difficulty																												
STPxx Step difficulty																												
FIT Rasch mean square fit index																												
GIRL Item difficulty difference to female																												
BOY Item difficulty difference to male																												
PB Item difficulty difference to distractor																												
PB International point-biserial corr. for correct distractor																												
DIFF International item difficulty																												
Countries in International Statistics:																												
AUT BEL BGR CHE CHL COL CYP ENG ESP FIN GRC GTM HKG IRL ITA KOR LIE LTU MEX NOR NZL POL RUS SVK SVN SWE THA TWN																												
Point bi-serial correlations																												
ITEM	N	DIFF	DISCR	0	1	2	3	4	OM	NR	INV	0	1	2	3	4	OM	EST	ERR	STP1	STP2	STP3	FIT	GIRL	BOY	EST	PB	DIFF
1 CII01M1	1469	78.4	0.5	10.8	78.4*	6.1	3.8	0.8	0.1	.	.	-0.33	0.52*	-0.30	-0.15	-0.10	-1.03	0.03	.	.	.	0.72	0.06	-0.06	-0.48	0.39	68.97	
2 CII10M1	1468	69.8	0.4	5.5	16.8	7.2	69.8*	0.3	0.1	0.4	.	-0.32	-0.18	-0.21	0.44*	-0.05	-0.50	0.03	.	.	.	0.99	0.13	-0.13	-0.06	0.39	61.61	
3 CII106M1	1465	73.3	0.5	8.9	12.6	73.3*	3.8	0.9	0.3	0.5	1	-0.32	-0.29	0.53*	-0.19	-0.06	-0.70	0.03	.	.	.	0.79	0.04	-0.04	-0.64	0.48	72.63	
4 CII108M1	1469	49.1	0.3	12.1	8.6	49.1*	28.1	1.5	0.8	1	.	-0.19	-0.23	0.30*	0.00	-0.12	0.56	0.03	.	.	.	1.14	0.13	-0.13	0.15	0.34	58.26	
5 CII109M1	1469	70.1	0.5	6.5	10.1	11.8	70.1*	1.2	0.3	1	.	-0.22	-0.30	-0.26	0.54*	-0.14	-0.52	0.03	.	.	.	0.85	-0.12	0.12	-0.13	0.43	63.61	
6 CII110M1	1469	58.5	0.5	8.9	14.4	16.6	58.5*	1.3	0.4	1	.	-0.17	-0.34	-0.17	0.50*	-0.09	0.10	0.03	.	.	.	0.95	0.19	-0.19	0.38	0.46	53.16	
7 CII113M1	1468	74.8	0.5	7.9	74.8*	7.3	8.8	0.8	0.1	0.4	1	-0.27	0.47*	-0.25	-0.20	-0.08	-0.79	0.03	.	.	.	0.89	-0.09	0.09	-0.04	0.41	61.73	
8 CII115M1	1468	51.4	0.5	20.2	51.4*	15.4	11.8	0.7	0.1	0.5	1	-0.13	0.45*	-0.30	-0.17	-0.06	0.45	0.03	.	.	.	1.01	0.07	-0.07	0.77	0.41	45.40	
9 CII119M1	1468	57.0	0.6	31.5	5.0	57.0*	5.3	0.3	0.1	0.8	1	-0.40	-0.23	0.60*	-0.21	-0.09	0.17	0.03	.	.	.	0.80	0.04	-0.04	0.05	0.47	60.43	
10 CII120M1	1467	57.9	0.4	30.1	5.9	57.9*	4.4	0.5	0.1	1.4	1	-0.17	-0.18	0.36*	-0.21	-0.04	0.13	0.03	.	.	.	1.10	-0.08	0.08	-0.83	0.37	76.15	
11 CII121M1	1467	34.9	0.3	34.9*	7.2	34.6	22.1	0.9	0.1	0.3	1	0.31*	-0.16	-0.17	-0.03	-0.05	1.27	0.03	.	.	.	1.20	0.31	-0.31	0.89	0.34	43.34	
12 CII127M1	1467	26.8	0.3	5.3	28.1	26.8*	37.8	1.7	0.1	0.3	1	-0.23	-0.02	0.26*	-0.09	-0.04	1.72	0.03	.	.	.	1.26	0.35	-0.35	1.17	0.28	38.05	



Table 11.6: National items excluded from scaling

Country	Item	Issue
Bulgaria	CI110M1	Translation error
Bulgaria	CI2FDM1	Translation error
Chile	CI2CCM2	Large item-by-country interaction
Chinese Taipei	CI2DLM1	Large item-by-country interaction
Chinese Taipei	CI2FSM1	Large item-by-country interaction
Chinese Taipei	CI2RCM1	Large item-by-country interaction
Chinese Taipei	CI2TGM1	Large item-by-country interaction
Colombia	CI101M1	Large item-by-country interaction
Colombia	CI113M1	Large item-by-country interaction
Czech Republic	CI2BIO1	Large item-by-country interaction
Dominican Republic	CI113M1	Large item-by-country interaction
Estonia	CI101M1	Translation error
Finland	CI2BIO1	Large item-by-country interaction
Guatemala	CI101M1	Large item-by-country interaction
Guatemala	CI113M1	Large item-by-country interaction
Hong Kong SAR	CI110M1	Large item-by-country interaction
Hong Kong SAR	CI2PCM2	Large item-by-country interaction
Indonesia	CI2BPM1	Translation error
Indonesia	CI2CCM1	Large item-by-country interaction
Indonesia	CI2SCM1	Translation error
Indonesia	CI2PCM1	Translation error
Indonesia	CI2CEM1	Translation error
Italy	CI2DLM1	Large item-by-country interaction
Korea, Republic of	CI104M1	Large item-by-country interaction
Korea, Republic of	CI110M1	Translation error
Korea, Republic of	CI128M1	Translation error
Korea, Republic of	CI2ASM2	Translation error
Korea, Republic of	CI2CEM2	Large item-by-country interaction
Korea, Republic of	CI2ETO1	Large item-by-country interaction
Korea, Republic of	CI2FSM1	Translation error
Korea, Republic of	CI2OMM1	Translation error
Korea, Republic of	CI2ORM1	Large item-by-country interaction
Korea, Republic of	CI2PGM2	Large item-by-country interaction
Korea, Republic of	CI2RCM1	Translation error
Korea, Republic of	CI2RDM2	Translation error
Korea, Republic of	CI2TGM2	Large item-by-country interaction
Korea, Republic of	CI2VOM3	Translation error
Latvia	CI2SCM1	Translation error
Liechtenstein	CI127M1	Large item-by-country interaction
Liechtenstein	CI2PCM1	Large item-by-country interaction
Liechtenstein	CI2PDO1	Large item-by-country interaction
Lithuania (Lithuanian version)	CI110M1	Printing problem
Mexico	CI2PCM1	Large item-by-country interaction
Norway	CI128M1	Large item-by-country interaction



Table 11.6: National items excluded from scaling (contd.)

Country	Item	Issue
Russian Federation	CI132M1	Incorrect translation of options
Slovak Republic	CI2CCM2	Large item-by-country interaction
Slovak Republic	CI2BIO1	Scoring problems
Slovenia	CI127M1	Large item-by-country interaction
Switzerland (German version)	CI121M1	Translation error
Switzerland (German version)	CI129M1	Translation error
Thailand	CI2BPM1	Large item-by-country interaction
Thailand	CI2PCM2	Large item-by-country interaction
Thailand	CI2SRM1	Large item-by-country interaction
Thailand	CI2VOM2	Large item-by-country interaction

The overall reliability of the international test, as obtained from the scaling model, was 0.84 (ACER ConQuest estimate). Table 11.9 shows the median reliabilities (Cronbach's alpha) and median item numbers for national samples across booklets. The median test reliability was 0.83 and ranged from 0.70 to 0.88. The median reliabilities were below 0.8 in only six countries. In these countries, the number of items had generally been reduced as a consequence of item deletions brought about by translation/printing errors or very large item-by-country interactions (see section above on item adjudication).

### International ability estimates

In many educational assessments, the purpose of testing is to obtain accurate estimates of individual domain-based cognitive abilities. The accuracy of measuring the latent ability  $\theta$  can be improved by using a larger number of test items. However, in large-scale surveys such as ICCS, the purpose is to obtain accurate population estimates by using instruments that cover a wider range of possible aspects of cognitive abilities.

The use of matrix-sampling design, where individual students are allocated booklets and respond to a set of items obtained from the main pool of items, has become standard in assessments of this type. However, reducing test length and administering subsets of items to individual students introduces a considerable degree of uncertainty at the individual level. Aggregated student abilities of this type can lead to bias in population estimates. However, this problem can be addressed by employing plausible value methodology that uses all available information from student tests and questionnaires, a process that leads to more accurate population estimates (Mislevy, 1991; Mislevy & Sheehan, 1987; von Davier, Gonzalez, & Mislevy, 2009).

Using item parameters anchored at their estimated values from the calibration sample makes it possible to randomly draw plausible values from the marginal posterior of the latent distribution for each individual. Estimations are based on the conditional item response model and the population model, which includes the regression on background variables used for conditioning. (For a detailed description, see Adams, Wu, & Macaskill, 1997; also Adams, 2002.) In order to obtain estimates of students' civic knowledge, ACER Conquest software was used, thereby allowing plausible values to be drawn (see Wu et al., 2007).





Table 11.7: Final item parameters used to scale the international test items

Item	Item Parameter	Step 1	Step 2	Item	Item Parameter	Step 1	Step 2
CI2COM1	-1.13 (0.03)			CI2PCM1	-0.40 (0.03)		
CI2MOM1	0.45 (0.03)			CI2PCM2	-1.35 (0.03)		
CI2MLM1	-0.89 (0.03)			CI2VOM1	-0.95 (0.03)		
CI2MLM2	0.37 (0.03)			CI2VOM2	-0.13 (0.03)		
CI2PDO1	0.67 (0.02)	-0.71 (0.03)	0.71 (0.03)	CI2VOM3	-1.68 (0.03)		
CI2RDM2	0.21 (0.03)			CI2DLM1	0.52 (0.03)		
CI2SHM1	-1.07 (0.03)			CI2HRM1	-0.89 (0.03)		
CI2SHM2	-2.26 (0.04)			CI2JOM1	-1.10 (0.03)		
CI2TGM1	0.18 (0.03)			CI2WFO2	0.47 (0.02)	1.30 (0.04)	-1.30 (0.04)
CI2TGM2	-0.06 (0.03)			CI2PGM1	-0.47 (0.03)		
CI2BPM1	-0.55 (0.03)			CI2PGM2	0.34 (0.03)		
CI2BPM2	-0.66 (0.03)			CI2ECM1	-1.17 (0.03)		
CI2GFM1	-1.07 (0.03)			CI2ECM2	-0.44 (0.03)		
CI2BIO1	1.61 (0.02)	-0.93 (0.03)	0.93 (0.03)	CI2CEM1	-1.20 (0.03)		
CI2GLM1	-1.09 (0.03)			CI2CEM2	0.04 (0.03)		
CI2GLM2	-0.56 (0.03)			CI2WFO1	0.41 (0.02)	-0.89 (0.02)	0.89 (0.02)
CI2FDM1	-0.68 (0.03)			CI2ORM1	-0.48 (0.03)		
CI2FSM1	-0.52 (0.03)			CI2RCM1	-0.44 (0.03)		
CI2SCM1	-0.23 (0.03)			CI2PJM1	0.01 (0.03)		
CI2SCM2	-0.32 (0.03)			CI2PJM2	-1.35 (0.03)		
CI2ASM1	-0.58 (0.03)			CI2REM2	-0.46 (0.03)		
CI2ASM2	-1.04 (0.03)			CI2REM3	-1.03 (0.03)		
CI2CNM1	-0.54 (0.03)			CI101M1	-0.97 (0.03)		
CI2CNM2	-1.45 (0.03)			CI109M1	-0.58 (0.03)		
CI2ETO1	1.42 (0.02)	-0.76 (0.03)	0.76 (0.03)	CI108M1	-0.34 (0.03)		
CI2ETM2	-0.37 (0.03)			CI128M1	-0.56 (0.03)		
CI2BCM1	-0.18 (0.03)			CI137M1	-0.99 (0.03)		
CI2PRM1	0.21 (0.03)			CI110M1	-0.12 (0.03)		
CI2CCM1	-0.86 (0.03)			CI113M1	-0.61 (0.03)		
CI2CCM2	-0.35 (0.03)			CI104M1	-0.38 (0.03)		
CI2SRM1	-0.59 (0.03)			CI115M1	0.34 (0.03)		
CI2SRM2	-0.06 (0.03)			CI119M1	-0.26 (0.03)		
CI2SRM3	-0.33 (0.03)			CI120M1	-1.26 (0.03)		
CI2OMM1	-0.08 (0.03)			CI121M1	0.43 (0.03)		
CI2OMM2	-0.69 (0.03)			CI127M1	0.56 (0.03)		
CI2OMM3	-0.36 (0.03)			CI132M1	-0.68 (0.03)		
CI2RRO1	0.70 (0.02)	-0.70 (0.03)	0.70 (0.03)	CI129M1	0.07 (0.03)		
CI2DCM1	-0.95 (0.03)			CI130M1	-0.57 (0.03)		
CI2PFM1	-1.46 (0.03)			CI106M1	-1.02 (0.03)		
CI2PFM2	-0.47 (0.03)						

Table 11.8: Estimated ICCS booklet parameters

	Estimated Booklet Effect	
Booklet 1	-0.01	(0.01)
Booklet 2	-0.02	(0.01)
Booklet 3	0.02	(0.01)
Booklet 4	0.02	(0.01)
Booklet 5	-0.03	(0.01)
Booklet 6	-0.01	(0.01)
Booklet 7	0.03	(0.01)

All available international student questionnaire variables as well as those derived from regional instruments were used for conditioning. Dealing with missing responses required substituting all missing responses in a variable with either the mode or the mean. Additional indicators for missing values were added as additional variables. Table D.1 in Appendix D lists all the international and regional student-level variables (along with their respective codings) that were used to condition the plausible values of civic knowledge.

Because of the large number of variables, the principal components of all student-level variables (single items or scale indices) were used as conditioning variables. These reflected 99 percent of the variance. At the student level, only gender and its missing indicator were used as a direct conditioning variable. Classroom indicators were added as further direct conditioning variables in order to account for differences among schools and classrooms.

After plausible values had been drawn, the resulting scale was transformed to a metric with a mean of 500 and a standard deviation of 100 for equally weighted ICCS countries that had met sampling requirements (Categories 1 and 2). This linear transformation can be computed by applying the formula

$$\theta'_n = 500 + 100 \left( \frac{\theta_n - \bar{\theta}}{\sigma_\theta} \right),$$

where  $\theta'_n$  are the student scores in the international metric,  $\theta_n$  are the original logit scores,  $\bar{\theta}$  is the international mean of student logit scores (-0.01) with equally weighted country subsamples, and  $\sigma_\theta$  is its corresponding international standard deviation (0.95). This transformation was applied to each of the five plausible values. Chapter 13 provides a description of how the plausible values were used to calculate imputation variance.

In addition to containing plausible values, the ICCS student database contains nationally standardized ability scores that were derived as weighted likelihood estimates. These estimates were computed by minimizing the equation

$$\sum_{i \in \Omega} \left[ \left( r_x + \frac{J_n}{2I_n} \right) - \sum_{j=1}^k \frac{\exp(\sum_{i=0}^x (\theta_n - \delta_i + \tau_{ij}))}{\sum_{b=0}^k \exp(\sum_{i=0}^x (\theta_n - \delta_i + \tau_{ij}))} \right] = 0$$

for each case  $n$ , where  $r_x$  is the sum score obtained from a set of  $k$  items with  $j$  categories.

Weighted likelihood estimates are obtained by applying the Newton-Raphson method. The term  $J_n/2I_n$  (with  $I_n$  being the information function for student  $n$  and  $J_n$  its derivative with respect to  $\theta$ ) is used as a weight function to account for the bias inherent in maximum likelihood estimation (Warm, 1989). ACER ConQuest was used to derive scale scores with the same international item parameters that were used to compute the plausible values.



Table 11.9: Median test reliabilities across booklets for national samples (Cronbach's alpha)

Country	Median Test Reliability across Booklets	Median Number of Items per Booklet
Austria	0.86	32
Belgium (Flemish)	0.82	32
Bulgaria	0.88	32
Chile	0.83	32
Chinese Taipei	0.83	31
Colombia	0.81	32
Cyprus	0.84	32
Czech Republic	0.81	32
Denmark	0.84	32
Dominican Republic	0.70	29
England	0.87	32
Estonia	0.84	32
Finland	0.81	32
Greece	0.87	32
Guatemala	0.78	32
Hong Kong SAR	0.84	30
Indonesia	0.72	27
Ireland	0.87	32
Italy	0.82	32
Korea, Republic of	0.77	27
Latvia	0.78	29
Liechtenstein	0.85	32
Lithuania	0.80	32
Luxembourg	0.86	31
Malta	0.85	30
Mexico	0.82	32
Netherlands	0.85	32
New Zealand	0.88	32
Norway	0.84	32
Paraguay	0.84	32
Poland	0.85	32
Russian Federation	0.82	32
Slovak Republic	0.82	32
Slovenia	0.83	32
Spain	0.84	32
Sweden	0.85	32
Switzerland	0.83	32
Thailand	0.78	31
ICCS median	0.83	32



Weighted likelihood estimates of civic knowledge were transformed to a national metric of 150 and a standard deviation of 10 for each country. The transformation was achieved by applying this formula:

$$\theta'_n = 150 + 10 \frac{\theta_n - \bar{\theta}_c}{\sigma_{\theta(c)}}$$

Here,  $\theta'_n$  are the scores in the national metric,  $\theta_n$  are the original weighted likelihood estimates in logits,  $\bar{\theta}_c$  is the country mean of logit scores, and  $\sigma_{\theta(c)}$  is the corresponding national standard deviation of the original scores.

A general recommendation for those drawing on ICCS data to analyze civic knowledge is to use plausible values that are scaled in an internationally comparative metric, as these provide the best possible population estimates.

### Development of proficiency levels for civic knowledge

One of the objectives of ICCS was to establish a described civic knowledge scale that would become a reference point for future international assessments in this learning area. Establishing proficiency levels of civic knowledge is an informative way of describing student performance across countries and also sets benchmarks for future surveys.

Students whose results are located within a particular level of proficiency are typically able to demonstrate certain understandings and skills that are associated with that level. These students also typically possess the understandings and skills defined as applying at lower proficiency levels.

Development of the proficiency levels required application of a method which ensured that the notion of “being at a level” could be interpreted consistently and would align with the fact that the achievement scale is a continuum. An attempt was therefore made to provide a common understanding about what being at a level meant and to ensure that this meaning was consistent across different proficiency levels. This method took the following three questions into account:

- What is the expected success of a student at a particular level on a test containing items at that level?
- What is the width of the levels in that scale?
- What is the probability that a student in the middle of a level will correctly answer an item of average difficulty for that level?

The following two parameters were adopted when defining the proficiency level:

- *The response probability for reporting item parameters:* this was set at  $rp = 0.62$ ;
- *The width of the proficiency levels:* this was set at 0.8 logits.

Use of these parameters made it possible to infer the following about students’ aptitude in relation to the proficiency levels:

- Students whose results placed them at the lowest possible point of the proficiency level were likely to correctly answer (on average) slightly over 50 percent of the items on a test made up of items spread uniformly across the level, from the easiest to the most difficult item.
- Students whose results placed them at the lowest possible point of the proficiency level had a 62 percent probability of giving the correct response to an item at the bottom end of the proficiency level.
- Students whose results placed them at the top of the proficiency level had a 78 percent probability of correctly responding to an item at the bottom end of the proficiency level.



The approach chosen was essentially an attempt to apply an appropriate choice of mastery by placing item locations at  $rp = 0.62$  while simultaneously ensuring that the approach would be understood by the readers of ICCS reports.

Three proficiency levels that could be used when reporting student performances from the assessment were identified. Table 11.10 shows the cut-points for these levels (in logits and final scale scores). The table also cites the percentage of students at each proficiency level across the participating ICCS countries.

*Table 11.10: Proficiency level cut-points and percentage of students at each level*

Proficiency Level	Cut-Points		Average Percentage of Students in Each Proficiency Level
	Logits	Scale scores	
Level 3	0.59	563	28
Level 2	-0.21	479	31
Level 1	-1.01	395	26
Below Level 1			16

In order to report released test items and to map them against proficiency levels, it was necessary to transform the location parameters of these items to a value that reflected a response probability of 62 percent. This was achieved by adding the natural log of the odds of 62 percent chance to the original log odds and then transforming the result to the international metric by applying the same transformation as for the (original) student scores. The standardized item difficulty  $d'_i$  obtained for each item was as follows:

$$d'_i = 500 + 100 \times \left( \frac{d_i + 1_n(0.62/0.38) - \bar{\theta}}{\sigma_\theta} \right).$$

Here,  $d_i$  is the item difficulty in its original metric,  $\bar{\theta}$  is the international mean of student logit scores (-0.01) with equally weighted country subsamples, and  $\sigma_\theta$  is its corresponding international standard deviation (0.95).<sup>2</sup>

## Estimation of changes between 1999 and 2009

Seventeen CIVED items were included in the international test to allow for the reporting of changes from the previous IEA civic and citizenship education survey (CIVED) in 1999. Data for estimating changes were available for 17 out of 38 countries. Norway, Slovenia, and Greece tested Grade 9 students in CIVED and collected data from additional Grade 9 student samples in ICCS. Italy tested both Grades 8 and 9 in CIVED, and the country's Grade 8 data were used for estimating change. Both England and Sweden tested at different times of the school year (England at the beginning of the following school year and Sweden at its start). Comparisons for these two participating countries are therefore reported in a separate section of the reporting table.

Modest positioning effects on item difficulties were observed in the CIVED data for a number of countries. Whereas in CIVED, the assessment consisted of only one booklet, with each item appearing in only one position within the test, ICCS used a rotated design, which ensured that students responded to link items in each of the three possible positions at the start, middle, or end of the assessment.



<sup>2</sup> Due to a transformation error, the location parameters for most example test items reported in Table 3.9 in the international report (Schulz, Ainley, Fraillon, Kerr, & Losito, 2010) were not quite correct: example item 1 (constructed item) should have been located at 529 scale points (instead of 521) for a partial credit and 717 (instead of 701) for a full credit. Example item 2 should have been at 440 (instead of 435), example item 4 at 445 (instead of 440), example item 5 at 517 (instead of 509). Example item 6 should have been at 600 (instead of 589), and example item 7 at 598 (instead of 587). However, these corrections did not change the mapping of example items to proficiency levels.

As a consequence, and after effort had been made to control for ability, the CIVED students had a higher probability than the ICCS students of giving correct answers to items that had been administered at the beginning. However, the CIVED students were less likely to know the answers to test questions administered toward the end of the assessment. This effect was, however, not very strong for the pooled international samples and was notable in only a few countries.

Given that the framework for the test domain was broader in ICCS than in CIVED and given that the majority of the link items represented only one content domain (civic systems and society), it was not surprising to find in the preliminary analysis some notable differences in the behavior of the CIVED link items and the new ICCS items. Effort to estimate comparable ICCS test scores using the newly established scale for the CIVED data was not deemed appropriate.

A decision was made to test an approach that involved using different equating methods to set test data for the link items (with reduced sample size, given these items appeared in only three out of the seven ICCS booklets) against the CIVED scale metric. However, comparisons of the differences in percentages correct for both surveys and the resulting trend estimates showed several inconsistencies that were probably a consequence of the set of link items including only two items from the subdimension “interpretative skills.” It was therefore decided to report only changes pertaining to the “civic content knowledge” subscale, for which 15 link items were available.

In order to review the link item characteristics, the adjusted item difficulty parameters (each standardized to have a mean of 0) were compared first at the international level and then for each national sample. Figure 11.6 shows the scatterplot between the item parameters from CIVED and those estimated for the trend sample consisting of 500 randomly selected students from each of the national samples with comparable data. As is apparent in the figure, five items were slightly outside the error bands derived from the respective standard errors of the item parameters from both calibrations. However, the figure also shows that the item parameters were generally highly similar; the correlation between item parameters was 0.96.

As was the case for the international scaling, some national items had to be omitted from scaling. These items included those that were excluded from the CIVED scaling (see Schulz & Sibberns, 2004), ICCS versions of items with translation errors or deviations, items that reflected printing problems, and items that showed very large differences in relative item difficulty between the two surveys. Table 11.11 shows the national items that were excluded from the equating procedures for ICCS data as well as the reasons for their exclusion.

For the final scaling, maximum likelihood estimates (MLEs) were derived using the same item parameters as in CIVED. They were then transformed to the same scale metric, which was set to have a mean of 100 and a standard deviation of 20 for the 28 CIVED countries. The transformation of scale scores was computed by applying this formula:

$$\theta'_n = 100 + 20 \left( \frac{\theta_n - \bar{\theta}}{\sigma_{\theta}} \right).$$

Here,  $\theta'_n$  are the student scores in the CIVED metric,  $\theta_n$  are the original logit scores (maximum likelihood estimates),  $\bar{\theta}$  is the CIVED mean of student logit scores (0.95) with equally weighted country subsamples, and  $\sigma_{\theta}$  is its corresponding CIVED standard deviation (1.36). Table 11.12 shows the item parameters used for scaling as well as the average percentages of correct responses for these items in the 17 countries in 2009 and 1999.



Figure 11.6: Scatterplot for link item parameter estimates from CIVED and for the ICCS trend sample

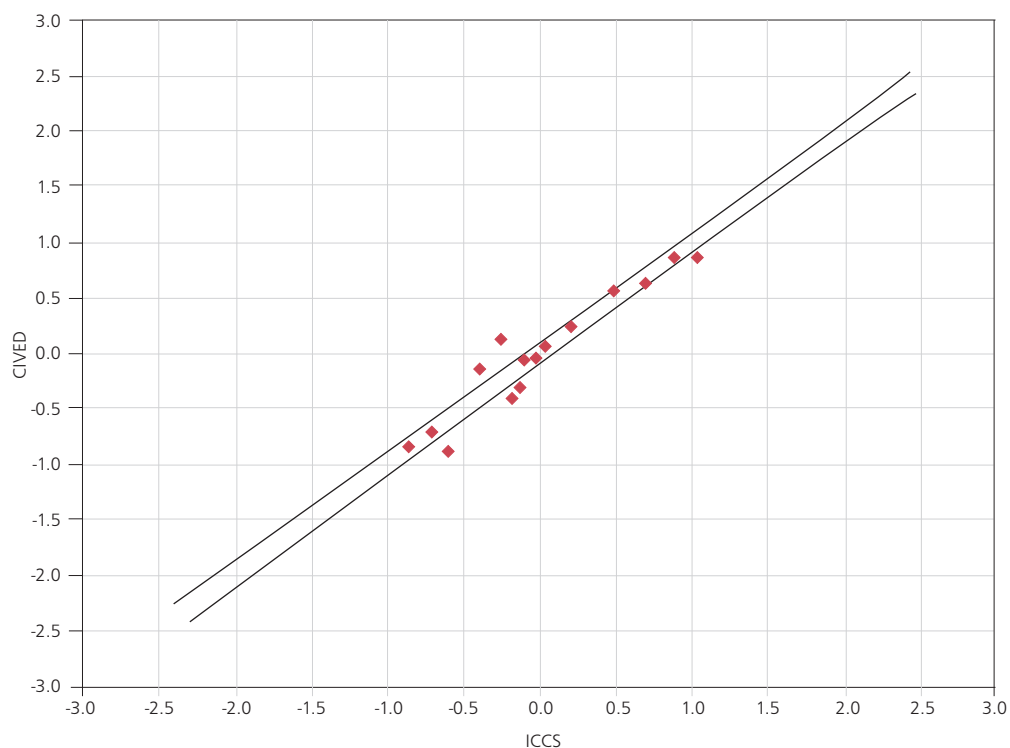


Table 11.11: National items excluded from the ICCS equating procedures

Country	Item	Reason
Bulgaria	CI110M1	Translation error in ICCS
Bulgaria	CI120M1	Not used in CIVED scaling
Switzerland (German version)	CI121M1	Translation error in ICCS
Switzerland (German version)	CI129M1	Translation error in ICCS
Estonia	CI101M1	Translation error in ICCS
Estonia (Russian version)	CI113M1	Incorrect translation in CIVED
Estonia	CI121M1	Not used in CIVED scaling
Estonia (Russian version)	CI128M1	Translation error in CIVED
Estonia (Russian version)	CI129M1	Translation error in CIVED
Finland	CI127M1	Not used in CIVED scaling
Lithuania (Lithuanian version)	CI110M1	Printing problem in ICCS
Slovak Republic	CI110M1	Different translation used in ICCS
Republic	CI120M1	Different translation used in ICCS
Slovenia	CI127M1	Not used in CIVED scaling
Sweden	CI106M1	Large difference in relative item difficulty
Sweden	CI121M1	Not used in CIVED scaling





Table 11.12: Item parameters and average percentage correct for link items

Item	Item parameters	Average Percentages of Correct Responses in:	
		ICCS 2009	CIVED 1999
CI101M1	-0.79 (.02)	72	78
CI104M1	-0.22 (.01)	63	70
CI106M1	-0.62 (.02)	72	79
CI108M1	0.15 (.01)	60	63
CI109M1	-0.32 (.01)	65	71
CI110M1	0.34 (.01)	57	58
CI113M1	-0.06 (.01)	68	67
CI115M1	0.73 (.01)	46	50
CI119M1	0.05 (.01)	61	66
CI120M1	-0.75 (.02)	75	77
CI121M1	0.95 (.01)	43	49
CI127M1	0.96 (.01)	38	47
CI128M1	0.22 (.01)	66	63
CI129M1	0.66 (.01)	51	54
CI130M1	0.02 (.01)	61	64

Scale scores could be derived only for those students who responded to the link item cluster (included in three out of the seven randomly allocated booklets) and only for those 17 national datasets where the respective student populations were comparable with the ones surveyed in CIVED in 1999. Table 11.13 records the scale reliabilities (Cronbach's alpha) for this subset of test items as well as the number of items that were used for scaling (after national item exclusions). The median reliability of this set of test items was 0.77, and the reliabilities ranged from 0.69 to 0.82 across the national samples.

Because the transformation equating the ICCS 2009 data with the CIVED 1999 data depended on the change in the degree of difficulty of each of the individual link items, the sample of link items chosen influenced the choice of transformation. This meant that the resulting transformation would have been slightly different if an alternative set of link items had been chosen. Uncertainty in the transformation thus relates to the sampling of the link item, in the same way that uncertainty in values such as country averages is an outcome of the particular sample of students that is used.

The uncertainty resulting from link-item sampling is referred to as linking error, and it is an error that analysts have to take into account when comparing the results arising out of different data collections (see Monseur & Berezner, 2007). As is the situation with the error that is introduced through the process of sampling students, the exact magnitude of this linking error cannot be determined. It is possible, however, to estimate the likely range of magnitudes for this error and to take it into account when interpreting results. As with sampling errors, the likely range of magnitude for the errors is represented as a standard error.

Because all link items were dichotomous and not clustered in units, it was possible to compute the linking error for ICCS by using the following simple formula:

$$\sigma_{(linking\_error)} = \frac{\sigma^2}{n}.$$



Here,  $\sigma^2$  represents the variance of the item parameter differences between 1999 and 2009 (using international calibration samples), and  $n$  denotes the number of link items used. The linking error for trend reporting from 1999 to 2009 was 0.65 score points in the final reporting metric (0.044 logits), and it was taken into account during estimation of the statistical significance of differences (see Chapter 13 for further details).

*Table 11.13: Test reliabilities for link items (Cronbach's alpha)*

Country	Reliability	Number of Items
Bulgaria	0.78	15
Chile	0.76	17
Colombia	0.76	17
Czech Republic	0.77	17
England	0.77	17
Estonia	0.70	15
Finland	0.80	16
Greece	0.82	17
Italy	0.80	17
Latvia	0.69	17
Lithuania	0.73	16
Norway	0.80	17
Poland	0.82	17
Slovak Republic	0.74	14
Slovenia	0.76	16
Sweden	0.78	15
Switzerland	0.70	15
ICCS median	0.77	17

## Regional cognitive items and scales

The ICCS instruments used for the European and Latin American regional modules included short cognitive tests, the development of which were guided by the ICCS assessment framework (Schulz et al., 2008). However, whereas the Latin American test items assessed conceptually highly similar aspects of civic knowledge with specific regionally relevant content, those developed for the European student instrument focused on more specific knowledge about the European Union and its policies and institutions.

Analyses that involved the use of multidimensional IRT models in ACER ConQuest showed a latent correlation between the European and international tests of 0.82 and between the Latin American and international tests of 0.86. There was thus a relatively high similarity between the dimensions measured with the international and regional test components, in particular those from the Latin American test.

Table 11.14 shows the results from a tentative calibration of the European test items. Although the item-fit statistics did not indicate a relatively large amount of item misfit, most of the European test items had item-total correlations of below 0.2, which suggested generally unsatisfactory measurement properties.

Table 11.15 presents the reliability estimates (Cronbach's alpha) for the European regional test items. For the pooled European sample, the estimate was 0.51. Estimates ranged from 0.34 (in Liechtenstein) to 0.61 (in Bulgaria). These results show that the regional test had generally low internal consistencies.



*Table 11.14: Item parameters, item fit, and item-total correlations from tentative calibration of European regional test items*

Item	Calibrated item parameter	Weighted item fit	Item-total correlation
ES2T01A	-3.51 (0.05)	0.99	0.15
ES2T01B	-1.80 (0.03)	0.98	0.19
ES2T01C	-0.68 (0.02)	1.06	0.01
ES2T02	-2.74 (0.04)	0.99	0.15
ES2T03	-0.29 (0.02)	0.99	0.20
ES2T04	0.41 (0.02)	1.01	0.15
ES2T05	-0.70 (0.02)	0.96	0.29
ES2T06	0.64 (0.02)	1.04	0.05
ES2T07	0.27 (0.02)	0.99	0.20
ES2T08	-0.30 (0.02)	0.97	0.25
ES2T09A	-0.61 (0.02)	1.00	0.17
ES2T09B	-2.23 (0.03)	0.97	0.22
ES2T09C	-1.87 (0.03)	0.98	0.21
ES2T09D	-0.88 (0.02)	1.01	0.13
ES2T09E	-0.09 (0.02)	1.04	0.08
ES2T10	0.92 (0.02)	1.04	0.04
ES2T11A	-0.86 (0.02)	0.96	0.27
ES2T11B	0.07 (0.02)	1.02	0.11
ES2T11C	-0.74 (0.02)	1.02	0.11
ES2T12	-0.68 (0.02)	0.98	0.23

Given the short test length, the unsatisfactory scaling properties of many of the European test items, and the rather low overall reliability of the European test, a decision was made to report only item percentages in the final report (see Kerr, Sturman, Schulz, & Burge, 2010).

The Latin American ICCS civic knowledge test consisted of 16 multiple-choice items that focused on specific aspects of knowledge relevant to the Latin American region. Although the test items were specifically developed for students in Latin American countries, and thus addressed aspects not relevant in other geographical regions of the world, the regional test items were designed to measure the same content and cognitive dimensions as those in the international test. Responses to these items therefore reflected the same latent construct of civic knowledge, which made it possible to calibrate the regional items on the ICCS international civic knowledge scale. That, in turn, meant that the item parameters could be compared with the ICCS international civic knowledge scale and reported against the international proficiency levels (see Schulz, Ainley, Friedman, & Lietz, 2011).

Table 11.16 shows the calibration results for the 16 Latin American test items when scaled alongside the international test items with fixed item parameters. All test items had satisfactory item statistics.

Table 11.17 records the reliabilities (Cronbach's alpha) for the Latin American regional tests. The internal consistency of this item set was 0.72 for the pooled ICCS sample; national reliabilities ranged from 0.64 (in the Dominican Republic) to 0.76 (in Chile). The test component thus had satisfactory internal consistency for the pooled sample and in three of the participating Latin American countries.



Table 11.15: Reliabilities for European test items

Country	Reliability (Cronbach's Alpha)
Austria	0.55
Belgium (Flemish)	0.44
Bulgaria	0.61
Cyprus	0.53
Czech Republic	0.43
Denmark	0.51
England	0.54
Estonia	0.42
Finland	0.48
Greece	0.55
Ireland	0.54
Italy	0.50
Latvia	0.46
Liechtenstein	0.34
Lithuania	0.48
Luxembourg	0.49
Malta	0.55
Netherlands	0.47
Poland	0.51
Slovak Republic	0.59
Slovenia	0.47
Spain	0.40
Sweden	0.55
Switzerland	0.35
ICCS average	0.51

Table 11.16: Item parameters, item fit, and item-total correlations for Latin American regional test items

Item	Calibrated Item Parameter	Weighted Item Fit	Item-Total Correlation
LS2T01	-0.34 (0.04)	0.93	0.43
LS2T02	-1.46 (0.04)	0.99	0.43
LS2T03	-1.73 (0.05)	0.93	0.40
LS2T04	-1.55 (0.04)	0.90	0.45
LS2T05	-0.87 (0.04)	1.06	0.40
LS2T06	0.67 (0.05)	1.05	0.41
LS2T07	-0.44 (0.04)	1.03	0.42
LS2T08	-0.83 (0.04)	0.87	0.30
LS2T09	-0.09 (0.04)	0.98	0.30
LS2T10	0.03 (0.04)	0.99	0.35
LS2T11	-0.30 (0.04)	1.02	0.37
LS2T12	0.40 (0.04)	1.02	0.30
LS2T13	-0.23 (0.04)	1.04	0.28
LS2T14	-0.98 (0.04)	0.98	0.29
LS2T15	-1.39 (0.04)	0.92	0.39
LS2T16	-0.85 (0.04)	0.98	0.33



Table 11.17: Reliabilities for Latin American test items

Country	Reliability (Cronbach's Alpha)
Chile	0.76
Colombia	0.67
Dominican Republic	0.64
Guatemala	0.66
Mexico	0.72
Paraguay	0.73
Latin American ICCS average	0.72

## Summary

The ICCS test items were scaled using item response modeling with the (one-parameter) Rasch model. An extensive analysis of scaling properties was carried out prior to scaling. This process included reviews of missing values, test coverage, assessment of item fit, differential item functioning by gender, and cross-national measurement equivalence.

Three proficiency levels were established, and test-item locations on the ICCS civic knowledge scale were used to describe these levels. Plausible values were generated as ability estimates, with full conditioning taking all available international and regional data at the student level into account.

In order to provide estimates of change in civic content knowledge over time, 15 CIVED items were used to obtain comparable test scores. In view of the low number of link items, anyone analyzing these data should take the equating error into account. Also, results need to be interpreted with caution given the change in test design between the two surveys.

ICCS used regional test components for Europe and Latin America. Because the European test items assessing specific student knowledge about the European Union did not form a consistent scale, results pertaining to them were reported at the item level. Furthermore, because the Latin American test items conceptually measured a cognitive dimension similar to the international ones, it was possible to scale them so that they could be used as regional descriptors of ICCS civic knowledge.



## References

- Adams, R. (2002). Scaling PISA cognitive data. In R. Adams & M. Wu (Eds.), *Technical report for the OECD Programme for International Student Assessment* (pp. 99–108). Paris, France: OECD Publications.
- Adams, R., Wu, M., & Macaskill, G. (1997). Scaling methodology and procedures for the mathematical and science scales. In M. O. Martin & D. L. Kelly (Eds.), *TIMSS technical report: Vol. II. Implementation and analysis: Primary and middle school years*. Chestnut Hill, MA: Boston College.
- Kerr, D., Sturman, L., Schulz, W., & Burge, B. (2010). *ICCS 2009 European report: Civic knowledge, attitudes, and engagement among lower-secondary students in 24 European countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- 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 and Berlin/Heidelberg, Germany: Springer.
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177–196.
- Mislevy, R. J., & Sheehan, K. M. (1987). Marginal estimation procedures. In A. E. Beaton (Ed.), *The NAEP 1983–1984 technical report* (Report No. 15-TR-20, pp. 293–360). Princeton, NJ: Educational Testing Service.
- Monseur, C., & Berezner, A. (2007). The computation of equating errors in international surveys in education. *Journal of Applied Measurement*, 8(3), 323–335.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.
- Rost, J., & von Davier, M. (1994). A conditional item-fit index for Rasch models. *Applied Psychological Measurement*, 18(2), 171–182.
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Friedman, T., & Lietz, P. (2011). *ICCS 2009 Latin American report: Civic knowledge and attitudes among lower-secondary students in six Latin American countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International civic and citizenship education study: Assessment framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., & Sibberns, H. (2004). Scaling procedures for cognitive items. In W. Schulz & H. Sibberns (Eds.), *IEA Civic Education Study: Technical report* (pp. 69–91). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series, Volume 2*, 9–36.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: Mesa Press.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ACER ConQuest: General item response modelling software* [computer program]. Camberwell, Victoria, Australia: Australian Council for Educational Research.



# Scaling procedures for ICCS questionnaire items

Wolfram Schulz and Tim Friedman

## Introduction

This chapter describes the procedures used to scale the ICCS questionnaire data (for students, teachers, and schools) and the indices based on them.

In general, it is possible to distinguish two general types of indices derived from the ICCS questionnaires:

1. Simple indices constructed through arithmetical transformation or recoding, for example, ratios between teachers and students; and
2. Scale indices derived from scaling of items, a process typically achieved by using item response modeling of dichotomous or Likert-type items.

The first part of this chapter lists the simple indices that were derived from the ICCS data and describes how they were created. The second part outlines the scaling procedures used in ICCS. The third and final part, lists the scaled indices, along with statistical information on item parameters, scale reliabilities, and the factor structure of related item sets.

The cross-country validity of item dimensionality and constructs was assessed during the field trial stage of ICCS. At this time, data were used to assess the extent to which measurement models held across participating countries. Extensive use was made of both confirmatory factor analysis and item response modeling, a process that made it possible to examine cross-national measurement equivalence before conducting the final selection of main survey questionnaire items (Schulz, 2009).

## Simple indices

### *Student questionnaire*

*Student age* (SAGE) 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 age were obtained from both the questionnaire and the student tracking forms. The formula for computing SAGE was

$$SAGE = (100 + T_y - S_y) + \frac{(T_m - S_m)}{12},$$

where  $T_y$  and  $S_y$  are, respectively, the year of the test and the year of birth of the tested student, in two-digit format (e.g., "06" or "92"), and where  $T_m$  and  $S_m$  are respectively the month of the test and the month of the student's birth. The result is rounded to two decimal places.

*Occupational data* for each student's parents were obtained by asking open-ended questions about the jobs of the student's mother and father. The responses were coded into four-digit ISCO codes (International Labour Organization, 1990) that were then mapped to the International Socioeconomic Index of Occupational Status (ISEI) (Ganzeboom, de Graaf, & Treiman, 1992). The three indices obtained from these scores were mother's occupational status (MSEI), father's occupational status (FSEI), and the highest occupational status of both parents (HISEI), with the latter corresponding to the higher ISEI score of either parent or to the only available parent's ISEI score. For all three indices, higher scores indicate higher levels of occupational status.

*Parental education* is another family background variable. The core difficulties with this variable relate to international comparability (education systems differ widely across countries and over time within countries) and response validity (students are often unable to accurately report





their parents' levels of education). ICCS classified levels of parental education according to the International Standard Classification of Education (ISCED) (UNESCO, 2006).

Indices of parental education were constructed by recoding educational qualifications into the following categories:

- (0) None;
- (1) ISCED 1 (primary education);
- (2) ISCED 2 (lower secondary);
- (3) ISCED 3 (upper secondary);
- (4) ISCED 4 (non-tertiary post-secondary) or ISCED 5B (vocational tertiary);
- (5) ISCED 5A (theoretically oriented tertiary) or ISCED 6 (post-graduate).

Indices with these categories were provided for each student's mother (MISCED) and father (FISCED). The index for *highest educational level of parental education* (HISCED) corresponded to the higher ISCED level of either parent.

For some ICCS analyses, a recoded version of HISCED was used in order to recode levels of education into approximate years of education (index PAREDYRS). Table D.1 in Appendix D shows, for each participating country, the number of years that were assumed to correspond to each category. Many cases of variation were found within the same levels, which made it necessary to choose approximate midpoints. For students who reported that their parents had not finished primary school, a value of two years was chosen on the assumption that most parents who had not finished primary school would have had at least some schooling.

*Students' expected education* was derived from a question that asked students which level of education they expected to achieve. The resulting index (SISCED) had the following categories:

- (0) No completion of ISCED 2;
- (1) Completion of ISCED 2 (lower secondary);
- (2) Completion of ISCED 3 (upper secondary);
- (3) Completion of ISCED 4 (non-tertiary post-secondary) or ISCED 5B (vocational tertiary);
- (5) Completion of ISCED 5A (theoretically oriented tertiary) or ISCED 6 (post-graduate).

For some analyses, this index was recoded into approximate years of further schooling based on the approximate years of schooling for the ISCED categories shown in Table D.2 in Appendix D. The index EXPEDYRS consisted of the approximate years of schooling minus the corresponding numbers of years in the target grade.

The ICCS student questionnaire collected information on the country of birth of the students and their parents. The index of *immigrant background* (IMMIG) that was created from these data had three categories:

- (1) Students without immigrant background (students born in the country of assessment or who had at least one parent born in the country);<sup>1</sup>
- (2) Students born in the country of assessment but whose parent(s) were born in another country;
- (3) Students born outside the country of assessment and whose parent(s) were born in another country.



<sup>1</sup> Students who were born abroad but had at least one parent born in the country of the test were also classified as students without immigrant background.

Missing values were assigned to students with missing responses for either their own place of birth, or that of their mother and father, or for all three questions. Some analyses involved use of a dichotomous indicator variable that distinguished between students with (Categories 2 and 3) and without immigrant background (Category 1).

The ICCS student questionnaire contained a question that asked students if the language spoken at home most of the time was the language of assessment or another language.<sup>2</sup> This information was used to derive an index on home language (TESTLANG), in which responses were grouped into two categories:

- (0) The language spoken at home most of the time differed from the language of assessment;
- (1) The language spoken at home most of the time was the language of assessment.

The ICCS student questionnaire included a question that asked students to indicate their *parents' level of interest in social and political issues*. The (recoded) indices for mother's (MINT) and father's interest (FINT) consisted of the following categories:

- (0) Not interested at all;
- (1) Not very interested;
- (2) Quite interested;
- (3) Very interested.

An index of the highest level of parental interest in political and social issues (PARINT) was created by computing the maximum value of both MINT and FINT. Some analyses involved use of a dichotomous index of parental interest. Here, the two categories were "not interested at all or not very interested" (0) and "quite interested or very interested" (1).

Data on students' *home literacy resources* were derived from a question that asked students how many books they had in their homes. The (recoded) index on home literacy (HOMELIT) had the following categories:

- (0) 0 to 10 books;
- (1) 11 to 25 books;
- (2) 26 to 100 books;
- (3) 101 to 200 books;
- (4) 201 to 500 books;
- (5) More than 500 books.

For some analyses, an index variable that reflected the approximate number of books at home was created by using midpoints of the range of books in each category (5, 18, 63, 151, 351, and 700). During the regression analyses, this variable was scaled to units of 100 books by dividing the new values by 100.

The ICCS student questionnaire included two questions that asked students if they liked a particular political party more than others and, if they did, how much they favored this party. The data from these two questions were combined to derive an index of *students' support for political parties* (SUPPART) that had four categories:

- (0) Does not like any political party more than others;
- (1) A little in favor;
- (2) To some extent in favor;
- (3) A lot in favor.



---

2 Some countries collected more detailed information on language use. This information is included in the database.

### *Teacher questionnaire*

Individual *teacher age* (TAGE) was calculated as the midpoint of the age ranges given in Question 7 of the teacher questionnaire. “Fewer than 25” was assigned a value of 23 and “60 or over” was coded as 63.

### *School questionnaire*

Individual *school principal age* (CAGE) was calculated as the midpoint of the age ranges given in Question 2 of the school questionnaire. “Fewer than 30” was assigned a value of 27 and “60 or over” was given a value of 63.

A simple addition of the total number of boys enrolled at school (IC2G20A) and the total number of girls enrolled at school (IC2G20B) was used to calculate school size (SCHLSIZE).

The same procedure was used to calculate school enrollment in the target grade (GRENROL). The number of boys (IC2G21A) and the number of girls (IC221B) at each target grade were added together.

To calculate the overall student–teacher ratio (CSTRATIO), the number of students at the school (SCHLSIZE) was divided by the number of teachers at the school (IC2G22B). The student–teacher ratio at the target-grade level (GSTRATIO) was derived by dividing the number of students enrolled in the target grade (GRENROL) by the number of teachers teaching classes at the target grade. The values for both indicators reflected the number of students per teacher; the lower values indicated better resourcing of the school.

To calculate the percentage of target-grade students at each school (TGPERC), the number of teachers instructing at the target-grade level (IC2G22A) was divided by the total number of teachers at the school (IC2G22B). This value was then multiplied by 100.

### *European regional questionnaire*

Two questions in the European student questionnaire were designed to capture students’ confidence in communicating in at least one other European language. The first of these questions asked students to give a simple “yes” or “no” as to whether they were able to communicate in, or understand, any languages spoken in European countries other than their own. Those students who answered yes were then asked to state how well they could communicate in these languages on a three-point scale of “not very well,” “well,” and “very well.” The two questions were combined to form an index of *students’ self-reported proficiency in another European language* (EURPLANG). The index had four categories, the values of which ranged from 0 to 4.

## **Scaling procedures**

### *Classic scaling analysis*

In this chapter, we report reliabilities both overall and for national samples and use Cronbach’s alpha coefficient as an estimate of the internal consistency of each scale.

### *Confirmatory factor analysis*

Structural equation modeling (SEM) (Kaplan, 2000) was used to confirm theoretically expected dimensions and, at the field-trial stage, to re-specify the dimensional structure. When using *confirmatory factor analysis*, it is necessary to acknowledge the need to employ a theoretical model of item dimensionality that can be tested via the collected data. Within the SEM framework,



latent variables link to observable variables via measurement equations. An observed variable  $x$  is thus modeled as

$$(1) \quad x = \Lambda_y \xi + \delta,$$

where  $\Lambda_y$  is a  $q \times k$  matrix of factor loadings,  $\xi$  denotes the latent variable(s), and  $\delta$  is a  $q \times 1$  vector of unique error variables. The expected covariance matrix is fitted according to the theoretical factor structure.

During the confirmatory factor analyses, selected model-fit indices were also used to measure the extent to which a model with an assumed a-priori structure “fitted the data.” For the ICCS analysis, model fit was assessed primarily through use of the *root-mean square error of approximation* (RMSEA), the *comparative fit index* (CFI), and the *non-normed fit index* (NNFI), all of which are less affected than other indices by sample size and model complexity (see Bollen & Long, 1993).

It was assumed, with respect to the analysis, that RMSEA values over 0.10 would suggest an unacceptable model fit while values below 0.05 would indicate a close model fit. As additional fit indices, CFI and NNFI are bound between 0 and 1. Values below 0.90 and 0.95 indicate a non-satisfactory model fit whereas values greater than 0.95 suggest a close model fit.

In addition to these fit indices, standardized factor loadings and residual variance were used to assess model structures for questionnaire data. Standardized factor loadings  $\lambda'$  can be interpreted in the same way as standardized regression coefficients if the indicator variable is regressed on the latent factor. The loadings also reflect the extent to which each indicator measures the underlying construct. Squared standardized factor loadings indicate how much variance in an indicator variable can be explained by the latent factor and are related to the (standardized) residual variance estimate  $\delta'$  (these provide an estimate of the unexplained proportion of variance) as

$$\delta' = (1 - \lambda'^2).$$

Multidimensional models were used to assess the estimated correlation(s) between latent factors and to review the similarity of the different dimensions measured by the item sets.

Generally, maximum likelihood estimation and covariance matrices are not appropriate for analyses of (categorical) questionnaire items because the approach treats items as if they are continuous. Weighted least squares estimation with polychoric correlations (see Jöreskog, 1990, 1994) were therefore used to estimate the confirmatory factor models. The software package that was used to do this was *LISREL 8.72* (Jöreskog & Sörbom, 2004).

A decision was made to use confirmatory factor analyses for sets of conceptually related questionnaire items that measured between one and four different factors. This approach made it possible to describe both the extent to which items measured underlying latent traits as well as the associations between latent factors. The analyses employed data from the (pooled) ICCS calibration samples of students, teachers, and schools, a process that ensured equal representations of countries in the analyses.

#### *Item response modeling*

Item response modeling was typically used to scale questionnaire items. The one-parameter (Rasch) model (Rasch, 1960) for dichotomous items models the probability of selecting Category 1 instead of 0 as

$$P_i(\theta) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}, \quad (1)$$

where  $P_i(\theta)$  is the probability of person  $n$  scoring 1 on item  $i$ ,  $\theta_n$  is the estimated latent trait of person  $n$ , and  $\delta_i$  is the estimated location of item  $i$  on this dimension. For each item, item responses are modeled as a function of the latent trait  $\theta_n$ .



In the case of items with more than two ( $k$ ) categories (as, for example, with Likert-type items), this model can be generalized to the *partial credit model* (Masters & Wright, 1997),<sup>3</sup> which takes the form of

$$P_{x_i}(\theta) = \frac{\exp \sum_{k=0}^x (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^h (\theta_n - \delta_i + \tau_{ij})} \quad x_i = 0, 1, \dots, m_i, \quad (2)$$

where  $P_{x_i}(\theta)$  denotes the probability of person  $n$  scoring  $x$  on item  $i$ ,  $\theta_n$  denotes the person's latent trait, the item parameter  $\delta_i$  gives the location of the item on the latent continuum, and  $\tau_{ij}$  denotes an additional step parameter.

The weighted mean-square statistic (*infit*), which is a residual-based fit statistic, was used to assess item fit. Weighted *infit* statistics were reviewed for both item and step parameters, and ACER Conquest software (Wu, Adams, Wilson, & Haldane, 2007) was used to estimate item parameters and to analyze item fit.

The international item parameters that were obtained came from the following calibrations.

- *Calibration of student item parameters*: subsamples of 500 students randomly selected from each (weighted) national database for the 36 countries that met sample participation requirements. The final calibration sample included data from 18,000 students.
- *Calibration of teacher item parameters*: subsamples of 250 teachers randomly selected from each (weighted) national database for the 27 countries that met sample participation requirements. The final calibration sample included data from 6,750 teachers.
- *Calibration of school item parameters*: national school samples weighted to have the same weight (set to values of 100 regardless of sample size) for each country that met sample participation requirements. The final calibration sample included data from all school principals.

After the international item parameter from the calibration sample had been estimated, weighted likelihood estimation was used to obtain individual student scores. Weighted likelihood estimations can be computed by minimizing the equation

$$\sum_{i \in \Omega} \left[ \left( r_x + \frac{J_n}{2I_n} \right) - \sum_{j=1}^k \frac{\exp(\sum_{j=0}^x \theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^h (\theta_n - \delta_i + \tau_{ij})} \right] = 0 \quad (3)$$

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  $\theta$ ) is used as a weight function to account for the bias inherent in maximum likelihood estimation (see Warm, 1989). ACER ConQuest software made it possible to pre-calibrate item parameters in order to derive scale scores.

The weighted likelihood estimates were transformed to an international metric with an ICCS average of 50 and a standard deviation of 10 for equally weighted datasets from the 36 countries that met sample participation requirements. The following formula was applied in order to achieve the transformation:

$$\theta'_n = 50 + 10 \frac{\theta_n - \bar{\theta}_{ICCS}}{\sigma_{\theta(ICCS)}},$$

3 An alternative is the rating scale model (RSM), which has the same step parameters for all items in a scale (see Andersen, 1997).



where  $\theta'_n$  are the scores in the international metric,  $\theta_n$  are the original weighted likelihood estimates in logits, and  $\bar{\theta}_{ICCS}$  is the international mean of logit scores with equally weighted country subsamples.  $\sigma_{\theta(ICCS)}$  is the corresponding international standard deviation of the original weighted likelihood estimates. Table D.4 in Appendix D presents the means and standard deviations used to transform the original scale scores for the international student, teacher, school, and regional (Asia, Europe, Latin America) questionnaires into the international metric.

#### *Describing questionnaire scale indices*

For the questionnaire scales, the weighted likelihood estimates (logits) for the latent dimensions were transformed to scales with an ICCS average of 50 and a standard deviation of 10 (with equally weighted samples). While these scores could be interpreted by comparing individual scores or group average scores with the ICCS average, the individual scores do not reveal anything about the actual item responses. Also, it would have been impossible to determine from the scale score values the extent to which respondents endorsed the items used to measure the latent variable. The scaling model that was used to derive individual scores made it possible to develop descriptions of these scales because scale scores could be mapped to (expected) item responses.<sup>4</sup>

It is possible to describe item characteristics by using the parameters of the partial credit model to provide an estimate for each category of its probability of being chosen as a minimum relative to all other categories. This process is equivalent to computing the odds of scoring higher than a particular category.

Figure 12.1 presents the results of plotting these cumulative probabilities against scale scores for a fictitious item. The three vertical lines denote those points on the latent continuum where it becomes more likely to score  $> 0$ ,  $> 1$ , or  $> 2$ . These locations  $\Gamma_k$  are *Thurstonian thresholds* that can be obtained through an iterative procedure that calculates summed probabilities for each category at each (decimal) point on the latent variable.

Summed probabilities are not identical to expected item scores and have to be understood in terms of the probability of scoring at least a particular category. Other ways of describing item characteristics based on the partial credit model are *item characteristic curves*, which involve plotting the individual category probabilities and the *expected item score curves* (for a detailed description, see Masters & Wright, 1997).

Thurstonian thresholds can be used to indicate for each item category those points on a scale at which respondents have a 0.5 probability of scoring this category or higher. For example, in the case of Likert-type items with the categories *strongly disagree* (SD), *disagree* (D), *agree* (A), and *strongly agree* (SA), we can determine at what point of a scale a respondent has a 50 percent likelihood of agreeing with the item.

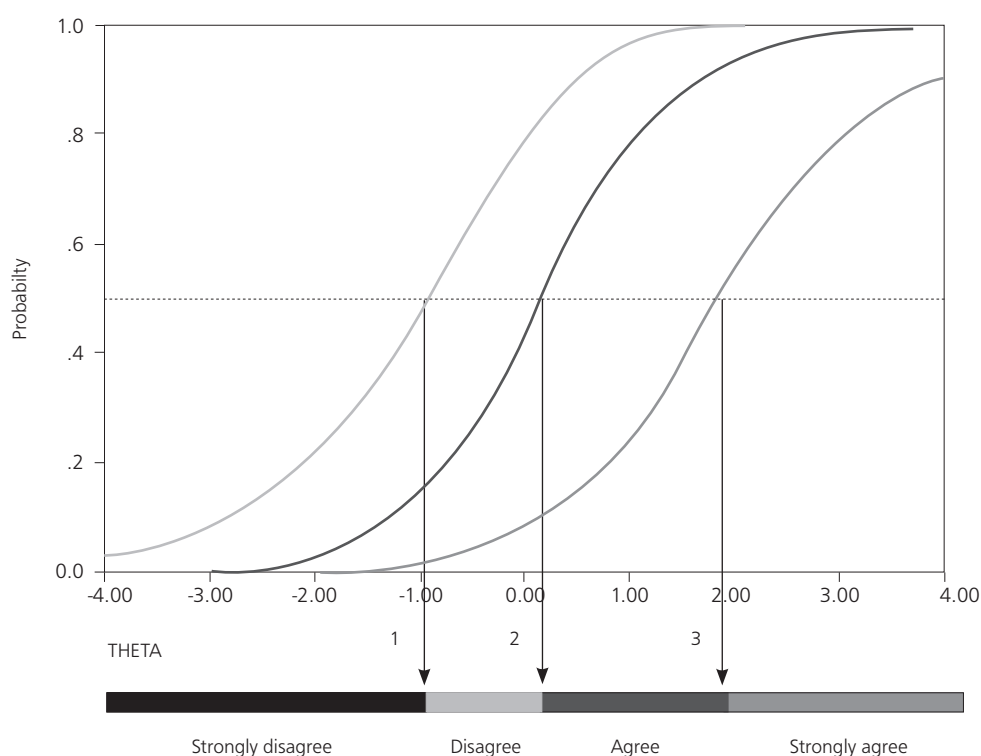
The item-by-score maps included in ICCS reports predict the minimum coded score (e.g., 0 = “strongly disagree,” 1 = “disagree,” 2 = “agree,” and 3 = “strongly agree”) a respondent would obtain on a Likert-type item. For example, we could predict that students with a certain scale score would have a 50 percent probability of agreeing (or strongly agreeing) with a particular item (see the example item-by-score map in Figure 12.2). For each item, it is thus possible to determine Thurstonian thresholds, the points at which a minimum item score becomes more likely than any lower score to occur and which determine the boundaries between item categories on the item-by-score map.



4 A similar approach was used in the IEA CIVED survey (see Schulz, 2004).



Figure 12.1: Summed category probabilities for fictitious item



This information can also be summarized by calculating the average thresholds across all items in a scale. This was usually done for the second threshold of the four-point Likert-type scales, a process that allows one to predict how likely it would be for a respondent with a certain scale score to have (on average across items) responses in the two lower or upper categories. Use of this approach in the case of items measuring agreement made it possible to distinguish between scale scores for respondents who were most likely to agree or disagree with the average item used for scaling.

National average scale scores were depicted as boxes that indicated their mean values plus/minus sampling error and that were set in graphical displays featuring two underlying colors. National average scores located in the area set in (say) light blue on average across items would indicate that student responses had resided in the lower item categories (“disagree or strongly disagree,” “not at all or not very interested,” “never or rarely”). If these scores were found in the darker blue area, however, then students’ average item responses would have been in the upper item response categories (“agree or strongly agree,” “quite or very interested,” “sometimes or often”).

## Scaled indices

### *International student questionnaire*

#### **Students’ behaviors**

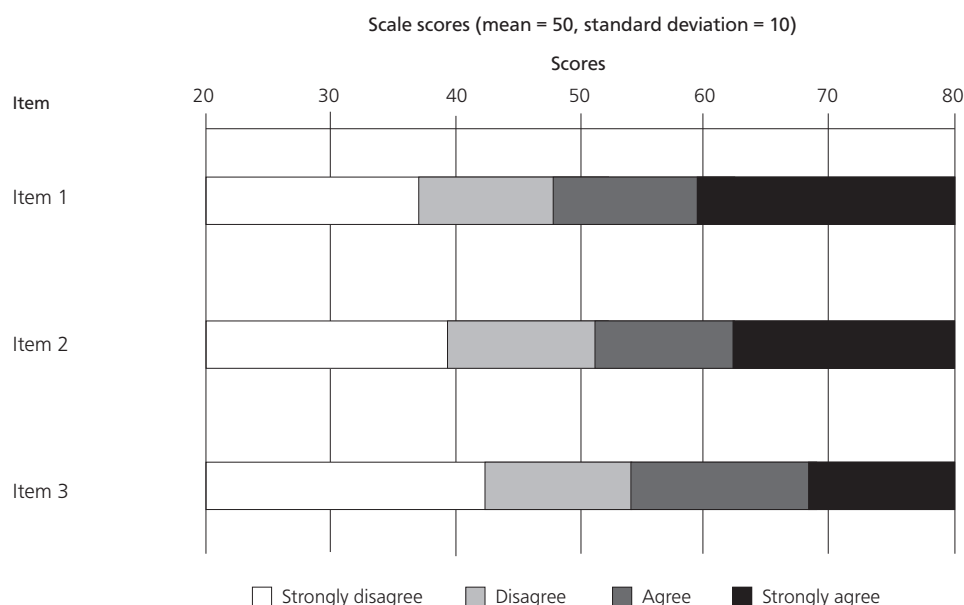
Three scales were derived from questions regarding student behavior outside of school. These scales, the reliabilities for which are reported in Table 12.1, are included in the ICCS student database. The scales are:

- Students’ discussion of political and social issues outside of school (POLDISC);
- Students’ civic participation in the wider community (PARTCOM);
- Students’ civic participation at school (PARTSCHL).





Figure 12.2: Example of questionnaire item-by-score map



#### Example of how to interpret the item-by-score map

- 1: A respondent with score 30 has more than a 50% probability of strongly disagreeing with all three items
- 2: A respondent with score 40 has more than a 50% probability of *not* strongly disagreeing with Items 1 and 2 but of strongly disagreeing with Item 3
- 3: A respondent with score 50 has more than a 50% probability of agreeing with Items 1 and of disagreeing with Items 2 and 3
- 4: A respondent with score 60 has more than a 50% probability of strongly agreeing with Items 1 and of at least agreeing with Items 2 and 3
- 5: A respondent with score 60 has more than a 50% probability of strongly agreeing with Items 1, 2, and 3

Question 13 of the ICCS student questionnaire asked students how often they took part in different activities outside of school. Four items asked them about their participation in discussion with friends and parents about political or social issues and events in other countries. Response categories were “never or hardly ever,” “monthly,” “weekly,” and “daily or almost daily.”

The resulting scale (POLDISC) had a reliability (Cronbach’s alpha) of 0.72 for the pooled ICCS sample. Across the participating countries, scale reliabilities ranged from 0.61 to 0.81 (see Table 12.1). Table 12.2 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect more frequent participation in discussions about political and social issues outside of school.

Question 14 asked students to state whether they had participated in eight different organizations, clubs, or groups in the wider community either “within the last 12 months,” “more than a year ago,” or “never.” Seven of these items were used to derive a scale reflecting *students’ civic participation in the wider community* (PARTCOM); the positive values on this scale reflect higher levels of civic participation. The scale reliability (Cronbach’s alpha) was 0.74 for the international sample. Scale reliabilities across countries ranged from 0.60 to 0.80 (see Table 12.1). Table 12.2 shows the item parameters that were used for scaling.



*Table 12.1: Reliabilities for scales reflecting students' behaviors*

Country	Political Discussion	Community Participation	School Participation
Austria	0.72	0.71	0.61
Belgium (Flemish)	0.71	0.65	0.70
Bulgaria	0.65	0.72	0.70
Chile	0.66	0.71	0.62
Chinese Taipei	0.74	0.68	0.66
Colombia	0.67	0.71	0.63
Cyprus	0.69	0.78	0.75
Czech Republic	0.71	0.67	0.64
Denmark	0.79	0.66	0.68
Dominican Republic	0.61	0.69	0.64
England	0.74	0.73	0.71
Estonia	0.75	0.68	0.68
Finland	0.81	0.62	0.66
Greece	0.66	0.72	0.63
Guatemala	0.68	0.71	0.61
Hong Kong SAR	0.79	0.74	0.72
Indonesia	0.64	0.66	0.56
Ireland	0.71	0.66	0.61
Italy	0.67	0.67	0.53
Korea, Republic of	0.76	0.76	0.75
Latvia	0.70	0.69	0.70
Liechtenstein	0.71	0.65	0.66
Lithuania	0.69	0.69	0.67
Luxembourg	0.69	0.76	0.64
Malta	0.63	0.70	0.00
Mexico	0.64	0.73	0.66
Netherlands	0.71	0.60	0.68
New Zealand	0.75	0.71	0.72
Norway	0.80	0.75	0.71
Paraguay	0.64	0.68	0.59
Poland	0.74	0.71	0.65
Russian Federation	0.74	0.73	0.69
Slovak Republic	0.71	0.67	0.62
Slovenia	0.69	0.73	0.70
Spain	0.68	0.74	0.62
Sweden	0.81	0.80	0.69
Switzerland	0.76	0.67	0.64
Thailand	0.76	0.67	0.64
<b>ICCS average</b>	<b>0.72</b>	<b>0.74</b>	<b>0.66</b>



Table 12.2: Item parameters for scales reflecting students' behaviors

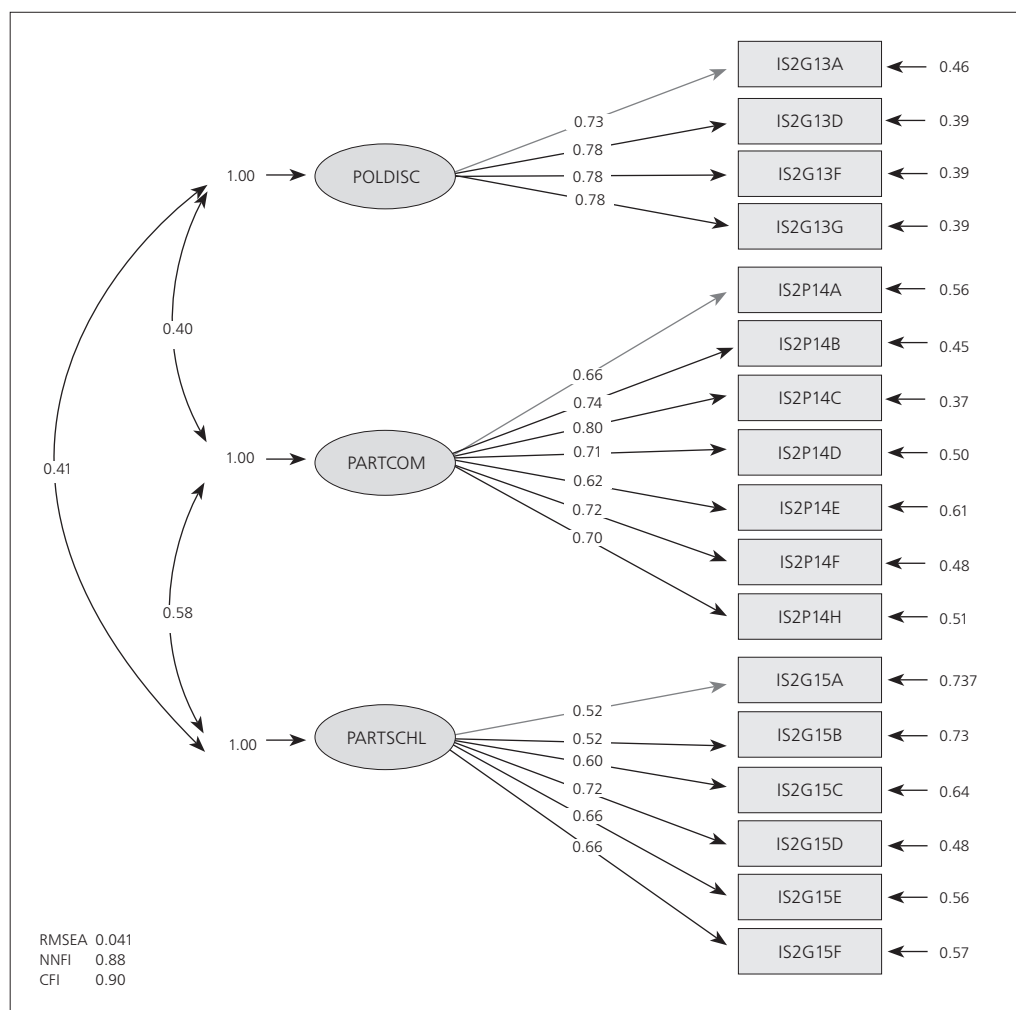
Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Political Discussion</b>	<b>How often are you involved in each of the following activities outside of school?</b>				
IS2G13A	Talking with your parent(s) about political or social issues	1.14	-0.70	-0.10	0.81
IS2G13D	Talking with friends about political and social issues	1.70	-0.69	-0.12	0.81
IS2G13F	Talking with your parent(s) about what is happening in other countries	0.46	-1.25	-0.05	1.31
IS2G13G	Talking with friends about what is happening in other countries	1.15	-1.15	-0.04	1.1
<b>Community Participation</b>	<b>Have you ever been involved in activities of any of the following organisations, clubs, or groups?</b>				
IS2P14A	Youth organisation affiliated with a political party or union	2.30	0.99	-0.99	
IS2P14B	Environmental organisation	1.57	-0.08	0.08	
IS2P14C	Human rights organisation	2.12	0.43	-0.43	
IS2P14D	A voluntary group doing something to help the community	1.28	-0.01	0.01	
IS2P14E	An organisation collecting money for a social cause	1.04	-0.09	0.09	
IS2P14F	A cultural organisation based on ethnicity	2.20	0.46	-0.46	
IS2P14H	A group of young people campaigning for an issue	1.43	0.19	-0.19	
<b>School Participation</b>	<b>At school, have you ever done any of the following activities?</b>				
IS2G15A	Voluntary participation in school-based music or drama activities outside of regular lessons	0.25	-0.27	0.27	
IS2G15B	Active participation in a debate	0.65	0.27	-0.27	
IS2G15C	Voting for <class representative> or <school parliament>	-0.49	0.15	-0.15	
IS2G15D	Taking part in decision-making about how the school is run	0.79	0.25	-0.25	
IS2G15E	Taking part in discussions at a <student assembly>	0.69	0.32	-0.32	
IS2G15F	Becoming a candidate for <class representative> or <school parliament>	0.70	0.29	-0.29	

Question 15 asked students if they had participated in six different civic-related activities at school either “within the last twelve months,” “more than a year ago,” or “never.” These items permitted derivation of a scale reflecting *students' civic participation at school* (PARTSCHL), with the positive values reflecting higher levels of civic participation. The scale reliability was 0.66 for the international sample, and the cross-national reliabilities ranged from 0.53 to 0.75. Table 12.2 shows the item parameters used for scaling.

Figure 12.3 shows the results of the confirmatory factor analysis of these items. The RMSEA indicated a close model fit whereas NNFI and CFI suggested some lack of fit. The CFA showed moderate positive correlations among the three latent traits, in particular between community- and school-based student participation.



Figure 12.3: Confirmatory factor analysis of items measuring students' behaviors



### Students' perceptions of the school context

Table 12.3 reports the four scales (and their reliabilities) derived from questions regarding students' perceptions. The scales, which are included in the ICCS student database, are:

- Students' perceptions of openness in classroom discussions (OPDISC);
- Students' perceptions of influence on decisions about school (STUDINF);
- Students' perceptions of student–teacher relations at school (STUTREL);
- Students' perceptions of the value of participation at school (VALPARTS).

Question 16 asked students how frequently (“never,” “rarely,” “sometimes,” “often”) they thought political and social issues were discussed during regular lessons. Six of the question items were used to derive the scale reflecting *students' perceptions of openness in classroom discussions* (OPDISC). The higher values on the scale reflect perceptions of higher levels of classroom discussions of political and social issues. The scale's reliability (Cronbach's alpha) was 0.76 for the pooled ICCS sample. The country reliabilities ranged from 0.65 to 0.84 (see Table 12.3). Table 12.4 shows the item parameters that were used for scaling.

Question 17 asked students to report the extent to which they thought their opinion was taken into account when decisions were being made about their school. The response options were “not at all,” “to a small extent,” “to a moderate extent,” and “to a large extent.”



*Table 12.3: Reliabilities for scales reflecting students' perceptions of the school context*

Country	Open Classroom	Student Influence	Student-Teacher Relations	Value of Participation
Austria	0.76	0.83	0.80	0.66
Belgium (Flemish)	0.74	0.88	0.78	0.69
Bulgaria	0.75	0.88	0.72	0.72
Chile	0.76	0.85	0.76	0.73
Chinese Taipei	0.80	0.86	0.85	0.81
Colombia	0.65	0.75	0.74	0.64
Cyprus	0.75	0.84	0.80	0.76
Czech Republic	0.71	0.86	0.78	0.70
Denmark	0.75	0.78	0.80	0.74
Dominican Republic	0.68	0.72	0.70	0.64
England	0.81	0.88	0.80	0.79
Estonia	0.74	0.87	0.75	0.73
Finland	0.74	0.87	0.80	0.81
Greece	0.69	0.85	0.77	0.70
Guatemala	0.70	0.81	0.66	0.68
Hong Kong SAR	0.84	0.87	0.84	0.77
Indonesia	0.66	0.87	0.53	0.54
Ireland	0.79	0.89	0.79	0.76
Italy	0.66	0.81	0.75	0.65
Korea, Republic of	0.81	0.90	0.79	0.81
Latvia	0.70	0.87	0.76	0.71
Liechtenstein	0.79	0.80	0.79	0.75
Lithuania	0.73	0.86	0.72	0.68
Luxembourg	0.75	0.87	0.80	0.73
Malta	0.75	0.85	0.79	0.73
Mexico	0.71	0.78	0.74	0.69
Netherlands	0.73	0.84	0.75	0.72
New Zealand	0.80	0.87	0.78	0.77
Norway	0.80	0.85	0.84	0.77
Paraguay	0.65	0.78	0.66	0.59
Poland	0.76	0.88	0.78	0.79
Russian Federation	0.75	0.86	0.75	0.71
Slovak Republic	0.68	0.86	0.79	0.64
Slovenia	0.77	0.86	0.76	0.75
Spain	0.72	0.88	0.75	0.72
Sweden	0.81	0.86	0.83	0.81
Switzerland	0.79	0.86	0.79	0.73
Thailand	0.79	0.86	0.79	0.73
ICCS average	0.76	0.88	0.78	0.73



Table 12.4: Item parameters for scales reflecting students' perceptions of the school context

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Open Classrooms</b>	<b>When discussing political and social issues during regular lessons, how often do the following things happen?</b>				
IS2G16B	Teachers encourage students to make up their own minds	-0.90	-0.77	-0.21	0.99
IS2G16C	Teachers encourage students to express their opinions	-1.21	-0.71	-0.13	0.84
IS2G16D	Students bring up current political events for discussion in class	0.38	-1.34	0.03	1.30
IS2G16E	Students express opinions in class even when their opinions are different from most of the other students	-0.72	-1.22	-0.10	1.33
IS2G16F	Teachers encourage students to discuss the issues with people having different opinions	-0.08	-1.06	-0.15	1.22
IS2G16G	Teachers present several sides of the issues when explaining them in class	-0.63	-0.96	-0.19	1.16
<b>Student Influence</b>	<b>In this school, how much is your opinion taken into account when decisions are made about the following issues?</b>				
IS2G17A	The way classes are taught	-0.16	-1.65	-0.19	1.85
IS2G17B	What is taught in classes	-0.01	-1.30	-0.25	1.55
IS2G17C	Teaching and learning materials	0.13	-1.35	-0.19	1.53
IS2G17D	The timetable	0.39	-0.59	-0.33	0.92
IS2G17E	Classroom rules	-0.47	-1.27	-0.19	1.47
IS2G17F	School rules	0.04	-0.67	-0.15	0.81
<b>Student-Teacher Relations</b>	<b>How much do you agree or disagree with the following statements about you and your school?</b>				
IS2G18A	Most of my teachers treat me fairly	-1.32	-1.64	-0.79	2.43
IS2G18B	Students get along well with most teachers	-0.87	-2.50	-0.33	2.84
IS2G18C	Most teachers are interested in students' wellbeing	-1.17	-1.93	-0.58	2.50
IS2G18E	Most of my teachers really listen to what I have to say	-0.95	-2.07	-0.62	2.68
IS2G18F	If I need extra help, I will receive it from my teachers	-1.29	-1.64	-0.84	2.49
<b>Value Participation</b>	<b>How much do you agree or disagree with the following statements about student participation at school?</b>				
IS2P19A	Student participation in how schools are run can make schools better	-1.50	-1.57	-0.81	2.37
IS2P19B	Lots of positive changes can happen in schools when students work together	-1.90	-1.35	-1.06	2.42
IS2P19C	Organising groups of students to express their opinions could help solve problems in schools	-1.52	-1.76	-0.80	2.55
IS2P19D	All schools should have a <school parliament>	-1.65	-1.26	-0.51	1.78
IS2P19E	Students can have more influence on what happens in schools if they act together rather than alone	-1.66	-1.24	-0.90	2.14

Six of the question items were used to form the scale reflecting *students' perceptions of influence on decisions about school* (STUDINF). Higher values on the scale correspond to greater perceived influence on decisions. The scale reliability (Cronbach's alpha) was 0.88 for the pooled ICCS sample, and the country reliabilities ranged from 0.72 to 0.90 (see Table 12.3). The item parameters that were used for scaling are shown in Table 12.4.

Question 18 contained items assessing the degree to which students agreed or disagreed with statements about relationships in their school. Response options ranged from "strongly agree" to "strongly disagree." Five of the seven items were used to derive the scale *students' perceptions of student-teacher relations at school* (STUTREL), which had a reliability (Cronbach's alpha) of 0.78 for the pooled ICCS sample. Country reliabilities ranged from 0.53 to 0.85 (see Table 12.3). Table 12.4 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect perceptions of strong relations between students and teachers at school.

In Question 19, students were asked to indicate their degree of agreement (range "strongly agree" to "strongly disagree") with statements about the value of participating in certain events at school. All five question items were included in the scale *students' perceptions of the value of participation at school* (VALPARTS). Higher scores on this scale correspond to a higher extent of agreement with statements about the value of participation at school. The reliability of this scale (Cronbach's alpha) was 0.73 for the pooled ICCS sample. Reliabilities across countries ranged from 0.54 to 0.81 (see Table 12.3). Table 12.4 shows the item parameters that were used for scaling.

Figure 12.4 presents the results for a model with a four-factor solution for all items related to school context. The RMSEA of 0.041 indicated a close model fit, and the NNFI and CFI also indicated satisfactory model fit. Inspection of the item factor loadings indicated that, for each of the factors, the items provided a good measurement of the respective underlying latent trait. The results also showed positive correlations between the four latent traits. Somewhat higher correlations were found between STUTREL (student-teacher relations) and STUDINF (perceptions of student influence) as well as between STUTREL and VALPARTS (value of student participation).

### **Students' democratic value beliefs**

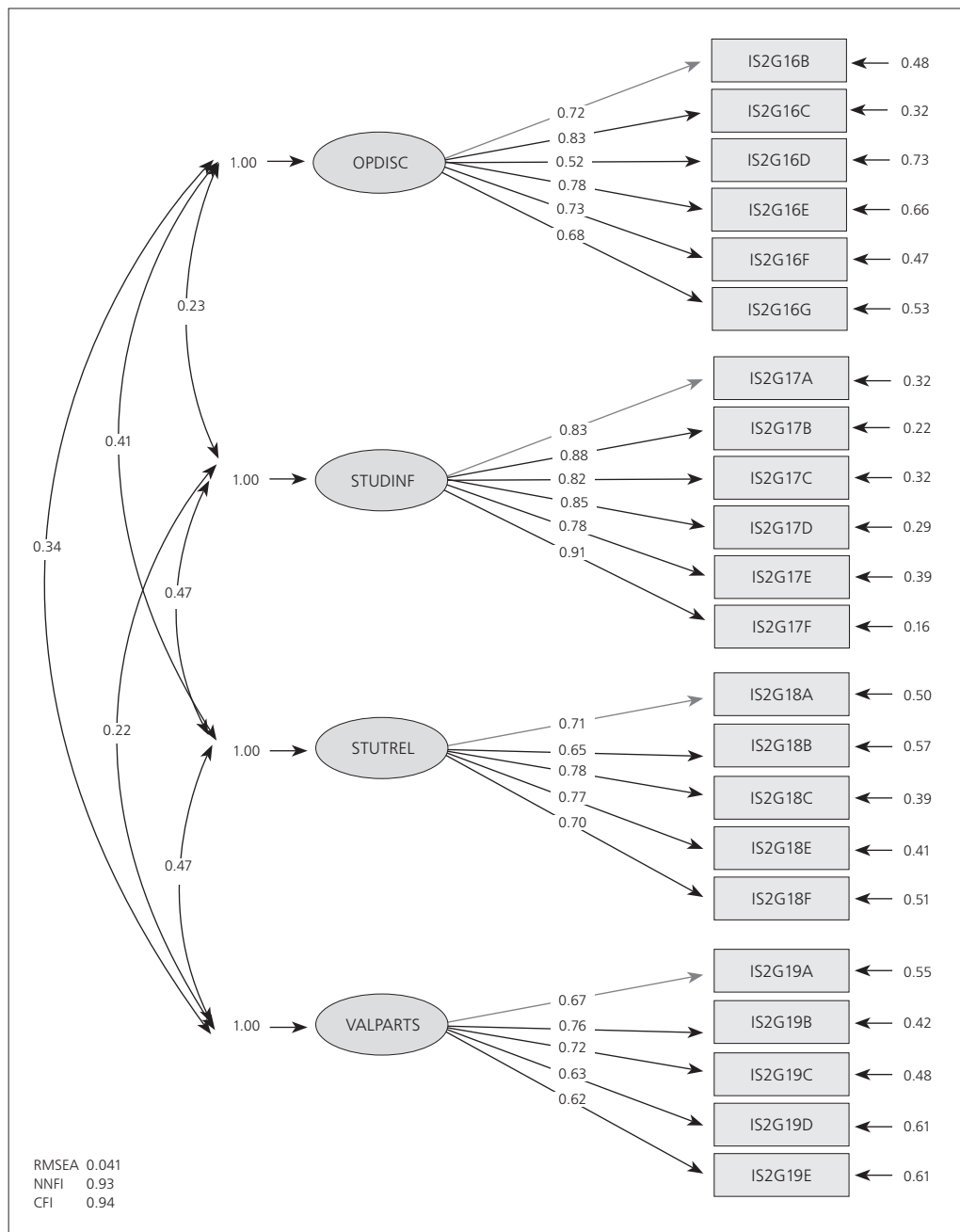
The ICCS student questionnaire included a set of items measuring students' beliefs about democratic values. Five items from Question 20 were used to derive the scale *students' support for democratic values* (DEMVAL). The question required students to state their level of agreement ("strongly agree" to "strongly disagree") with statements about what a society should be like; the higher values on the scale correspond to greater support for democratic values. The scale's reliability (Cronbach's alpha) was 0.65 for the pooled ICCS sample, and the country reliabilities ranged from 0.56 to 0.78 (see Table 12.5). Table 12.6 shows the item parameters that were used for scaling.

The results of the confirmatory factor analysis of these items (see Figure 12.5) showed satisfactory model fit for the one-factor solution. Item IS2P20F ("people should always be free to criticize the government publicly") had a somewhat lower factor loading than the other items in this scale.





Figure 12.4: Confirmatory factor analysis of items measuring students' perceptions of the school context



*Table 12.5: Item parameters for scales reflecting students' support for democratic value beliefs*

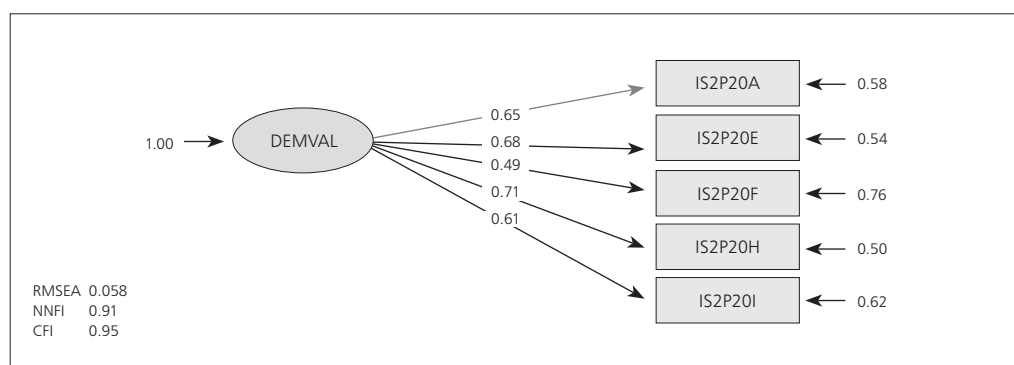
Country	Democratic Values
Austria	0.68
Belgium (Flemish)	0.62
Bulgaria	0.66
Chile	0.62
Chinese Taipei	0.67
Colombia	0.57
Cyprus	0.69
Czech Republic	0.65
Denmark	0.70
Dominican Republic	0.56
England	0.73
Estonia	0.61
Finland	0.68
Greece	0.70
Guatemala	0.56
Hong Kong SAR	0.72
Indonesia	0.62
Ireland	0.67
Italy	0.60
Korea, Republic of	0.78
Latvia	0.62
Liechtenstein	0.60
Lithuania	0.65
Luxembourg	0.67
Malta	0.57
Mexico	0.64
Netherlands	0.66
New Zealand	0.72
Norway	0.72
Paraguay	0.58
Poland	0.71
Russian Federation	0.63
Slovak Republic	0.63
Slovenia	0.59
Spain	0.63
Sweden	0.75
Switzerland	0.62
Thailand	0.62
<b>ICCS average</b>	<b>0.65</b>



Table 12.6: Item parameters for scale reflecting students' democratic value beliefs

	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
Democratic Values	There are different views about what a society should be like. We are interested in your views on this. How much do you agree or disagree with the following statements?				
IS2P20A	Everyone should always have the right to express their opinions freely	-2.64	-0.13	-1.34	1.48
IS2P20E	All people should have their social and political rights respected	-2.06	-0.53	-1.06	1.59
IS2P20F	People should always be free to criticize the government publicly	-1.22	-1.70	-0.21	1.90
IS2P20H	All citizens should have the right to elect their leaders freely	-2.12	-0.76	-0.84	1.61
IS2P20I	People should be able to protest if they believe a law is unfair	-1.78	-1.16	-0.77	1.93

Figure 12.5: Confirmatory factor analysis of items measuring students' democratic values



### Students' perceptions of good citizenship

Question 21 of the ICCS student questionnaire contained items relating to being a good adult citizen. Students were asked to rate the importance ("very important," "quite important," "not very important," "not important at all") of a series of possible citizenship behaviors. The two scales that were derived from this question, and which are included in the student database, are:

- Students' perceptions of the importance of conventional citizenship (CITCON);
- Students' perceptions of the importance of social-movement-related citizenship (CITSOC).

The first six of these items were used to construct the scale *students' perceptions of the importance of conventional citizenship* (CITCON). The reliability of this scale (Cronbach's alpha) was 0.71 for the pooled ICCS sample; the cross-national reliabilities ranged from 0.54 to 0.77 (see Table 12.7). Table 12.8 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale denote stronger degrees of importance placed on conventional citizenship behaviors.

Another four items in this question were used to derive the second scale—*students' perceptions of the importance of social-movement-related citizenship* (CITSOC). Higher values on this scale correspond to greater perceived importance of social-movement-related citizenship. The scale had a reliability (Cronbach's alpha) of 0.74 for the pooled ICCS sample, while the country reliabilities ranged from 0.51 to 0.81 (see Table 12.7). The item parameters that were used for scaling are shown in Table 12.8.



Table 12.7: Reliabilities for scales reflecting students' perceptions of the importance of citizenship behaviors

Country	Conventional Citizenship	Soc.-Mov.-Citizenship
Austria	0.71	0.72
Belgium (Flemish)	0.69	0.76
Bulgaria	0.69	0.72
Chile	0.70	0.71
Chinese Taipei	0.76	0.76
Colombia	0.61	0.58
Cyprus	0.72	0.77
Czech Republic	0.71	0.74
Denmark	0.67	0.78
Dominican Republic	0.54	0.51
England	0.74	0.79
Estonia	0.68	0.69
Finland	0.75	0.77
Greece	0.63	0.71
Guatemala	0.59	0.57
Hong Kong SAR	0.77	0.75
Indonesia	0.59	0.61
Ireland	0.71	0.77
Italy	0.64	0.71
Korea, Republic of	0.74	0.81
Latvia	0.63	0.64
Liechtenstein	0.69	0.73
Lithuania	0.66	0.71
Luxembourg	0.73	0.75
Malta	0.69	0.71
Mexico	0.66	0.66
Netherlands	0.70	0.70
New Zealand	0.74	0.77
Norway	0.75	0.76
Paraguay	0.60	0.54
Poland	0.73	0.73
Russian Federation	0.73	0.72
Slovak Republic	0.67	0.70
Slovenia	0.72	0.71
Spain	0.70	0.75
Sweden	0.77	0.81
Switzerland	0.70	0.74
Thailand	0.70	0.74
ICCS average	0.71	0.74

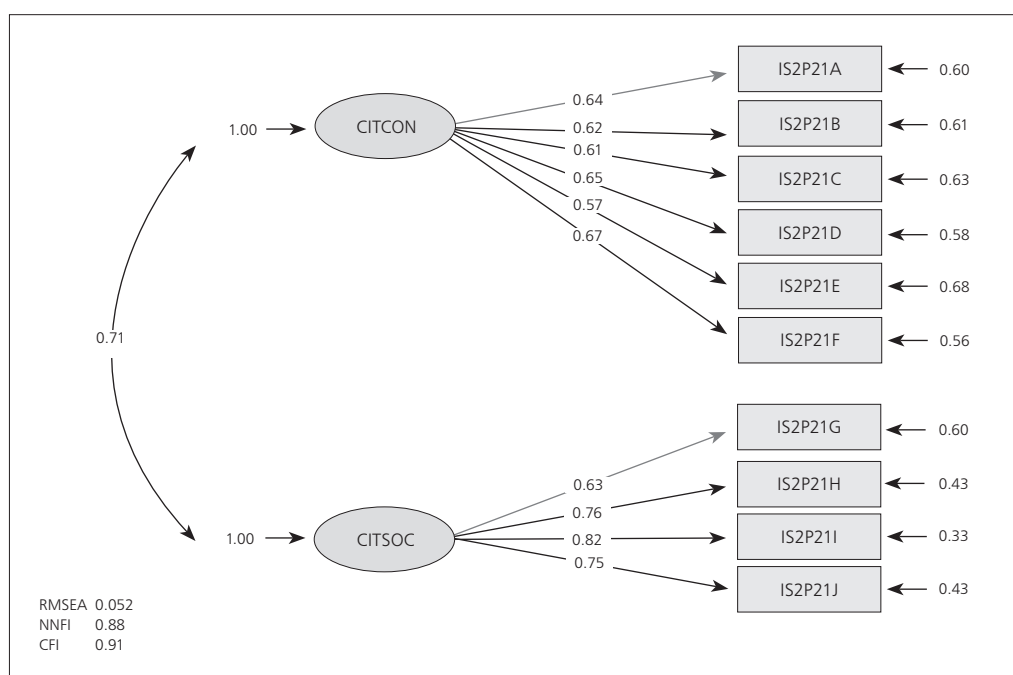


Table 12.8: Item parameters for scales reflecting students' perceptions of public service

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Conventional Citizenship</b>	How important are the following behaviours for being a good adult citizen?				
IS2P21A	Voting in every national election	-1.28	-1.45	-0.06	1.52
IS2P21B	Joining a political party	0.31	-2.07	0.62	1.46
IS2P21C	Learning about the country's history	-1.05	-1.39	-0.06	1.44
IS2P21D	Following political issues in the newspaper, on the radio, on TV or on the internet	-0.86	-1.66	-0.13	1.79
IS2P21E	Showing respect for government representatives	-0.91	-1.31	-0.41	1.72
IS2P21F	Engaging in political discussions	0.08	-2.00	0.34	1.65
<b>Soc.-Mov.-Citizenship</b>	How important are the following behaviors for being a good adult citizen?				
IS2P21G	Participating in peaceful protests against laws believed to be unjust	-0.62	-1.85	0.03	1.81
IS2P21H	Participating in activities to benefit people in the <local community>	-1.40	-1.89	-0.32	2.20
IS2P21I	Taking part in activities promoting human rights	-1.57	-1.81	-0.29	2.10
IS2P21J	Taking part in activities to protect the environment	-1.65	-1.59	-0.29	1.88

Figure 12.6 shows the results of the confirmatory factor analysis for these items. RMSEA and CFI suggested a satisfactory model fit for the two-factor solutions. Both latent factors were highly correlated with 0.71, and the item factor loadings indicated that the items had good measurement qualities for both latent traits.

Figure 12.6: Confirmatory factor analysis of items measuring students' perceptions of good citizenship



### Students' civic-related self-beliefs

Table 12.9 reports the three scales, and their reliabilities, that were derived from questions regarding students' civic-related self-beliefs. The three scales, which are included in the ICCS student database, are:

- Students' interest in politics and social issues (INTPOLS);
- Students' sense of internal political efficacy (INPOLEF);
- Students' citizenship self-efficacy (CITEFF).

Question 22 required students to indicate their interest in a series of issues. Item responses included "very interested," "quite interested," "not very interested," "not interested at all." Five of the question items were used to construct the scale *students' interest in politics and social issues* (INTPOLS); the higher scale scores correspond to greater interest in politics and social issues. The reliability of this scale (Cronbach's alpha) was 0.86 for the pooled ICCS sample. Country reliabilities ranged from 0.75 to 0.92 (see Table 12.7). Table 12.10 shows the item parameters that were used for scaling.

Question 23 asked students to state their degree of agreement or disagreement with a series of statements about their thoughts on political matters. Six items from this question were used to form the scale *students' sense of internal political efficacy* (INPOLEF), which had a reliability (Cronbach's alpha) of 0.84 for the pooled ICCS sample and ranged from 0.72 to 0.89 (see Table 12.9). Table 12.10 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect a higher sense of internal political efficacy.

Question 31 of the student questionnaire asked students how well they thought they would perform several listed activities ("very well," "fairly well," "not very well," "not at all"). Together, the question items derived the scale *students' citizenship self-efficacy* (CITEFF), and the higher values on it denote higher levels of confidence with respect to this form of self-efficacy. The reliability of this scale (Cronbach's alpha) was 0.82 for the pooled ICCS sample, and the country reliabilities ranged from 0.70 to 0.88 (see Table 12.9). Table 12.10 shows the item parameters that were used for scaling.

Figure 12.7 sets out the results from the confirmatory factor analysis for the three item sets. All three fit indices suggested a satisfactory model fit, and the factor loadings indicated good measurement qualities for the items that were used to derive indices of the three latent constructs. High correlations emerged between the three latent factors, in particular between INTPOLS (student interest in political and social issues) and INPOLEF (internal political efficacy).

### Students' attitudes toward equal rights

The three scales (and their reliabilities) that were derived from questions regarding students' attitudes toward equal rights are reported in Table 12.11 and included in the ICCS student database. They are:

- Students' attitudes toward gender equality (GENEQL);
- Students' attitudes toward equal rights for all ethnic/racial groups (ETHRGHT);
- Students' attitudes toward equal rights for immigrants (IMMRGHT).

Question 24 presented a series of items about the roles of women and men in society. Students were asked to indicate their level of agreement (ranging from "strongly agree" to "strongly disagree") with each statement. The first six question items were used to form the scale *students' attitudes toward gender equality* (GENEQL), the reliability of which (Cronbach's alpha) was 0.79 for the pooled ICCS sample. The country reliabilities ranged from 0.56 to 0.88 (see Table 12.11).



Table 12.9: Reliabilities for scales reflecting students' civic-related self-beliefs

Country	Interest Politics/ Soc. Issues	Internal Political Efficacy	Citizenship Self-Efficacy
Austria	0.83	0.85	0.77
Belgium (Flemish)	0.88	0.85	0.78
Bulgaria	0.83	0.81	0.81
Chile	0.84	0.82	0.83
Chinese Taipei	0.87	0.84	0.84
Colombia	0.83	0.78	0.80
Cyprus	0.87	0.83	0.81
Czech Republic	0.85	0.82	0.80
Denmark	0.90	0.89	0.84
Dominican Republic	0.78	0.74	0.70
England	0.90	0.87	0.87
Estonia	0.83	0.81	0.80
Finland	0.91	0.89	0.85
Greece	0.83	0.77	0.76
Guatemala	0.75	0.72	0.78
Hong Kong SAR	0.88	0.83	0.88
Indonesia	0.75	0.76	0.79
Ireland	0.87	0.86	0.84
Italy	0.85	0.84	0.80
Korea, Republic of	0.86	0.84	0.87
Latvia	0.82	0.79	0.78
Liechtenstein	0.84	0.87	0.80
Lithuania	0.84	0.79	0.80
Luxembourg	0.88	0.86	0.81
Malta	0.83	0.82	0.83
Mexico	0.80	0.78	0.80
Netherlands	0.87	0.87	0.84
New Zealand	0.89	0.87	0.88
Norway	0.92	0.87	0.86
Paraguay	0.77	0.75	0.77
Poland	0.88	0.84	0.81
Russian Federation	0.83	0.80	0.80
Slovak Republic	0.84	0.83	0.80
Slovenia	0.89	0.86	0.83
Spain	0.86	0.83	0.81
Sweden	0.92	0.89	0.88
Switzerland	0.84	0.86	0.79
Thailand	0.84	0.86	0.79
<b>ICCS average</b>	<b>0.86</b>	<b>0.84</b>	<b>0.82</b>





Table 12.10: Item parameters for scales reflecting students' civic-related self-beliefs

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Interest Politics/ Soc. Issues</b>	<b>How much are you interested in the following issues?</b>				
IS2P22A	Political issues within your <local community>	0.26	-2.73	0.40	2.34
IS2P22B	Political issues in your country	-0.14	-2.69	0.17	2.52
IS2P22C	Social issues in your country	-0.44	-2.73	0.02	2.71
IS2P22D	Politics in other countries	1.03	-2.69	0.56	2.14
IS2P22E	International politics	0.61	-2.57	0.42	2.16
<b>Internal Political Efficacy</b>	<b>How much do you agree or disagree with the following statements about you and politics?</b>				
IS2P23A	I know more about politics than most people my age	0.74	-2.56	0.63	1.94
IS2P23B	When political issues or problems are being discussed, I usually have something to say	0.24	-2.38	-0.15	2.53
IS2P23C	I am able to understand most political issues easily	0.09	-2.51	-0.14	2.66
IS2P23D	I have political opinions worth listening to	0.19	-2.40	0.04	2.36
IS2P23E	As an adult, I will be able to take part in politics	0.20	-2.04	-0.13	2.18
IS2P23F	I have a good understanding of the political issues facing this country	0.01	-2.28	-0.16	2.43
<b>Citizenship Self-Efficacy</b>	<b>How well do you think you would do the following activities?</b>				
IS2P30A	Discuss a newspaper article about a conflict between countries	-0.42	-2.24	-0.03	2.28
IS2P30B	Argue your point of view about a controversial political or social issue	-0.49	-2.13	-0.01	2.14
IS2P30C	Stand as a candidate in a <school election>	-0.27	-1.76	0.06	1.71
IS2P30D	Organise a group of students in order to achieve changes at school	-0.57	-1.77	-0.10	1.88
IS2P30E	Follow a television debate about a controversial issue	-0.20	-1.94	-0.01	1.94
IS2P30F	Write a letter to a newspaper giving your view on a current issue	-0.28	-1.68	-0.07	1.74

Table 12.12 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect stronger agreement with the notion of gender equality.

Question 25 contained a series of views on the rights and responsibilities of different ethnic/racial groups in society. Students were asked to indicate their level of agreement (ranging from “strongly agree” to “strongly disagree”) with each one. Five question items were used to construct the *students' attitudes toward equal rights for all ethnic/racial groups* (ETHRGHT) scale, where higher scores corresponded to a greater degree of agreement with the idea that ethnic and racial groups should have the same rights as other citizens in a society. The reliability of this scale (Cronbach's alpha) was 0.83 for the pooled ICCS sample. The reliabilities across countries ranged from 0.64 to 0.91 (see Table 12.11). The item parameters that were used for scaling are shown in Table 12.12.



Figure 12.7: Confirmatory factor analysis of items measuring students' civic-related self-beliefs

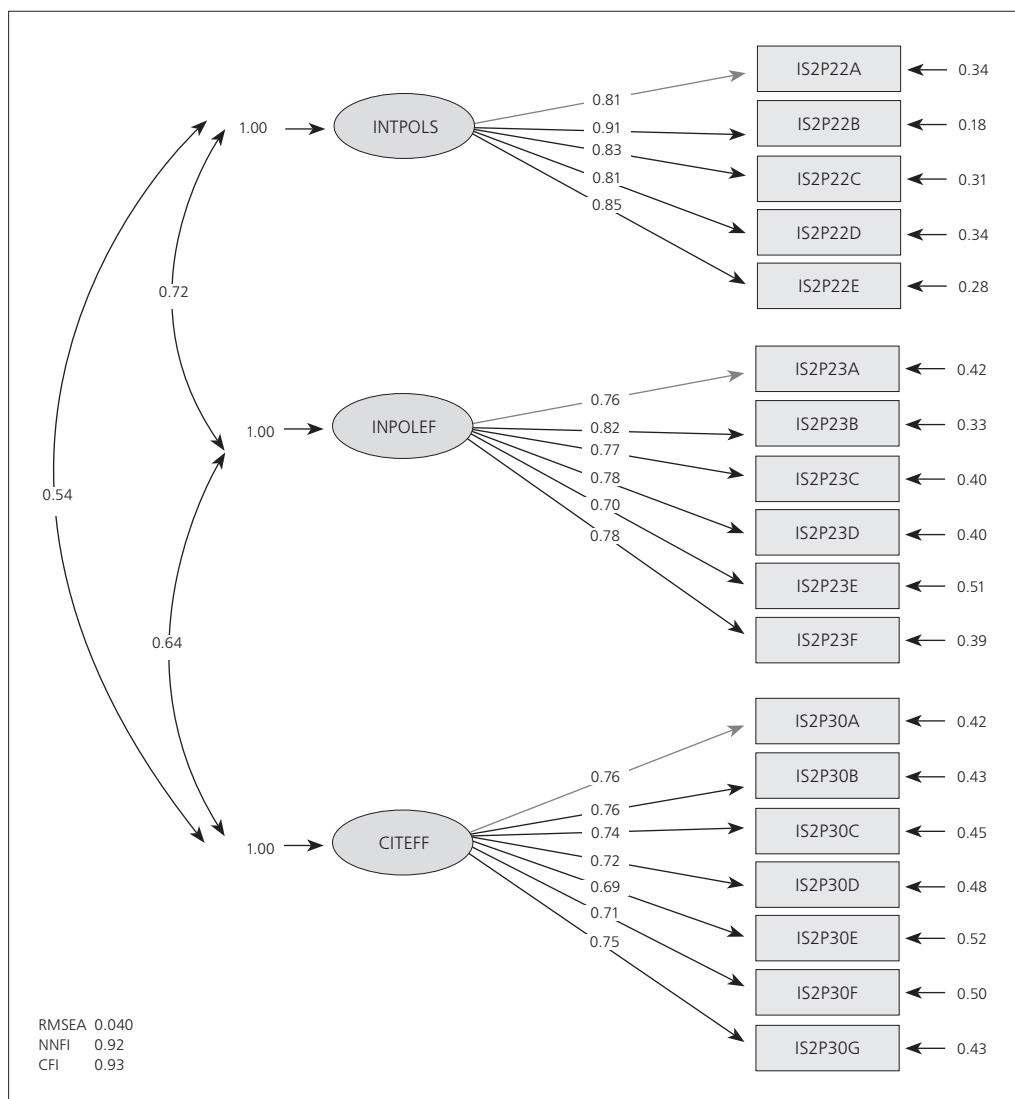


Table 12.11: Reliabilities for scales reflecting students' attitudes toward equal rights

Country	Gender Equality	Ethnic/Racial Equality	Immigrants' Rights
Austria	0.82	0.82	0.82
Belgium (Flemish)	0.79	0.84	0.77
Bulgaria	0.73	0.84	0.76
Chile	0.75	0.81	0.74
Chinese Taipei	0.81	0.82	0.84
Colombia	0.72	0.77	0.73
Cyprus	0.80	0.80	0.80
Czech Republic	0.77	0.83	0.76
Denmark	0.85	0.87	0.82
Dominican Republic	0.57	0.64	0.64
England	0.84	0.90	0.87
Estonia	0.74	0.80	0.76
Finland	0.87	0.89	0.85
Greece	0.82	0.79	0.77
Guatemala	0.71	0.75	0.72
Hong Kong SAR	0.81	0.91	0.84
Indonesia	0.62	0.66	0.50
Ireland	0.84	0.87	0.82
Italy	0.80	0.83	0.80
Korea, Republic of	0.67	0.87	0.77
Latvia	0.74	0.75	0.00
Liechtenstein	0.88	0.87	0.84
Lithuania	0.78	0.80	0.74
Luxembourg	0.79	0.85	0.82
Malta	0.76	0.78	0.79
Mexico	0.56	0.76	0.77
Netherlands	0.81	0.86	0.78
New Zealand	0.83	0.88	0.84
Norway	0.84	0.88	0.85
Paraguay	0.70	0.69	0.66
Poland	0.80	0.84	0.75
Russian Federation	0.67	0.82	0.78
Slovak Republic	0.77	0.84	0.74
Slovenia	0.83	0.80	0.80
Spain	0.75	0.83	0.81
Sweden	0.87	0.90	0.89
Switzerland	0.83	0.84	0.83
Thailand	0.83	0.84	0.83
ICCS average	0.79	0.83	0.80



Table 12.12: Item parameters for scales reflecting students' attitudes toward equal rights

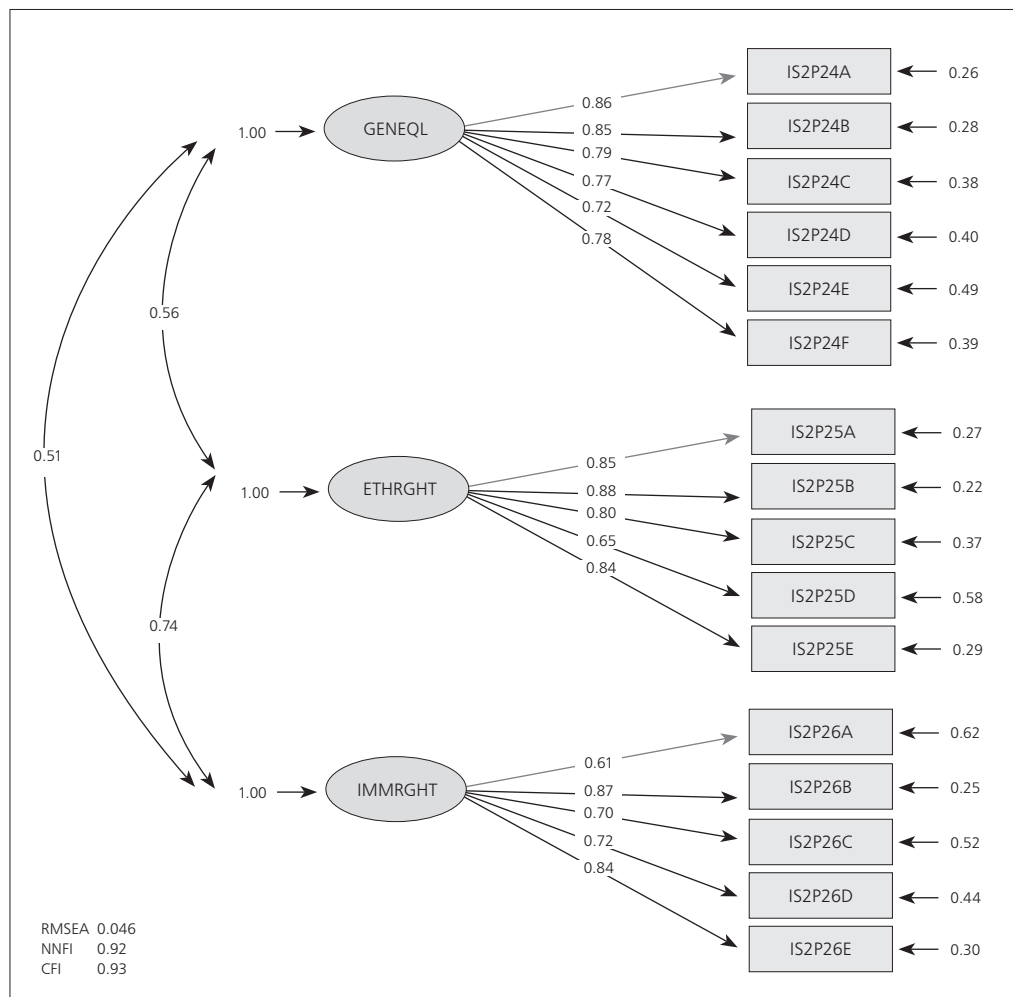
Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Gender Equality</b>	<b>There are different views about the roles of women and men in society. How much do you agree or disagree with the following statements?</b>				
IS2P24A	Men and women should have equal opportunities to take part in government	-2.49	-0.34	-0.97	1.32
IS2P24B	Men and women should have the same rights in every way	-2.43	-0.99	-0.36	1.34
IS2P24C	Women should stay out of politics	-1.63	-0.62	-0.66	1.27
IS2P24D	When there are not many jobs available, men should have more right to a job than women	-1.16	-0.88	-0.25	1.12
IS2P24E	Men and women should get equal pay when they are doing the same jobs	-2.13	-0.70	-0.37	1.06
IS2P24F	Men are better qualified to be political leaders than women	-0.99	-1.08	-0.22	1.30
<b>Ethnic/Racial Equality</b>	<b>There are different views on the rights and responsibilities of different &lt;ethnic/racial groups&gt; in society. How much do you agree or disagree with the following statements?</b>				
IS2P25A	All <ethnic/racial groups> should have an equal chance to get a good education in <country of test>	-2.75	-1.30	-1.29	2.60
IS2P25B	All <ethnic/racial groups> should have an equal chance to get good jobs in <country of test>	-2.61	-1.94	-0.87	2.80
IS2P25C	Schools should teach students to respect members of all <ethnic/racial groups>	-2.51	-1.71	-0.75	2.46
IS2P25D	<Members of all ethnic/racial groups> should be encouraged to run in elections for political office	-1.24	-2.57	-0.44	3.00
IS2P25E	<Members of all ethnic/racial groups> should have the same rights and responsibilities	-2.39	-1.41	-1.00	2.42
<b>Immigrants' Rights</b>	<b>People are increasingly moving from one country to another. How much do you agree or disagree with the following statements about &lt;immigrants&gt;?</b>				
IS2P26A	<Immigrants> should have the opportunity to continue speaking their own language	-1.06	-1.68	-0.54	2.22
IS2P26B	<Immigrant> children should have the same opportunities for education that other children in the country have	-2.20	-1.09	-1.04	2.14
IS2P26C	<Immigrants> who live in a country for several years should have the opportunity to vote in elections	-1.32	-1.81	-0.43	2.24
IS2P26D	<Immigrants> should have the opportunity to continue their own customs and lifestyle	-1.21	-1.62	-0.64	2.26
IS2P26E	<Immigrants> should have all the same rights that everyone else in the country has	-1.73	-1.44	-0.58	2.03



Question 26 presented students with a series of statements about immigrants and immigration. Students were asked to indicate their level of agreement (ranging from “strongly agree” to “strongly disagree”) with each one. Five of the items were used to construct the scale *students’ attitudes toward equal rights for immigrants* (IMMRGHT). Students with higher scores on this scale were those who agreed that immigrants should have equal rights. The reliability of this scale (Cronbach’s alpha) was 0.80 for the pooled ICCS sample, and the country reliabilities ranged from 0.50 to 0.89 (see Table 12.11). Table 12.12 shows the item parameters that were used for scaling.

Figure 12.8 shows the results of the confirmatory factor analysis that assumed a three-factor model. The fit indices indicated a satisfactory model fit, and the factor loadings showed that items generally measured the underlying latent traits in a consistent manner. High positive correlations were found between latent factors, in particular between students’ attitudes toward equal rights for all ethnic/racial groups (ETHRGHT) and for immigrants (IMMRGHT).

Figure 12.8: Confirmatory factor analysis of items measuring students’ attitudes toward equal rights



### Students' attitudes toward institutions and their country

Two scales were derived from questions regarding students' attitudes toward institutions and their country. Both were included in the student database. The two scales, the reliabilities of which are reported in Table 12.13, are:

- Students' trust in civic institutions (INTRUST);
- Students' attitudes toward their country (ATTCNT).

*Table 12.13: Reliabilities for scales reflecting students' attitudes toward civic institutions and their country*

Country	Trust. Civ. Institutions	Attitudes Toward Country
Austria	0.80	0.80
Belgium (Flemish)	0.84	0.75
Bulgaria	0.87	0.79
Chile	0.82	0.80
Chinese Taipei	0.86	0.85
Colombia	0.84	0.79
Cyprus	0.84	0.79
Czech Republic	0.82	0.81
Denmark	0.84	0.81
Dominican Republic	0.82	0.76
England	0.81	0.80
Estonia	0.79	0.88
Finland	0.86	0.84
Greece	0.83	0.74
Guatemala	0.84	0.76
Hong Kong SAR	0.84	0.86
Indonesia	0.77	0.76
Ireland	0.83	0.79
Italy	0.79	0.79
Korea, Republic of	0.84	0.79
Latvia	0.78	0.85
Liechtenstein	0.87	0.87
Lithuania	0.79	0.83
Luxembourg	0.84	0.82
Malta	0.76	0.81
Mexico	0.83	0.81
Netherlands	0.83	0.77
New Zealand	0.84	0.83
Norway	0.85	0.82
Paraguay	0.80	0.74
Poland	0.83	0.83
Russian Federation	0.81	0.84
Slovak Republic	0.83	0.84
Slovenia	0.86	0.83
Spain	0.80	0.81
Sweden	0.89	0.82
Switzerland	0.83	0.82
Thailand	0.83	0.82
ICCS average	0.84	0.82



Question 27 required students to indicate their level of trust (“completely,” “quite a lot,” “a little,” “not at all”) in up to 14 different institutions. The level of trust reported for six of the items was used to derive the scale *students’ trust in civic institutions* (INTRUST). The scale had a reliability (Cronbach’s alpha) of 0.84 for the pooled ICCS sample, and the country reliabilities ranged from 0.76 to 0.89 (see Table 12.13). Table 12.14 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect greater trust in civic institutions.

The items in Question 28 were a series of statements about the country of the test. Students were asked to indicate their level of agreement (“strongly agree,” “agree,” “disagree,” “strongly disagree”) with those statements. Seven of the question items were used to form a scale reflecting *students’ attitudes toward their country* (ATTCNT). The higher scores on the scale are from students who held the more favorable attitudes toward their country. The reliability of this scale (Cronbach’s alpha) was 0.82 for the pooled ICCS sample; the range for country reliabilities was 0.74 to 0.88 (see Table 12.13). Table 2.14 presents the item parameters that were used for scaling.

Figure 12.9 presents the results of the confirmatory factor analysis of the two item sets. The fit indices indicated not a close but an acceptable model fit for the two-factor solution. The factor loadings suggested that the latent traits could be measured with these two items sets. A high level of reliability as well as a high positive correlation were found between the two latent factors of  $r = 0.70$ .

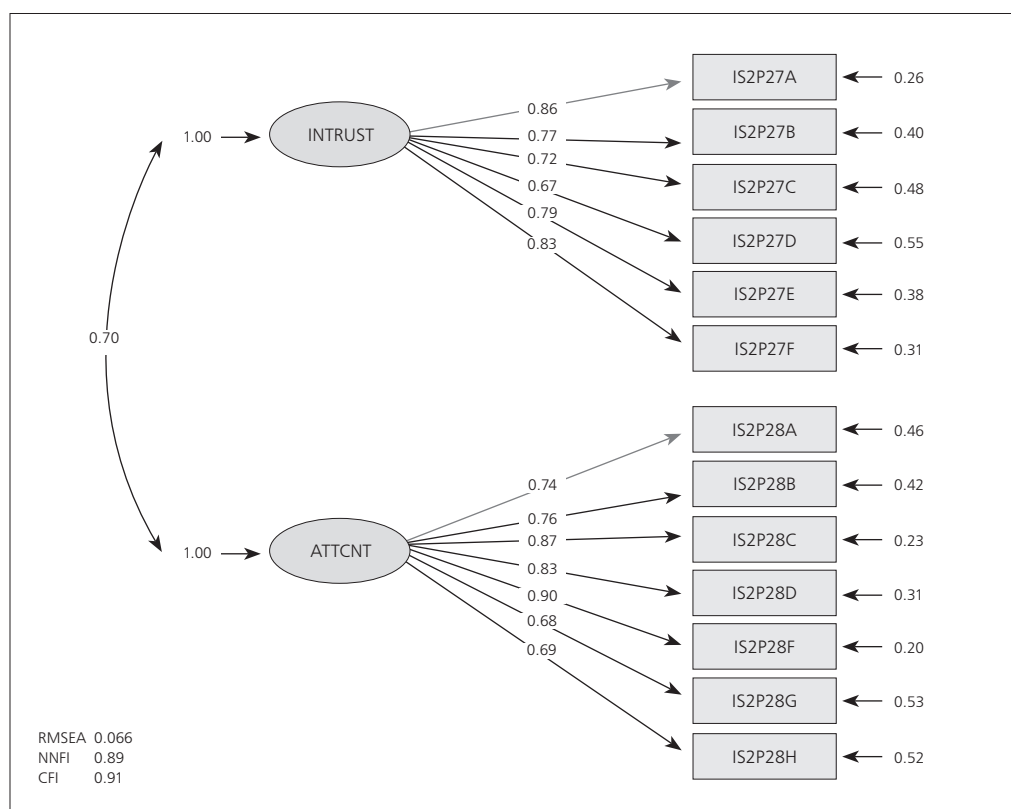
Table 12.14: Item parameters for scales reflecting students’ attitudes toward civic institutions and their country

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Trust Civ. Institutions</b>	How much do you trust each of the following institutions?				
IS2P27A	The <national government> of <country of test>	-0.53	-2.21	-0.15	2.37
IS2P27B	The <local government> of your town or city	-0.59	-2.45	-0.26	2.70
IS2P27C	Courts of justice	-0.78	-2.20	-0.22	2.41
IS2P27D	The police	-0.65	-1.64	-0.27	1.91
IS2P27E	Political parties	0.47	-2.49	-0.01	2.49
IS2P27F	<National parliament>	-0.10	-2.14	-0.18	2.32
<b>Attitudes Toward Country</b>	How much do you agree or disagree with the following statements about <country of test>?				
IS2P28A	The <flag of country of test> is important to me	-1.60	-1.13	-0.41	1.55
IS2P28B	The political system in <country of test> works well	-0.43	-2.07	-0.52	2.59
IS2P28C	I have great respect for <country of test>	-1.81	-1.30	-0.69	1.99
IS2P28D	In <country of test>, we should be proud of what we have achieved	-1.72	-1.10	-0.74	1.85
IS2P28F	I am proud to live in <country of test>	-1.77	-1.28	-0.42	1.69
IS2P28G	<Country of test> shows a lot of respect for the environment	-0.43	-1.97	-0.12	2.09
IS2P28H	Generally speaking, <country of test> is a better country to live in than most other countries	-0.96	-1.56	-0.21	1.78





Figure 12.9: Confirmatory factor analysis of items measuring students' attitudes toward civic institutions and their country



### Students' expected participation in political protest

Two scales were derived from questions regarding students' expected participation in political protest. The scales, the reliabilities of which are reported in Table 12.15, are included in the ICCS student database. They are:

- Students' expected participation in future legal protest (LEGPROT);
- Students' expected adult electoral participation (ILLPROT).

Question 31 contained a list of ways that citizens can use to protest about matters they believe are wrong ("I would certainly do this," "I would probably do this," "I would probably *not* do this," and "I would certainly *not* do this"). The first six of these items, which deal with legal protests, were used to construct the scale *students' expected participation in future legal protest* (LEGPROT). The scale had a reliability (Cronbach's alpha) of 0.79 for the pooled ICCS sample; across countries, the reliabilities ranged from 0.70 to 0.86 (see Table 12.15). Table 12.16 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect a greater likelihood of participation in future legal protests.

The remaining three items from Question 31 were used to form the scale *students' expected participation in future illegal protest* (ILLPROT). This scale had a reliability (Cronbach's alpha) of 0.83 for the pooled ICCS sample, and the country reliabilities ranged from 0.68 to 0.91 (see Table 12.15). The item parameters that were used for scaling are shown in Table 12.16.



*Table 12.15: Reliabilities for scales reflecting students' expected participation in political protest*

Country	Legal Protest	Illegal Protest
Austria	0.76	0.83
Belgium (Flemish)	0.79	0.86
Bulgaria	0.78	0.80
Chile	0.80	0.77
Chinese Taipei	0.82	0.90
Colombia	0.73	0.82
Cyprus	0.79	0.81
Czech Republic	0.80	0.84
Denmark	0.78	0.87
Dominican Republic	0.76	0.78
England	0.83	0.84
Estonia	0.74	0.81
Finland	0.82	0.87
Greece	0.70	0.78
Guatemala	0.75	0.80
Hong Kong SAR	0.82	0.91
Indonesia	0.73	0.68
Ireland	0.82	0.83
Italy	0.75	0.76
Korea, Republic of	0.86	0.84
Latvia	0.74	0.82
Liechtenstein	0.76	0.89
Lithuania	0.76	0.82
Luxembourg	0.80	0.85
Malta	0.78	0.83
Mexico	0.76	0.83
Netherlands	0.78	0.87
New Zealand	0.84	0.84
Norway	0.83	0.86
Paraguay	0.75	0.73
Poland	0.81	0.85
Russian Federation	0.77	0.77
Slovak Republic	0.79	0.84
Slovenia	0.78	0.84
Spain	0.76	0.82
Sweden	0.82	0.88
Switzerland	0.77	0.86
Thailand	0.77	0.86
ICCS average	0.79	0.83

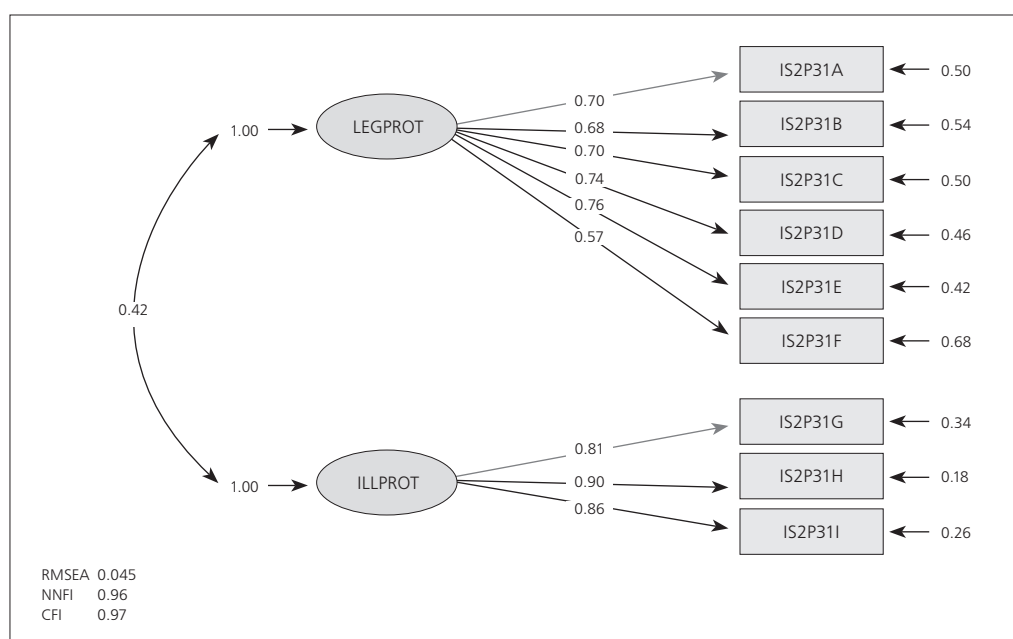


The confirmatory factor analysis, which assumed a two-factor model, suggested a close model fit, and the factor loadings indicated a high degree of item reliability (see Figure 12.10). Item IS2P31F (“choosing not to buy certain products”) had a somewhat lower factor loading than the other items measuring LEGPROT (“expected participation in legal protest”). There was a moderate positive correlation between the two latent factors of  $r = 0.42$ .

Table 12.16: Item parameters for scales reflecting students’ expected participation in political protest

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Legal Protest</b>	There are many different ways how citizens may protest against things they believe are wrong. Would you take part in any of the following forms of protest in the future?				
IS2P31A	Writing a letter to a newspaper	-0.24	-1.74	-0.01	1.74
IS2P31B	Wearing a badge or t-shirt expressing your opinion	-0.05	-1.60	0.10	1.51
IS2P31C	Contacting an <elected representative>	0.35	-1.83	0.28	1.55
IS2P31D	Taking part in a peaceful march or rally	-0.11	-1.50	-0.04	1.54
IS2P31E	Collecting signatures for a petition	-0.26	-1.55	-0.01	1.55
IS2P31F	Choosing not to buy certain products	-0.15	-1.54	-0.09	1.64
<b>Illegal Protest</b>	There are many different ways how citizens may protest against things they believe are wrong. Would you take part in any of the following forms of protest in the future?				
IS2P31G	Spray-painting protest slogans on walls	1.37	-2.06	0.60	1.47
IS2P31H	Blocking traffic	1.85	-2.14	0.73	1.41
IS2P31I	Occupying public buildings	1.92	-2.02	0.64	1.37

Figure 12.10: Confirmatory factor analysis of items measuring students’ expected participation in protest activities



### Students' expected political participation

Three scales were derived from questions regarding students' expected political participation. The scales, the reliabilities of which are given in Table 12.17, are included in the ICCS student database. They are:

- Students' expected adult electoral participation (ELECPART);
- Students' expected adult participation in political activities (POLPART);
- Students' expected future informal political participation (INFPART).

Question 32 listed several different ways that adults can take an active part in political life. Students were asked to state what they thought they would do on reaching adulthood ("I would certainly do this," "I would probably do this," "I would probably not do this," and "I would certainly not do this"). The first three items for the question were used to construct the scale *students' expected adult electoral participation* (ELECPART). A higher score for this scale meant a greater expectancy of adult electoral participation. This scale had a reliability (Cronbach's alpha) of 0.82 for the pooled ICCS sample; the country reliabilities ranged from 0.69 to 0.90 (see Table 12.17). Table 12.18 shows the item parameters that were used for scaling.

Four other Question 32 items were used to construct the POLPART scale (*students' expected adult participation in political activities*), which had a reliability (Cronbach's alpha) of 0.81 for the pooled ICCS sample. The country reliabilities ranged from 0.69 to 0.86 (see Table 12.17). The item parameters that were used for scaling are shown in Table 12.18.

Question 33 contained items listing actions in which young people can participate in the near future. The response scale was "I would certainly do this," "I would probably do this," "I would probably *not* do this," and "I would certainly *not* do this". Four of the five items made up the scale *students' expected future informal political participation* (INFPART), which had a reliability of 0.82 for the pooled ICCS sample and country reliabilities that ranged from 0.73 to 0.85 (see Table 12.17). Table 12.18 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect a greater likelihood of future informal political participation.

Figure 12.11 shows the confirmatory factor analysis assuming a three-factor solution for these item sets. The model indices suggested a good fit, and the factor loadings for the items used to measure the three latent constructs were relatively high. Positive correlations were found between the latent factors, in particular between POLPART (expected participation in political activities) and INFPART (expected informal political participation).

### Students' attitudes toward the influence of religion in society

Question 36 was part of an international option and consisted of a number of statements about what role religion should have in society. Students were asked to indicate their level of agreement with these statements ("strongly agree," "agree," "disagree," "strongly disagree"). Five of the question items were used to form a scale reflecting *students' attitudes toward the influence of religion in society* (RELINF). Higher scores on this scale indicate stronger agreement with the notion that religion should play an important role in shaping society. The reliability of this scale (Cronbach's alpha) was 0.89 for the pooled ICCS sample; country reliabilities ranged from 0.62 to 0.91 (see Table 12.19). The item parameters that were used for scaling are shown in Table 12.20.

Figure 12.12 shows the results of the confirmatory factor analysis of these items. The model fit indices suggested a satisfactory fit, while the high factor loadings indicated a high degree of measurement reliability.



Table 12.17: Reliabilities for scales reflecting students' expected political participation

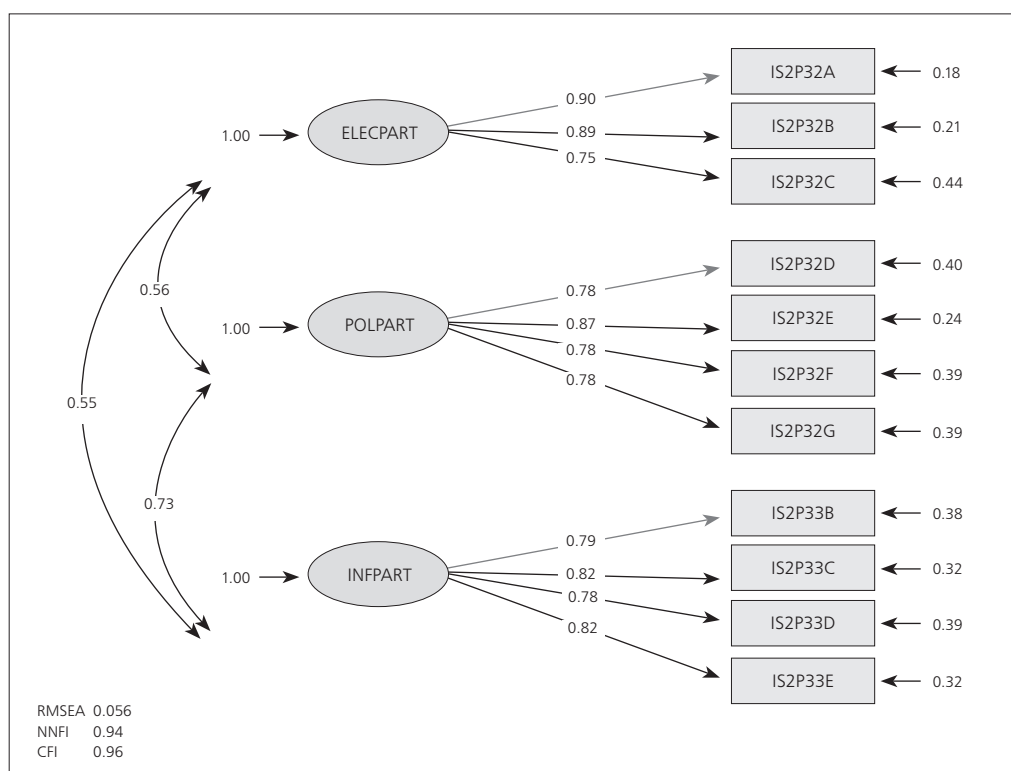
Country	Electoral Participation	Political Participation	Informal Political Participation
Austria	0.77	0.76	0.80
Belgium (Flemish)	0.78	0.82	0.82
Bulgaria	0.82	0.82	0.76
Chile	0.90	0.83	0.85
Chinese Taipei	0.86	0.80	0.85
Colombia	0.77	0.83	0.82
Cyprus	0.79	0.81	0.82
Czech Republic	0.87	0.81	0.79
Denmark	0.79	0.69	0.82
Dominican Republic	0.72	0.82	0.80
England	0.87	0.84	0.85
Estonia	0.81	0.80	0.77
Finland	0.83	0.79	0.84
Greece	0.79	0.70	0.73
Guatemala	0.72	0.86	0.82
Hong Kong SAR	0.89	0.83	0.84
Indonesia	0.69	0.76	0.77
Ireland	0.84	0.78	0.83
Italy	0.82	0.81	0.79
Korea, Republic of	0.80	0.82	0.83
Latvia	0.80	0.78	0.76
Liechtenstein	0.78	0.73	0.81
Lithuania	0.80	0.81	0.78
Luxembourg	0.84	0.82	0.83
Malta	0.71	0.80	0.83
Mexico	0.77	0.84	0.80
Netherlands	0.83	0.75	0.82
New Zealand	0.82	0.85	0.84
Norway	0.87	0.78	0.85
Paraguay	0.72	0.77	0.80
Poland	0.80	0.77	0.78
Russian Federation	0.82	0.84	0.82
Slovak Republic	0.84	0.80	0.77
Slovenia	0.83	0.77	0.76
Spain	0.84	0.81	0.82
Sweden	0.86	0.79	0.84
Switzerland	0.85	0.79	0.80
Thailand	0.85	0.79	0.80
ICCS average	0.82	0.81	0.82



Table 12.18: Item parameters for scales reflecting students' expected political participation

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Electoral Participation</b>	Listed below are different ways adults can take an active part in political life. When you are an adult, what do you think you will do?				
IS2P32A	Vote in <local elections>	-1.81	-1.64	-0.54	2.18
IS2P32B	Vote in <national elections>	-1.81	-1.73	-0.41	2.15
IS2P32C	Get information about candidates before voting in an election	-1.46	-1.89	-0.28	2.16
<b>Political Participation</b>	Listed below are different ways adults can take an active part in political life. When you are an adult, what do you think you will do?				
IS2P32D	Help a candidate or party during an election campaign	0.28	-2.51	0.37	2.15
IS2P32E	Join a political party	1.03	-2.16	0.54	1.63
IS2P32F	Join a trade union	0.86	-2.36	0.35	2.01
IS2P32G	Stand as a candidate in <local elections>	1.08	-2.04	0.43	1.62
<b>Informal Political Participation</b>	Listed below are different actions that you as a young person could take during the next few years. What do you expect that you will do?				
IS2P33B	Talk to others about your views on political and social issues	-0.24	-2.75	0.02	2.73
IS2P33C	Write to a newspaper about political and social issues	0.63	-2.79	0.50	2.28
IS2P33D	Contribute to an online discussion forum about social and political issues	0.52	-2.64	0.31	2.33
IS2P33E	Join an organisation for a political or social cause	0.71	-2.52	0.46	2.07

Figure 12.11: Confirmatory factor analysis of items measuring students' expected political participation



*Table 12.19: Reliabilities for scale reflecting students' attitudes toward the influence of religion in society*

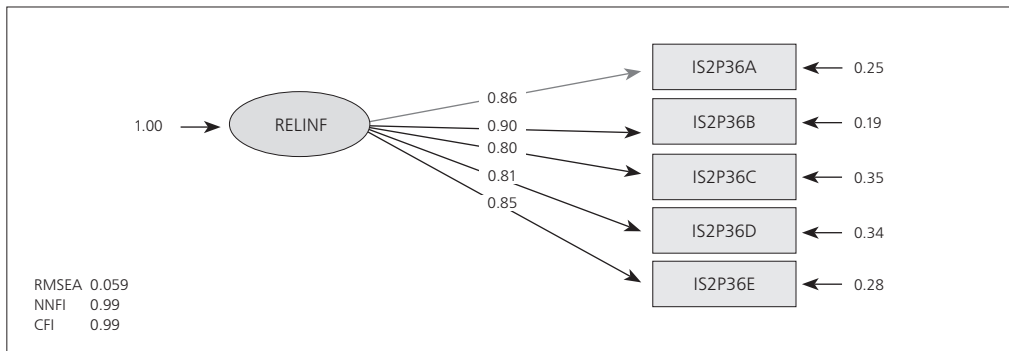
Country	Influence of Religion
Austria	0.88
Belgium (Flemish)	0.86
Bulgaria	0.84
Chile	0.84
Chinese Taipei	0.84
Colombia	0.76
Cyprus	0.79
Czech Republic	0.89
Denmark	0.88
Dominican Republic	0.68
England	0.91
Estonia	N/A
Finland	N/A
Greece	0.80
Guatemala	0.70
Hong Kong SAR	0.88
Indonesia	0.62
Ireland	N/A
Italy	N/A
Korea, Republic of	0.90
Latvia	0.87
Liechtenstein	0.90
Lithuania	0.84
Luxembourg	0.91
Malta	0.81
Mexico	N/A
Netherlands	0.89
New Zealand	0.90
Norway	0.90
Paraguay	0.71
Poland	0.86
Russian Federation	0.84
Slovak Republic	0.88
Slovenia	N/A
Spain	N/A
Sweden	0.90
Switzerland	0.91
Thailand	0.91
<b>ICCS average</b>	<b>0.89</b>



Table 12.20: Item parameters for scales reflecting students' attitudes toward the influence of religion in society

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
How much do you agree or disagree with the following statements about religion?					
IS2P36A	Religion is more important to me than what is happening in national politics	0.00	-2.35	0.32	2.04
IS2P36B	Religion helps me to decide what is right and what is wrong	0.02	-2.34	0.10	2.23
IS2P36C	Religious leaders should have more power in society	1.00	-2.92	0.39	2.52
IS2P36D	Religion should influence people's behavior towards others	-0.07	-2.33	-0.29	2.63
IS2P36E	Rules of life based on religion are more important than civil laws	0.67	-2.71	0.31	2.40

Figure 12.12: Confirmatory factor analysis of items measuring students' attitudes toward the influence of religion in society



### National index of students' socioeconomic background

The *national index of students' socioeconomic background* (NISB) was derived from the following three indices: highest occupational status of parents (HISEI), highest educational level of parents in approximate years of education according to the ISCED classification (PAREDYRS), and the approximate number of books at home (with midpoints of category ranges as values).

The process of imputing values for students who had missing data for only one of the three indicators involved use of predicted values. It also involved a random component based on a regression on the other two variables that had been estimated for students with values on all three variables. This imputation procedure was carried out separately for each national sample.

After the resulting variables, including the imputed values, had been converted into *z*-standardized variables (with a mean of 0 and a standard deviation of 1 for each national dataset), they were used for a principal component analysis that again was undertaken separately for each weighted national sample.

The final NISB scores, which were obtained as factor scores for the first principal component, had a mean of 0 and a standard deviation of 1 for each country. Table 12.21 shows the factor loadings and reliabilities for each country.





Table 12.21: Factor loadings and reliabilities for the national index of students' socioeconomic background

Country	Highest Parental Occupation	Highest Parental Education	Number of Books at Home	Reliability (Cronbach's alpha)
Austria	0.79	0.77	0.71	0.63
Belgium (Flemish)	0.81	0.81	0.59	0.58
Bulgaria	0.83	0.83	0.67	0.67
Chile	0.85	0.86	0.56	0.64
Chinese Taipei	0.80	0.82	0.66	0.64
Colombia	0.82	0.82	0.60	0.61
Cyprus	0.80	0.79	0.53	0.52
Czech Republic	0.80	0.78	0.65	0.60
Denmark	0.81	0.79	0.65	0.62
Dominican Republic	0.80	0.80	0.51	0.50
England	0.79	0.75	0.68	0.59
Estonia	0.79	0.80	0.54	0.52
Finland	0.83	0.82	0.61	0.63
Greece	0.85	0.84	0.67	0.70
Guatemala	0.83	0.85	0.54	0.61
Hong Kong SAR	0.80	0.82	0.67	0.64
Indonesia	0.85	0.86	0.34	0.51
Ireland	0.78	0.77	0.66	0.58
Italy	0.83	0.83	0.67	0.68
Korea, Republic of	0.77	0.78	0.68	0.59
Latvia	0.79	0.79	0.55	0.52
Liechtenstein	0.76	0.82	0.73	0.65
Lithuania	0.81	0.81	0.63	0.62
Luxembourg	0.84	0.82	0.76	0.73
Malta	0.81	0.81	0.60	0.59
Mexico	0.85	0.85	0.56	0.63
Netherlands	0.79	0.74	0.63	0.54
New Zealand	0.76	0.72	0.64	0.50
Norway	0.79	0.77	0.63	0.57
Paraguay	0.83	0.81	0.67	0.66
Poland	0.82	0.84	0.71	0.70
Russian Federation	0.77	0.71	0.65	0.50
Slovak Republic	0.84	0.84	0.65	0.68
Slovenia	0.82	0.82	0.56	0.58
Spain	0.84	0.85	0.67	0.69
Sweden	0.79	0.73	0.65	0.55
Switzerland	0.83	0.82	0.71	0.69
Thailand	0.81	0.83	0.54	0.58



### *Teacher questionnaire*

#### **Teachers' perceptions of school governance**

Two scales were derived from questions regarding school governance. Table 12.22 presents the reliabilities for these scales, both of which are included in the ICCS teacher database. The scales are:

- Teachers' participation in school governance (TCHPART);
- Teachers' perceptions of students' influence on decisions about school (TSTUDINF).

Question 11 presents a series of statements for teachers. These were used to form the scale *teachers' participation in school governance* (TCHPART). Respondents were asked to indicate the number of teachers in their respective schools who were participating in each of the activities indicated by the statements (response options were "all or nearly all," "most of them," "some of them," "none or hardly any"). This scale had a reliability (Cronbach's alpha) of 0.86 for the pooled ICCS sample, and the country reliabilities ranged from 0.72 to 0.91 (see Table 12.22). Table 12.23 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect greater teacher participation in school governance.

Question 13 of the teacher questionnaire asked respondents to indicate the extent to which students' opinions were taken into account when decisions were being made about school-related matters ("to a large extent," "to a moderate extent," "to a small extent," "not at all"). The first four items were used to construct the scale *teachers' perceptions of students' influence on decisions about school* (TSTUDINF). The larger scale values reflect perceptions of higher degrees of student influence. The scale had a reliability (Cronbach's alpha) of 0.80 for the pooled ICCS sample; the country reliabilities ranged from 0.60 to 0.87 (see Table 12.22). Table 12.23 presents the item parameters used for scaling.

The confirmatory factor analysis of these two items sets showed an acceptable model fit (see Figure 12.13). The size of the factor loadings illustrated that these items measured the underlying construct well. The two latent factors were positively correlated at 0.49.

#### **Teachers' perceptions of teaching in classes**

The reliabilities for the three scales that were derived from questions regarding teaching in classrooms are reported in Table 12.24. These scales, which are included in the ICCS teacher database, are:

- Confidence in teaching methods (CONFTECH);
- Teachers' use of assessment (TCASSESS);
- Teachers' reports of students' participation in class activities (TSTCLACT).

The six items relating to Question 10 in the teacher questionnaire were used to form the scale *confidence in teaching methods* (CONFTECH). This question asked teachers to rate their confidence in using a variety of teaching methods and approaches ("very confident," "quite confident," "not very confident," "not confident at all"). The scale reliability of CONFTECH was 0.73 for the pooled ICCS sample, and the country reliabilities ranged from 0.58 to 0.82 (see Table 12.24). Table 12.25 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect higher levels of confidence in teaching methods.

Question 18 was designed to determine the extent of *teachers' use of assessment* (TCASSESS). Response categories were "to a large extent," "to a moderate extent," "to a small extent or not at all". The six items that formed the scale listed a range of purposes for which assessment tasks could be used. Higher scale scores corresponded to greater use of assessments. This scale had a reliability of 0.77 for the pooled ICCS sample, while the country reliabilities ranged from 0.55 to 0.85 (see Table 12.24). Table 12.25 shows the item parameters that were used for scaling.



Table 12.22: Reliabilities for scales reflecting teachers' perceptions of school governance

Country	Participation in Schl. Governance	Students' Influence
Austria	0.86	0.64
Belgium (Flemish)	0.82	0.81
Bulgaria	0.87	0.79
Chile	0.86	0.83
Chinese Taipei	0.86	0.77
Colombia	0.85	0.79
Cyprus	0.87	0.79
Czech Republic	0.86	0.69
Denmark	0.80	0.64
Dominican Republic	0.79	0.78
England	0.82	0.80
Estonia	0.79	0.76
Finland	0.83	0.68
Greece	0.83	0.68
Guatemala	0.84	0.84
Hong Kong SAR	0.84	0.87
Indonesia	0.72	0.84
Ireland	0.85	0.80
Italy	0.87	0.78
Korea, Republic of	0.88	0.75
Latvia	0.79	0.78
Liechtenstein	0.81	0.63
Lithuania	0.84	0.79
Luxembourg	0.91	0.78
Malta	0.82	0.81
Mexico	0.91	0.81
Netherlands	0.91	0.81
New Zealand	0.82	0.76
Norway	0.85	0.63
Paraguay	0.80	0.74
Poland	0.86	0.72
Russian Federation	0.80	0.73
Slovak Republic	0.84	0.79
Slovenia	0.85	0.73
Spain	0.89	0.76
Sweden	0.81	0.73
Switzerland	0.81	0.60
Thailand	0.85	0.83
<b>ICCS average</b>	<b>0.86</b>	<b>0.80</b>



Table 12.23: Reliabilities for scales reflecting teachers' perceptions of school governance

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Particip. in Schl. Gov.</b>	<b>With reference to the current school year, how many teachers in this school ...</b>				
IT2G11A	support good discipline throughout the school even with students not belonging to their own class or classes?	-2.51	-3.34	0.13	3.21
IT2G11B	work collaboratively with one another in devising teaching activities?	-1.25	-3.39	0.44	2.94
IT2G11C	act to resolve conflict situations arising among students in the school?	-2.17	-3.56	0.41	3.14
IT2G11D	take on tasks and responsibilities in addition to teaching (tutoring, school projects, etc.)?	-1.26	-3.71	0.73	2.99
IT2G11E	actively take part in school <development/improvement activities>?	-1.53	-3.66	0.63	3.02
IT2G11F	encourage students' active participation in school life?	-2.14	-3.62	0.44	3.17
IT2G11G	cooperate in defining and drafting the <school development plan>?	-1.01	-3.05	0.68	2.38
<b>Students' Influence</b>	<b>At this school, how much are students' opinions taken into account when decisions are made about the following issues?</b>				
IT2G13A	Teaching/learning materials	0.29	-1.87	-0.13	2.00
IT2G13B	The timetable	0.86	-1.51	-0.18	1.70
IT2G13C	Classroom rules	-1.27	-1.87	-0.37	2.23
IT2G13D	School rules	-0.36	-1.77	-0.34	2.11

Figure 12.13: Confirmatory factor analysis of items measuring teachers' perceptions of school governance

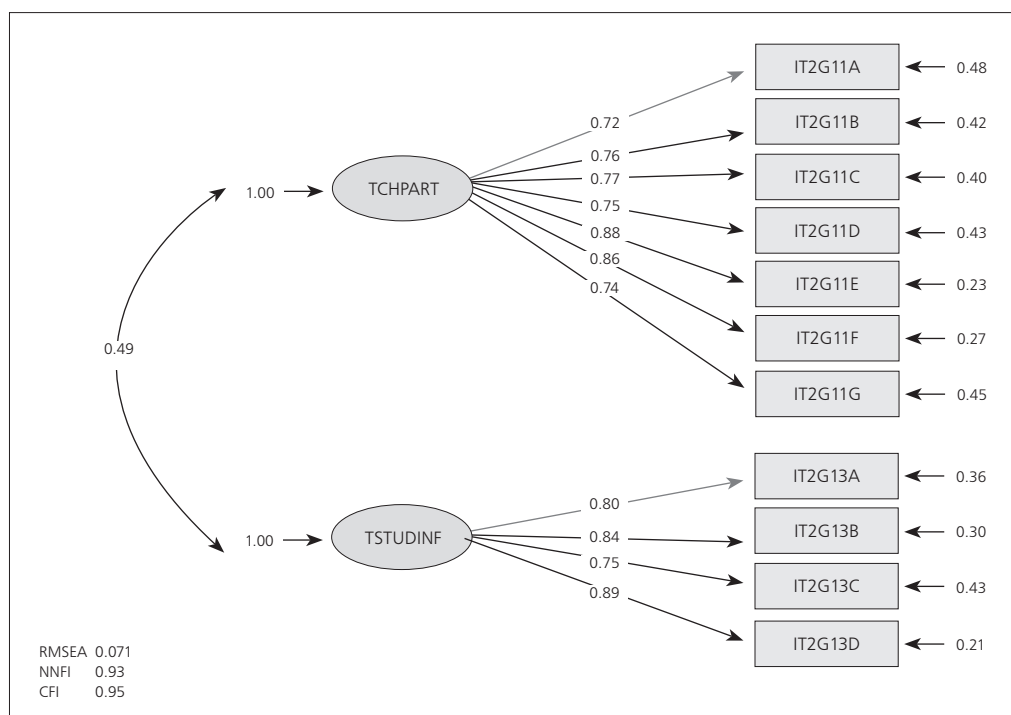


Table 12.24: Reliabilities for scales reflecting teachers' perceptions of teaching in classes

Country	Teaching Confidence	Use of Assessment	Students' Participation in Class
Austria	0.68	0.77	0.77
Belgium (Flemish)	0.58	0.78	0.81
Bulgaria	0.75	0.69	0.84
Chile	0.76	0.74	0.82
Chinese Taipei	0.82	0.84	0.86
Colombia	0.72	0.73	0.86
Cyprus	0.73	0.76	0.82
Czech Republic	0.67	0.79	0.82
Denmark	0.66	0.84	0.77
Dominican Republic	0.74	0.55	0.79
England	0.66	0.77	0.79
Estonia	0.69	0.69	0.80
Finland	0.65	0.76	0.72
Greece	0.65	0.76	0.72
Guatemala	0.70	0.69	0.83
Hong Kong SAR	0.77	0.84	0.85
Indonesia	0.69	0.79	0.85
Ireland	0.65	0.81	0.80
Italy	0.74	0.75	0.81
Korea, Republic of	0.77	0.85	0.88
Latvia	0.64	0.75	0.79
Liechtenstein	0.68	0.69	0.60
Lithuania	0.73	0.76	0.84
Luxembourg	0.68	0.85	0.85
Malta	0.60	0.76	0.78
Mexico	0.71	0.75	0.81
Netherlands	0.71	0.75	0.81
New Zealand	0.70	0.79	0.80
Norway	0.65	0.67	0.78
Paraguay	0.67	0.67	0.83
Poland	0.70	0.78	0.76
Russian Federation	0.73	0.78	0.81
Slovak Republic	0.67	0.73	0.78
Slovenia	0.73	0.77	0.80
Spain	0.79	0.82	0.75
Sweden	0.72	0.83	0.77
Switzerland	0.69	0.76	0.69
Thailand	0.80	0.82	0.84
ICCS average	0.73	0.77	0.83



Table 12.25: Item parameters for scales reflecting teachers' perceptions of teaching in classes

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Teaching Confidence</b>	How confident do you feel about using the following teaching methods and approaches?				
IT2G10A	Group work	-2.07	-1.60	-0.51	2.10
IT2G10B	Problem solving	-1.78	-1.89	-0.45	2.33
IT2G10C	Role playing, simulation	-0.89	-1.81	-0.03	1.84
IT2G10D	Classroom discussion	-1.92	-1.46	-0.41	1.86
IT2G10E	Research work	-0.95	-1.41	-0.19	1.59
IT2G10F	Lecturing	-1.72	-0.93	-0.44	1.37
IT2G10G	Laboratory activities	-0.24	-0.58	-0.34	0.93
IT2G10H	<Information and communication technology (ICT)> supported activities	-0.78	-1.44	-0.11	1.54
<b>Use of Assessment</b>	To what extent do you use the performance of your <target grade> students on assessment tasks for the following purposes?				
IT2G18A	Providing feedback to your students	-2.47	-1.68	1.68	
IT2G18B	Allowing your students to reflect on their learning processes	-2.18	-1.72	1.72	
IT2G18C	Allowing your students to reflect on their behavior	-1.38	-1.16	1.16	
IT2G18D	Identifying your students' learning difficulties	-1.88	-1.54	1.54	
IT2G18E	Providing feedback to parents	-0.34	-1.18	1.18	
IT2G18F	Illustrating learning objectives to your students	-1.72	-1.49	1.49	
IT2G18G	Planning future lessons	-1.85	-1.39	1.39	
IT2G18H	Improving your teaching	-2.49	-1.60	1.60	
<b>Students' Participation in Class</b>	In your lessons for <target grade>, how many students ...				
IT2G19A	suggest class activities?	0.50	-2.93	0.78	2.15
IT2G19B	negotiate the learning objectives with the teacher?	1.35	-2.19	0.27	1.93
IT2G19C	propose topics/issues for class discussion?	0.94	-2.68	0.63	2.04
IT2G19D	freely state their own views on school problems?	-0.48	-2.73	0.37	2.37
IT2G19E	know how to listen to and respect opinions even if different from their own?	-0.59	-3.24	0.17	3.07
IT2G19F	freely express their opinion even if different from those of the majority?	-0.65	-3.31	0.62	2.68
IT2G19G	feel comfortable during class discussions because they know their views will be respected?	-0.95	-2.89	0.16	2.74



In Question 19, teachers were asked to indicate how many students from their lessons were participating in a range of class activities (“all or nearly all,” “most of them,” “some of them,” “none or hardly any”). The eight items for this question were used to form the scale *teachers’ reports of students’ participation in class activities* (TSTCLACT), which had a reliability of 0.83 for the pooled ICCS sample. Country reliabilities ranged from 0.60 to 0.88 (see Table 12.24). Table 12.25 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect higher reported levels of student participation in class activities.

Figure 12.14 shows the results of the confirmatory factor analysis of the three item sets. The RMSEA fit index showed a close model fit for the three-factor solution whereas the NNFI and CFI had relatively low values, suggesting lack of fit for this model. Generally, the factor loadings were quite high, which indicates high reliabilities of measurement at the item level. All three latent factors were positively correlated with one another.

### **Teachers’ perceptions of participation in the community**

The reliabilities of the two scales that were derived from questions regarding teachers’ perceptions of participation in the community are reported in Table 12.26. The two scales, both of which are included in the ICCS teacher database, are:

- Teachers’ perceptions of students’ activities in the community (TSTUDACT);
- Teachers’ personal participation in activities outside school (TCHACT).

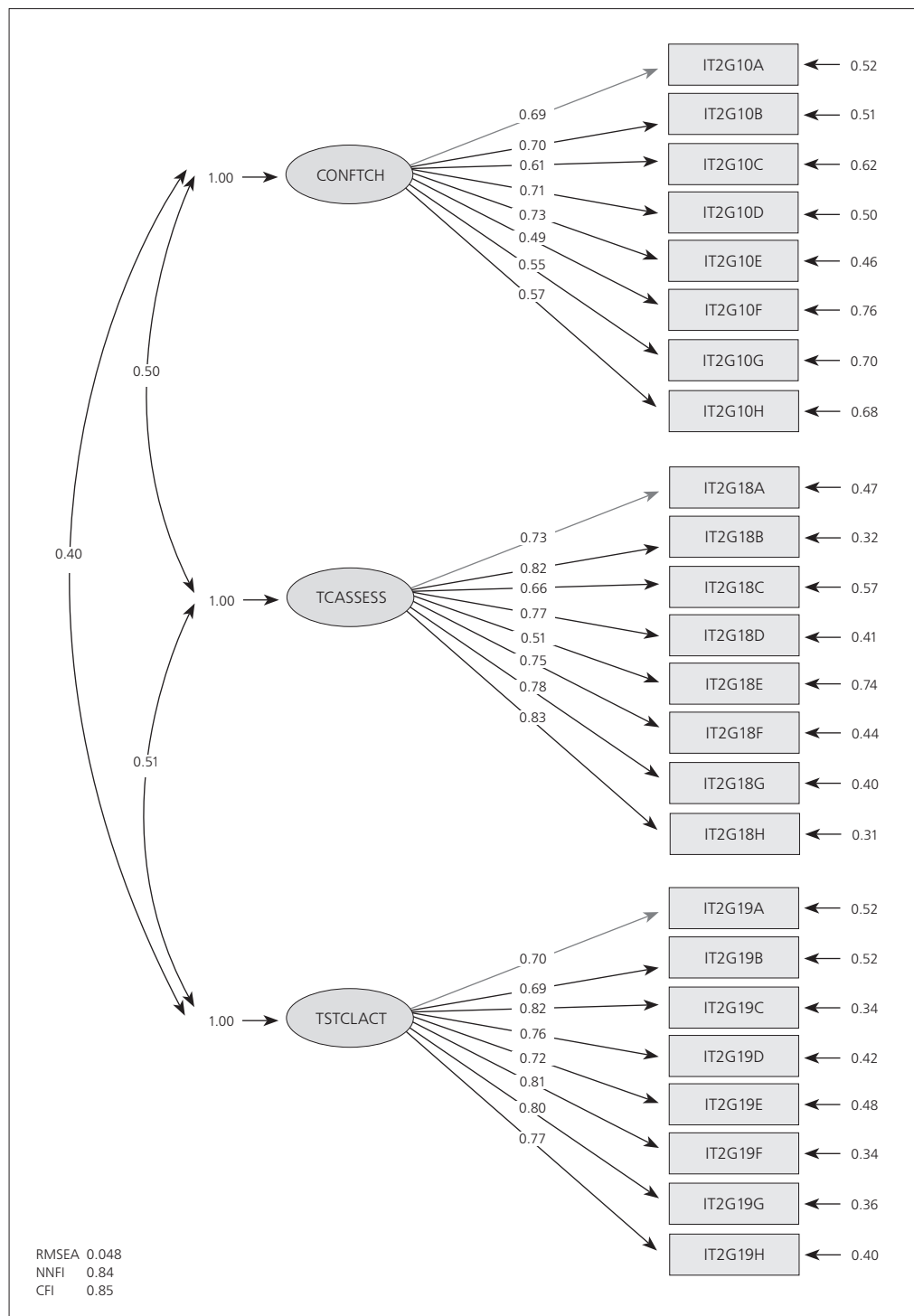
Seven of the eight items accompanying Question 15 were used to construct the scale *teachers’ perceptions of students’ activities in the community* (TSTUDACT). The items in the question required teachers to indicate whether, during the past year, their classes had taken part in activities that could be carried out by the school in cooperation with external groups/organizations (“yes” or “no”). Higher scale scores reflect greater perceived student participation in activities in the community. This scale had a reliability (Cronbach’s alpha) of 0.75 for the pooled ICCS sample and reliabilities that ranged from 0.54 to 0.79 across the participating countries (see Table 12.26). The item parameters that were used for scaling are shown in Table 12.27.

The 11 items associated with Question 16 required teachers to rate how often they personally participated in activities conducted by organizations/groups outside of their school work (“never,” “a few times,” “about once a month,” “more than once a month”). These items formed the scale *teachers’ personal participation in activities outside school* (TCHACT), with higher scores relating to greater levels of teacher participation. The reliability of TCHACT was 0.80 for the pooled ICCS sample; reliabilities across the participating countries ranged from 0.66 to 0.85 (see Table 12.26). Table 12.27 presents the item parameters that were used for scaling.

Figure 12.15 shows the results of the confirmatory factor analysis for these item sets. The RMSEA value of 0.052 suggested a satisfactory model fit for the two-factor solution. However, both the NNFI and CFI had relatively low fit values. Factor loadings tended to be quite high for both item sets, and the two latent factors were positively correlated at 0.64, which suggests that teachers who reported having themselves been involved in community activities also recorded higher frequencies of community participation with their target-grade classes.



Figure 12.14: Confirmatory factor analysis of items measuring teachers' perceptions of teaching in classrooms





*Table 12.26: Reliabilities for scales reflecting teachers' perceptions of participation in the community*

Country	Students' Activities	Teachers' Activities
Austria	0.64	0.76
Belgium (Flemish)	0.67	0.81
Bulgaria	0.68	0.78
Chile	0.66	0.81
Chinese Taipei	0.77	0.75
Colombia	0.69	0.85
Cyprus	0.68	0.82
Czech Republic	0.75	0.67
Denmark	0.58	0.70
Dominican Republic	0.73	0.84
England	0.74	0.76
Estonia	0.67	0.68
Finland	0.58	0.72
Greece	0.58	0.72
Guatemala	0.73	0.79
Hong Kong SAR	0.78	0.78
Indonesia	0.72	0.72
Ireland	0.75	0.77
Italy	0.69	0.76
Korea, Republic of	0.71	0.73
Latvia	0.69	0.74
Liechtenstein	0.59	0.66
Lithuania	0.71	0.73
Luxembourg	0.74	0.78
Malta	0.70	0.75
Mexico	0.65	0.80
Netherlands	0.65	0.80
New Zealand	0.74	0.76
Norway	0.54	0.69
Paraguay	0.69	0.77
Poland	0.69	0.74
Russian Federation	0.79	0.78
Slovak Republic	0.63	0.73
Slovenia	0.72	0.74
Spain	0.70	0.81
Sweden	0.63	0.80
Switzerland	0.57	0.71
Thailand	0.71	0.83
<b>ICCS average</b>	<b>0.75</b>	<b>0.80</b>

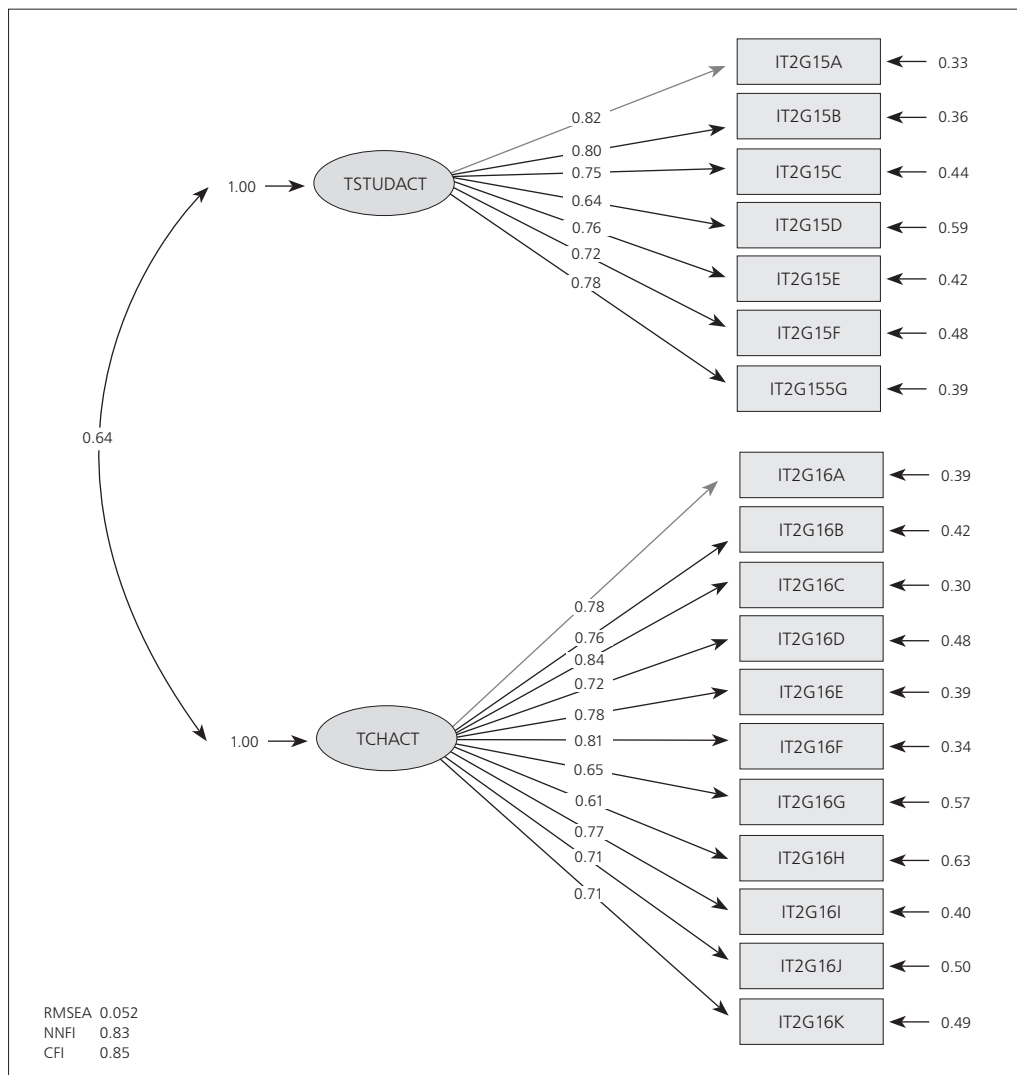


Table 12.27: Item parameters for scales reflecting teachers' perceptions of participation in the community

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
	Below is a list of activities that can be carried out by the school in cooperation with external groups/ organisations.				
<b>Students' Activities</b>	<b>During the current school year, have you and your &lt;target grade&gt; classes taken part in any of these activities?"</b>				
IT2G15A	Activities related to the environment, geared to the local area	0.07	0.00		
IT2G15B	Human rights projects	1.17	0.00		
IT2G15C	Activities related to underprivileged people or groups	1.03	0.00		
IT2G15D	Cultural activities (for example, theatre, music, cinema)	-1.20	0.00		
IT2G15E	Multicultural and intercultural activities within the <local community>	0.78	0.00		
IT2G15F	Campaigns to raise people's awareness, such as <AIDS World Day, World No Tobacco Day>	0.01	0.00		
IT2G15G	Activities related to improving facilities for the <local community> (for example, public gardens, libraries, health centres, recreation centres, community hall)	1.02	0.00		
<b>Teachers' Activities</b>	<b>Besides the activities carried out as part of your school work, how often in the last 12 months have you personally taken part in activities promoted by the following organisations/groups?</b>				
IT2G16A	Environmental organizations (<for example, WWF, Greenpeace, other national or local environmental organizations>)	1.91	-0.93	1.02	-0.10
IT2G16B	Cultural and/or educational organizations (<for example, UNESCO>)	1.88	-1.04	0.97	0.07
IT2G16C	Human rights organizations (<for example, Amnesty International>)	2.38	-0.57	0.80	-0.22
IT2G16D	Political parties or organizations	2.12	-0.38	0.69	-0.32
IT2G16E	Groups helping disadvantaged people	1.49	-1.20	1.17	0.03
IT2G16F	Cultural groups promoting the integration of ethnic minorities	2.28	-0.47	0.86	-0.40
IT2G16G	Associations promoting culture in the <local community> (<for example, exhibitions, theater performances>)	1.06	-1.51	1.12	0.40
IT2G16H	Groups run by religious organisations	1.25	-0.24	0.84	-0.59
IT2G16I	Health/disability organizations	1.90	-0.93	0.96	-0.02
IT2G16J	Trade unions	1.70	-0.69	0.66	0.04
IT2G16K	Teachers' associations	1.29	-1.30	0.72	0.57



Figure 12.15: Confirmatory factor analysis of items measuring teachers' perceptions of participation in the community



### Teachers' perceptions of school and classroom climate

The reliabilities for the three scales that were derived from questions regarding teachers' perceptions of school and classroom climate are reported in Table 12.28. The scales, as included in the ICCS teacher database, are:

- Teachers' perceptions of social problems at school (TSCPROB);
- Teachers' perceptions of student behavior at school (TSTSBEH);
- Teachers' perceptions of classroom climate (TCLCLIM).

Question 14 of the teacher questionnaire asked teachers to indicate how frequently students at their school experienced social problems ("never," "sometimes," "often," "very often"). The last two categories were collapsed for scaling. Nine items within the question were used to form the scale *teachers' perceptions of social problems at school* (TSCPROB). The scale had a reliability of 0.82 for the international sample, and the cross-national reliabilities ranged from 0.61 to 0.86 (see Table 12.28). Table 12.29 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect perceptions of a high incidence of social problems amongst students.



Table 12.28: Reliabilities for scales reflecting teachers' perceptions of classroom climate

Country	Social Problems	Students' Behavior	Classroom Climate
Austria	0.81	0.78	0.87
Belgium (Flemish)	0.80	0.84	0.91
Bulgaria	0.78	0.87	0.87
Chile	0.84	0.88	0.88
Chinese Taipei	0.79	0.89	0.89
Colombia	0.83	0.86	0.88
Cyprus	0.84	0.87	0.84
Czech Republic	0.77	0.84	0.90
Denmark	0.80	0.84	0.90
Dominican Republic	0.79	0.80	0.77
England	0.82	0.89	0.92
Estonia	0.77	0.79	0.83
Finland	0.76	0.83	0.88
Greece	0.76	0.83	0.88
Guatemala	0.83	0.82	0.80
Hong Kong SAR	0.81	0.89	0.91
Indonesia	0.61	0.87	0.81
Ireland	0.86	0.90	0.92
Italy	0.74	0.87	0.88
Korea, Republic of	0.77	0.88	0.88
Latvia	0.71	0.79	0.82
Liechtenstein	0.73	0.73	0.83
Lithuania	0.79	0.85	0.86
Luxembourg	0.85	0.85	0.93
Malta	0.83	0.87	0.89
Mexico	0.84	0.86	0.83
Netherlands	0.84	0.86	0.83
New Zealand	0.86	0.88	0.90
Norway	0.77	0.88	0.86
Paraguay	0.75	0.81	0.82
Poland	0.80	0.83	0.88
Russian Federation	0.71	0.80	0.86
Slovak Republic	0.81	0.87	0.84
Slovenia	0.79	0.88	0.86
Spain	0.82	0.89	0.90
Sweden	0.80	0.84	0.91
Switzerland	0.76	0.75	0.87
Thailand	0.82	0.87	0.82
ICCS average	0.82	0.87	0.87



Table 12.29: Item parameters for scales reflecting teachers' perceptions of classroom climate

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Social Problems</b>	Please indicate how frequently each of the following problems occurs among students at this school				
IT2G14A	Vandalism	0.66	-1.79	1.79	
IT2G14B	Truancy	-0.93	-2.54	2.54	
IT2G14C	Racism	2.29	-1.52	1.52	
IT2G14D	Religious intolerance	3.00	-1.31	1.31	
IT2G14E	Bullying	0.18	-1.89	1.89	
IT2G14F	Violence	0.76	-2.00	2.00	
IT2G14G	Sexual harassment	3.21	-1.62	1.62	
IT2G14H	Drug abuse	2.96	-1.62	1.62	
IT2G14I	Alcohol abuse	2.25	-1.60	1.60	
<b>Students' Behavior</b>	In your opinion, how many students in this school ...				
IT2G17A	are well behaved on entering and leaving the school premises?	-3.35	-4.94	-0.67	5.60
IT2G17B	have a positive attitude towards their own school?	-3.02	-5.12	-0.38	5.49
IT2G17C	have a good relationship with the school teachers and staff?	-4.41	-5.78	-0.35	6.13
IT2G17D	show care for school facilities and equipment?	-2.32	-4.98	-0.45	5.43
IT2G17E	are well behaved during breaks?	-3.17	-5.59	-0.38	5.97
IT2G17F	show they feel part of the school community?	-2.79	-4.87	-0.08	4.95
<b>Classroom Climate</b>	In your opinion, how many of your <target grade> students ...				
IT2G20A	get on well with their classmates?	-5.33	-6.28	-0.97	7.26
IT2G20B	are well integrated in the class?	-5.25	-6.82	-0.44	7.27
IT2G20C	respect their classmates even if they are different?	-3.93	-6.14	-0.15	6.29

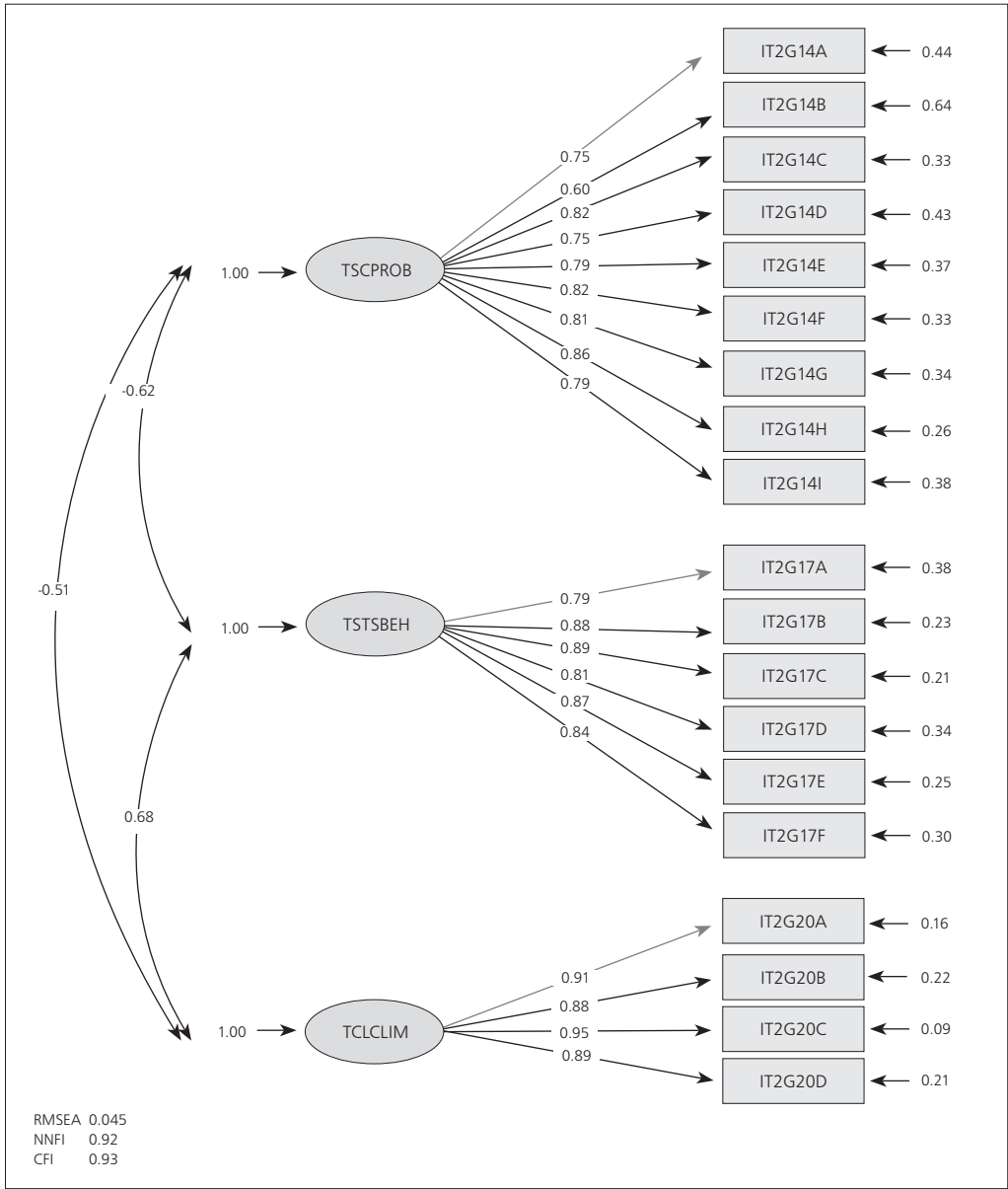
Question 17 related to issues concerning students' behavior at school. Teachers were asked to state how many students exhibited such behaviors ("all or nearly all," "most of them," "some of them," "none or hardly any"). The items for this question were used to form the scale *teachers' perceptions of students' behavior at school* (TSTSBEH). The higher scores on this scale, which had a reliability of 0.87 for the pooled ICCS sample and country reliabilities ranging from 0.73 to 0.90 (see Table 12.28), related to larger numbers of students exhibiting positive behaviors. Table 12.29 shows the item parameters that were used for scaling.

In Question 20, teachers were asked to rate how many of their students interacted with the class and other students ("all or nearly all," "most of them," "some of them," "none or hardly any"). The question items, all of which concerned student relations and integration, were used to construct the scale *teachers' perceptions of classroom climate* (TCLCLIM). The higher TCLCLIM scores indicate a more positive classroom climate. Scale reliability was 0.87 for the pooled ICCS sample; the national reliabilities ranged from 0.77 to 0.93 (see Table 12.28). Table 12.29 shows the item wording as well as the item parameters that were used for scaling.



Figure 12.16 shows the result of the confirmatory factor analysis for these three item sets. The three fit indices indicated a satisfactory model, while the factor loadings tended to be very high for all three factors. Negative correlations were found between TSCPROB (perceptions of social problems at school) and the two other factors. However, positive correlations were also found between teacher perceptions of (positive) student behavior (TSTSBEH) and (positive) perceptions of school climate (TCLCLIM).

Figure 12.16: Confirmatory factor analysis of items measuring teachers' perceptions of school climate



### Teachers' reports of teaching civic and citizenship education

Two scales were derived from questions that asked teachers to provide information about the teaching of civic and citizenship education in their schools. These (optional) questions were directed only at those teachers who were teaching subjects related to this learning area. The reliabilities for these two scales, both of which are included in the ICCS teacher database (see immediately below), are reported in Table 12.30.

- Teachers' reports on civic and citizenship education activities in class (TCIVACT);
- Teachers' confidence in civic and citizenship education teaching (TCIVCONF).

Question 25 listed a range of activities likely to occur during class. Teachers were asked to rate how often ("never," "sometimes," "often," "very often") these activities occurred during their classes featuring civic and citizenship education. Six of the question items were used to construct the scale *teachers' reports on civics and citizenship activities in class* (TCIVACT). The higher scores on this scale correspond to a higher occurrence of civic and citizenship activities. The scale had a reliability of 0.78 for the pooled ICCS sample, and the range of reliabilities across the participating countries was 0.45 to 0.83 (see Table 12.30). Table 12.31 shows the item parameters that were used for scaling.

Question 28 asked teachers how confident they felt about teaching a range of topics associated with civic and citizenship education ("very confident," "quite confident," "not very confident," "not confident at all"). Fourteen of these items were used to form the scale *teachers' confidence in teaching civic and citizenship education* (TCIVCONF), which had a reliability of 0.90 for the pooled ICCS sample and scale reliabilities (cross-country) ranging from 0.83 to 0.93 (see Table 12.30). Table 12.31 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale denote greater teacher confidence in teaching topics related to civic and citizenship education.

The results of the confirmatory factor analysis of the two item sets suggested a satisfactory model fit (see Figure 12.17) and showed that most items had strong factor loadings on both factors. The correlation between both factors was positive at 0.53.



Table 12.30: Reliabilities for scales reflecting teachers' perceptions of classroom climate

Country	Class Activities	Teacher Confidence
Austria	0.76	0.83
Belgium (Flemish)	0.71	0.91
Bulgaria	0.78	0.88
Chile	0.78	0.92
Chinese Taipei	0.76	0.93
Colombia	0.78	0.90
Cyprus	0.63	0.93
Czech Republic	0.77	0.87
Denmark	0.65	0.89
Dominican Republic	0.72	0.89
England	0.76	0.91
Estonia	N/A	N/A
Finland	0.70	0.91
Greece	0.70	0.91
Guatemala	N/A	N/A
Hong Kong SAR	0.83	0.91
Indonesia	0.78	0.84
Ireland	0.70	0.91
Italy	0.73	0.89
Korea, Republic of	0.76	0.92
Latvia	0.63	0.86
Liechtenstein	0.79	0.88
Lithuania	0.82	0.91
Luxembourg	N/A	N/A
Malta	0.73	0.91
Mexico	0.75	0.89
Netherlands	0.75	0.89
New Zealand	0.72	0.89
Norway	0.45	0.91
Paraguay	0.76	0.84
Poland	0.71	0.84
Russian Federation	0.72	0.89
Slovak Republic	0.73	0.86
Slovenia	0.76	0.91
Spain	0.72	0.92
Sweden	0.63	0.89
Switzerland	0.65	0.83
Thailand	0.72	0.93
ICCS average	0.78	0.90



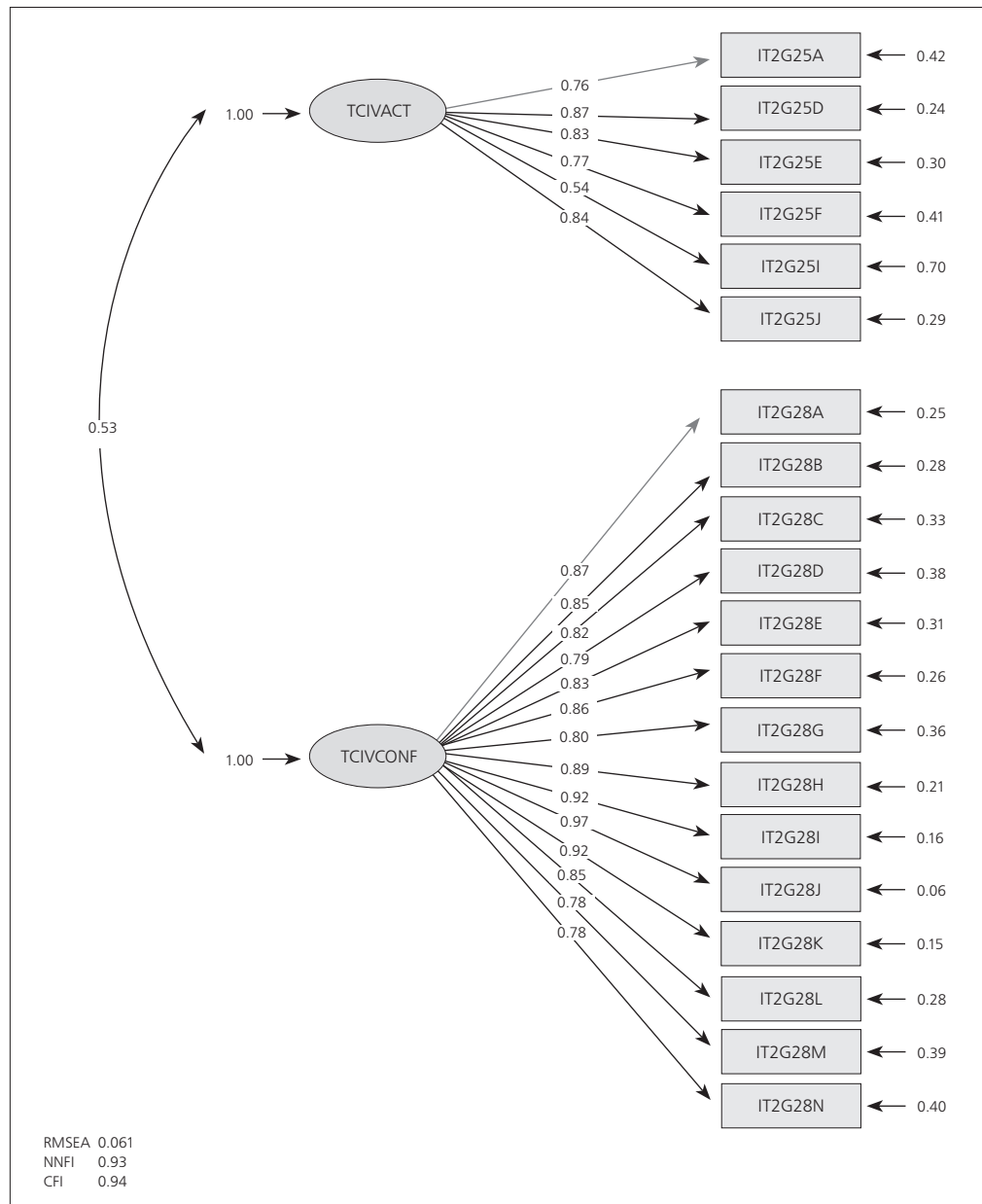


Table 12.31: Item parameters for scales reflecting teachers' perceptions of classroom climate

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Civ. Ed. in Class</b>	<b>How often do the following activities occur during your &lt;civic and citizenship education&gt; classes at &lt;target grade&gt;?</b>				
IT2G25A	Students work on projects that involve gathering information outside of school	0.41	-3.04	0.77	2.28
IT2G25D	Students work in groups on different topics and prepare presentations	-0.55	-2.65	0.42	2.24
IT2G25E	Students work individually on different topics and prepare presentations	0.10	-2.68	0.42	2.25
IT2G25F	Students participate in role play and simulations	0.67	-2.44	0.58	1.85
IT2G25I	The teacher includes discussion on controversial issues in class	-0.79	-2.76	0.40	2.35
IT2G25J	Students research and analyse information from different sources	-0.44	-2.72	0.45	2.26
<b>Confidence Teaching Civ. Ed.</b>	<b>How confident do you feel about teaching the following topics?</b>				
IT2G28A	Human rights	-2.18	-2.08	-0.65	2.72
IT2G28B	Different cultures and ethnic groups	-1.67	-2.48	-0.30	2.77
IT2G28C	Voting and elections	-1.71	-1.95	-0.25	2.20
IT2G28D	The economy and business	-0.53	-2.52	0.24	2.29
IT2G28E	Rights and responsibilities at work	-1.81	-2.29	-0.22	2.51
IT2G28F	The global community and international organizations	-1.18	-2.78	0.11	2.66
IT2G28G	The environment	-2.33	-2.20	-0.49	2.68
IT2G28H	Emigration and immigration	-1.43	-2.46	-0.12	2.57
IT2G28I	Equal opportunities for men and women	-2.31	-1.97	-0.68	2.64
IT2G28J	Citizens' rights and responsibilities	-2.40	-1.83	-0.64	2.47
IT2G28K	The constitution and political systems	-1.36	-1.97	-0.11	2.09
IT2G28L	Media communication	-1.78	-2.19	-0.41	2.61
IT2G28M	Volunteering	-1.15	-2.84	0.29	2.55
IT2G28N	Legal institutions and courts	-0.47	-2.47	0.27	2.19



Figure 12.17: Confirmatory factor analysis of items reflecting teachers' reports on civic and citizenship education



### *School questionnaire*

#### **Principals' reports on school governance**

Four scales were derived from questions regarding principals' reports on school governance. Table 12.32 presents the reliabilities of the four scales, all of which are included in the ICCS school database.

- Principals' perceptions of school autonomy (SCAUTON);
- Principals' perceptions of teachers' participation in school governance (SCTCPART);
- Principals' perceptions of parents' participation in school life (SCPARACT);
- Principals' perceptions of students' influence on decisions about school (CSTUDINF).

Question 4 of the school questionnaire asked principals to give their perceptions of the extent of autonomy—"full," "quite a lot," "little," "none"—that their school had with respect to a range of matters. The 12 question items were used to construct the scale *principals' perceptions of school autonomy* (SCAUTON). The scale had a reliability (Cronbach's alpha) of 0.87 for the international sample and reliabilities ranging from 0.54 to 0.93 across the participating countries (see Table 12.32). Table 12.33 shows the item wording and the item parameters that were used for scaling. Higher values on this scale reflect perceptions of relatively high incidences of school autonomy.

Question 5 incorporated a series of statements about teachers' participation in running the school. Principals were asked to indicate the number of teachers who they thought participated in each of the actions ("all or nearly all," "most of them," "some of them," "none or hardly any," "not applicable"). The last category was treated as a missing value during scaling. Seven of the question items were used to form the scale *principals' perceptions of teachers' participation in school governance* (SCTCPART). Here, the higher scale scores relate to perceptions of a considerable amount of teacher participation. This scale had a reliability (Cronbach's alpha) of 0.86 for the pooled ICCS sample and cross-country reliabilities that ranged from 0.72 to 0.95 (see Table 12.32). Table 12.33 shows the item parameters that were used for scaling.

In Question 8, principals provided their opinion on the number of parents who participated in five different activities ("all or nearly all," "most of them," "some of them," "none or hardly any," "not applicable"). The last category was treated as a missing value during scaling. All five items were used to form the scale *principals' perceptions of parents' participation in school life* (SCPARACT). This scale had a reliability (Cronbach's alpha) of 0.77 for the pooled ICCS sample and national reliabilities that ranged from 0.38 to 0.84 (see Table 12.28). Table 12.29 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect perceptions of a high incidence of parental participation in school life.

Question 10 asked principals to rate how much students' opinions were taken into account during consideration of school-based issues ("to a large extent," "to a moderate extent," "to a small extent," "not at all"). Four of the five items in this question were used to construct the scale *principals' perceptions of students' influence on decisions about school* (CSTUDINF). This scale had a reliability (Cronbach's alpha) of 0.75 for the international sample; country reliabilities ranged from 0.14 to 0.86 (see Table 12.28). Table 12.29 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale signify perceptions of a higher degree of student-based influence on decisions about school.

Figure 12.18 shows the results of the confirmatory factor analysis for the four item sets. The fit indices indicated a satisfactory model fit, and the size of the factor loadings indicated generally good item reliabilities. All four factors were positively correlated, with estimates ranging from 0.37 to 0.59.



Table 12.32: Reliabilities for scales reflecting principals' reports on school governance

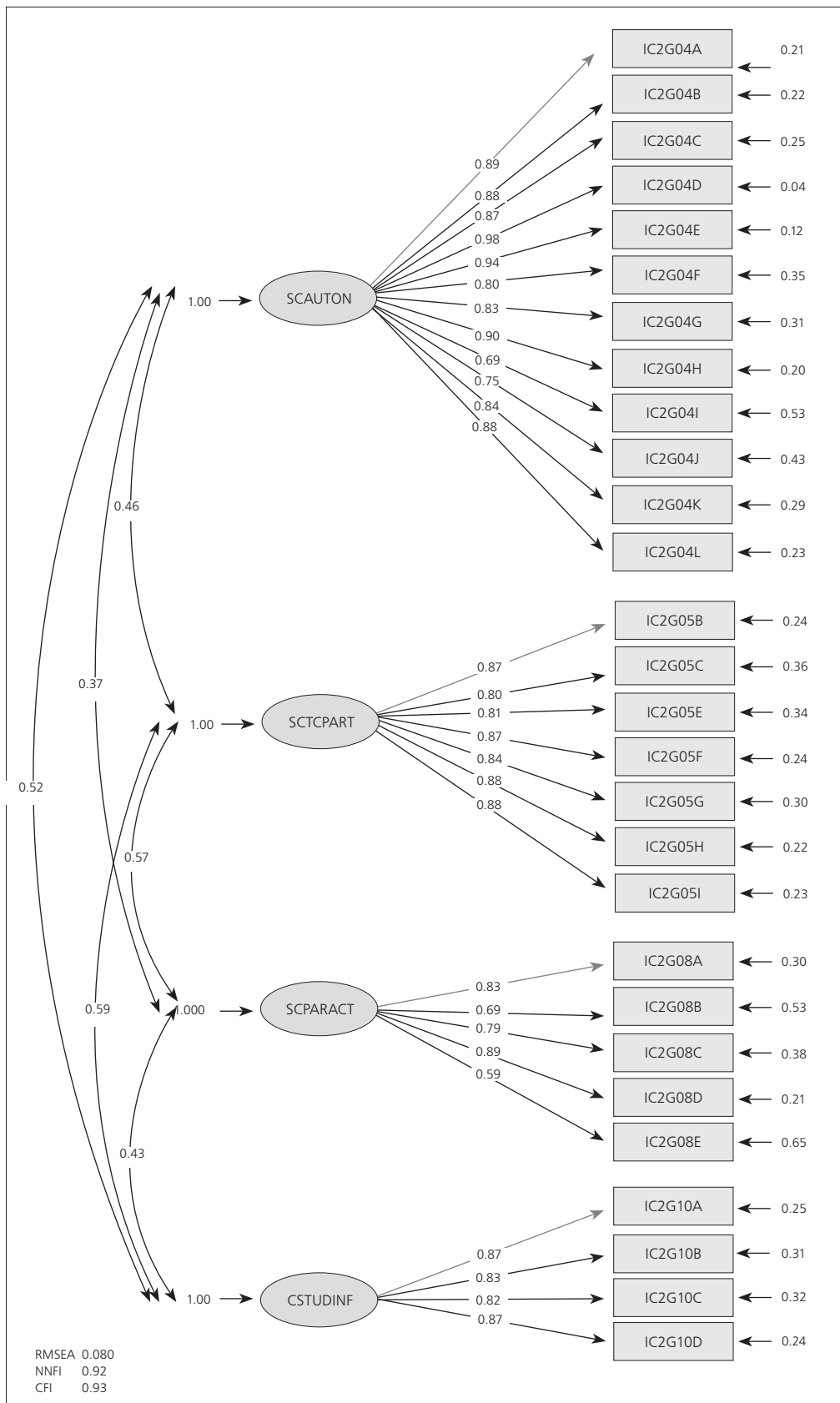
Country	School Autonomy	Teachers' Participation	Parents' Participation	Students' Influence
Austria	0.71	0.81	0.73	0.57
Belgium (Flemish)	0.54	0.75	0.52	0.73
Bulgaria	0.67	0.79	N/A	0.78
Chile	0.86	0.91	0.74	0.83
Chinese Taipei	0.87	0.87	0.79	0.70
Colombia	0.88	0.89	0.73	0.71
Cyprus	0.67	0.90	0.71	0.80
Czech Republic	0.82	0.81	0.70	0.67
Denmark	0.79	0.73	0.56	0.59
Dominican Republic	0.89	0.83	0.74	0.79
England	0.82	0.77	0.67	0.78
Estonia	0.80	0.80	N/A	0.71
Finland	0.78	0.85	0.52	0.66
Greece	0.77	0.83	0.84	0.72
Guatemala	0.88	0.82	0.77	0.86
Hong Kong SAR	0.82	0.87	0.47	0.84
Indonesia	0.75	0.83	0.70	0.85
Ireland	0.66	0.82	0.71	0.72
Italy	0.78	0.86	0.76	0.68
Korea, Republic of	0.84	0.85	0.78	0.77
Latvia	0.73	0.79	0.74	0.69
Liechtenstein	0.85	0.89	0.42	0.14
Lithuania	0.79	0.85	0.73	0.66
Luxembourg	0.90	0.81	0.76	0.45
Malta	0.93	0.78	0.69	0.72
Mexico	0.87	0.95	0.80	0.77
Netherlands	0.87	0.72	0.58	0.43
New Zealand	0.79	0.74	0.57	0.73
Norway	0.56	0.81	0.60	0.29
Paraguay	0.81	0.87	0.78	0.70
Poland	0.84	0.78	N/A	0.66
Russian Federation	0.79	0.81	0.82	0.70
Slovak Republic	0.76	0.84	0.74	0.74
Slovenia	0.81	0.83	0.74	0.65
Spain	0.77	0.91	0.81	0.56
Sweden	0.84	0.88	0.59	0.73
Switzerland	0.71	0.81	0.38	0.32
Thailand	0.81	0.89	0.80	0.85
ICCS average	<b>0.87</b>	<b>0.86</b>	<b>0.77</b>	<b>0.75</b>



Table 12.33: Item parameters for scales reflecting principals' reports on school governance

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>School Autonomy</b>	How much autonomy does this school have in relation to the following issues?				
IC2G04A	Curriculum planning	-0.47	-0.80	-0.50	1.31
IC2G04B	Curriculum delivery	-1.21	-1.21	-0.45	1.65
IC2G04C	Choice and use of textbooks	-1.20	-0.60	-0.16	0.77
IC2G04D	Appointing teachers	-0.40	0.54	-0.63	0.08
IC2G04E	Dismissing teachers	0.15	-0.05	-0.11	0.15
IC2G04F	Establishing student assessment policies	-0.89	-1.30	-0.44	1.75
IC2G04G	Determining the content of in-service professional development programs for teachers	-0.73	-0.83	-0.44	1.26
IC2G04H	Teacher appraisal	-0.84	-0.54	-0.47	1.00
IC2G04I	Budget allocations within the school	-0.89	-0.89	-0.31	1.21
IC2G04J	<Extracurricular activities>	-1.95	-1.50	-0.14	1.65
IC2G04K	Student admittance policies	-0.33	-0.41	-0.18	0.59
IC2G04L	Establishing teachers' salaries	0.99	0.02	-0.18	0.15
	The following statements refer to teachers' participation in running the school.				
<b>Teachers' Participation</b>	In your opinion, how many teachers in this school ...				
IC2G05B	make their own contribution to solving school problems?	-1.80	-4.06	0.42	3.65
IC2G05C	put forward useful suggestions for improving school governance?	-0.64	-3.70	0.79	2.91
IC2G05E	contribute to establishing school priorities?	-1.02	-3.62	0.45	3.16
IC2G05F	support good discipline throughout the school even with students not belonging to their own class or classes?	-2.27	-3.66	0.13	3.52
IC2G05G	act to resolve conflict situations arising among the students in the school?	-2.41	-3.83	0.27	3.57
IC2G05H	actively take part in school <development/improvement activities>?	-2.07	-4.30	0.58	3.71
IC2G05I	encourage students' active participation in school life?	-2.30	-4.17	0.45	3.72
<b>Parents' Participation</b>	In your opinion, how many parents of students in this school participate in the following activities?				
IC2G08A	Taking part actively in the school parent <association, assembly, committee>	0.47	-2.81	0.60	2.20
IC2G08B	Voting in <school council, school representative body> elections	0.43	-1.99	0.04	1.96
IC2G08C	Supporting school projects within the <local community>	0.17	-2.60	0.43	2.16
IC2G08D	Attending school parent <association, assembly, committee> meetings	0.07	-2.40	0.22	2.17
IC2G08E	Attending parent-teacher meetings	-1.40	-2.39	-0.19	2.57
<b>Students' Influence</b>	In this school, how much are students' opinions taken into account when decisions are made about the following issues?				
IC2G10A	Teaching/learning materials	0.51	-1.89	-0.1	1.99
IC2G10B	The timetable	0.52	-1.63	-0.07	1.71
IC2G10C	Classroom rules	-1.83	-1.76	-0.27	2.04
IC2G10D	School rules	-1.00	-1.83	-0.36	2.20

Figure 12.18: Confirmatory factor analysis of items reflecting teachers' reports on school governance



### Principals' reports on the local community

The three scales that were derived from questions regarding principals' reports on the local community are included in the ICCS school database. Their reliabilities are reported in Table 12.34.

- Principals' perceptions of students' opportunities to participate in community activities (SCSTUDOP);
- Availability of resources in local community (RESCOM);
- Principals' perceptions of social tension in the community (COMSOCT).

Principals were asked in Question 6 to indicate how many students during the current school year had received the opportunity to take part in activities that could be carried out by the school in cooperation with external groups/organizations. The response categories were "all or nearly all," "most of them," "some of them," "none or hardly any," and "not offered at school." During scaling of these items, the last category was combined with the category "none or hardly any" and both were assigned a value of 0.

Seven of the question's items were used to form the scale *principals' perceptions of students' opportunities to participate in community activities* (SCSTUDOP); higher scale scores correspond to perceptions that students had a good many opportunities to participate in these activities. This scale had a reliability (Cronbach's alpha) of 0.75 for the pooled ICCS sample; reliabilities ranged from 0.30 to 0.88 across the participating countries (see Table 12.34). Table 12.35 shows the item parameters that were used for scaling.

Question 13 asked principals to respond with "yes" or "no" as to whether different resources were available in the local area in which their school was located. Six of the question items were used to construct the scale *availability of resources in the local community* (RESCOM). This scale had a reliability (Cronbach's alpha) of 0.80 for the pooled ICCS sample and reliabilities that ranged from 0.52 to 0.86 across the participating countries (see Table 12.34). Table 12.35 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale denote greater availability of resources within the school's local community.

Question 14 required principals to rate the extent—"to a large extent," "to a moderate extent," "to a small extent," "not at all"—to which a series of issues were a source of social tension in the school's locality. All question items were used to form the scale *principals' perceptions of social tension in the community* (COMSOCT). The larger COMSOCT scores relate to increased levels of social tension. This scale had a reliability (Cronbach's alpha) of 0.88 for the international sample. Country reliabilities ranged from 0.74 to 0.93 (see Table 12.34). Table 12.35 shows the item parameters that were used for scaling.

Figure 12.19 presents the results of the confirmatory factor analysis for these three item sets. The model fit was satisfactory, and the factor loadings indicated good measurement properties for most of the items. Correlations between the three latent factors were weak.



Table 12.34: Reliabilities for scales reflecting principals' reports on the local community

Country	Students' Particip. Opps.	Resources	Social Tension
Austria	0.65	0.71	0.89
Belgium (Flemish)	0.57	0.70	0.89
Bulgaria	0.68	0.80	0.76
Chile	0.79	0.82	0.87
Chinese Taipei	0.85	0.67	0.90
Colombia	0.80	0.80	0.90
Cyprus	0.80	0.80	0.91
Czech Republic	0.72	0.77	0.91
Denmark	0.81	0.83	0.90
Dominican Republic	0.75	0.80	0.86
England	0.59	0.68	0.92
Estonia	0.62	0.78	0.89
Finland	0.66	0.72	0.83
Greece	0.73	0.73	0.89
Guatemala	0.87	0.79	0.86
Hong Kong SAR	0.66	0.52	0.91
Indonesia	0.83	0.81	0.82
Ireland	0.75	0.76	0.93
Italy	0.66	0.72	0.86
Korea, Republic of	0.80	0.77	0.88
Latvia	0.67	0.77	0.85
Liechtenstein	0.30	0.70	0.93
Lithuania	0.74	0.70	0.86
Luxembourg	0.81	0.69	0.92
Malta	0.73	0.65	0.93
Mexico	0.81	0.83	0.83
Netherlands	0.70	0.73	0.83
New Zealand	0.82	0.68	0.91
Norway	0.72	0.75	0.87
Paraguay	0.84	0.82	0.83
Poland	0.77	0.84	0.84
Russian Federation	0.77	0.65	0.82
Slovak Republic	0.73	0.83	0.85
Slovenia	0.68	0.86	0.88
Spain	0.65	0.75	0.90
Sweden	0.77	0.77	0.93
Switzerland	0.59	0.74	0.86
Thailand	0.88	0.85	0.74
ICCS average	0.75	0.80	0.88

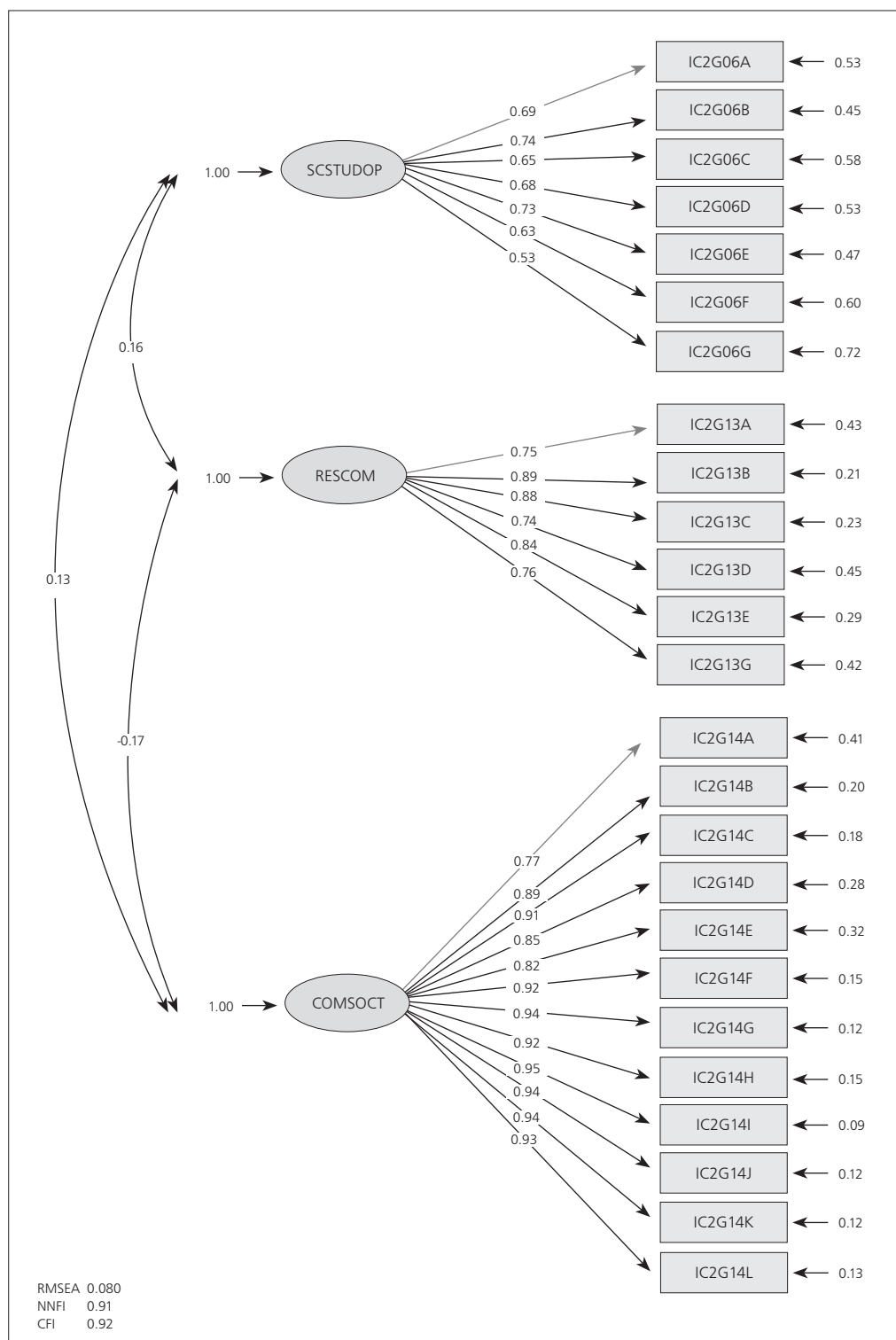




Table 12.35: Item parameters for scales reflecting principals' reports on the local community

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
	Below is a list of activities that may be carried out by the school in cooperation with external groups/organizations.				
<b>Students' Particip. Opps.</b>	<b>During the current school year, how many &lt;target grade&gt; students in this school have had the opportunity to take part in any of these activities?</b>				
IC2G06A	Activities related to the environment, geared to the local area	-0.55	-1.54	0.46	1.08
IC2G06B	Human rights projects	0.15	-1.14	0.38	0.76
IC2G06C	Activities related to underprivileged people or groups	0.02	-1.40	0.64	0.77
IC2G06D	Cultural activities (for example, theater, music, cinema)	-1.32	-1.76	0.55	1.22
IC2G06E	Multicultural and intercultural initiatives within the <local community>	0.12	-1.42	0.42	1.01
IC2G06F	Campaigns to raise people's awareness, such as <AIDS World Day, World No Tobacco Day>	-0.53	-1.03	0.20	0.82
IC2G06G	Activities related to improving facilities for the <local community> (for example, public gardens, libraries, health centers, recreation centers, community hall)	0.27	-1.21	0.51	0.70
<b>Resources</b>	<b>Are the following resources available in the local area where this school is located?</b>				
IC2G13A	Public library	-2.20	0.00		
IC2G13B	Cinema	0.70	0.00		
IC2G13C	Theater or concert hall	0.51	0.00		
IC2G13D	Language school	1.02	0.00		
IC2G13E	Museum or art gallery	0.45	0.00		
IC2G13G	Public garden or park	-2.12	0.00		
<b>Social Tension</b>	<b>To what extent are any of the following issues a source of social tension in the area in which this school is located?</b>				
IC2G14A	Immigration	1.05	-1.22	0.24	0.99
IC2G14B	Poor quality of housing	0.91	-1.60	0.25	1.35
IC2G14C	Unemployment	-0.13	-1.77	0.46	1.31
IC2G14D	Religious intolerance	2.23	-1.70	0.37	1.32
IC2G14E	Ethnic conflicts	2.36	-1.51	0.27	1.23
IC2G14F	Extensive poverty	0.84	-1.29	0.34	0.94
IC2G14G	Organized crime	1.75	-1.12	0.52	0.61
IC2G14H	Youth gangs	1.46	-1.52	0.43	1.08
IC2G14I	Petty crime	0.96	-2.39	0.75	1.63
IC2G14J	Sexual harassment	2.24	-1.74	0.74	1.01
IC2G14K	Drug abuse	1.18	-1.92	0.52	1.41
IC2G14L	Alcohol abuse	0.53	-2.07	0.21	1.85

Figure 12.19: Confirmatory factor analysis of items measuring principals' perceptions of the local community



### Principals' reports on school climate

Table 12.36 presents the reliabilities (Cronbach's alpha) for the two scales that were derived from questions regarding principals' reports on school climate. The two scales, which are included in the ICCS school database, were named:

- Principals' perceptions of students' behavior at school (CSTUDBEH);
- Principals' perceptions of social problems at school (CSCPROB).

When answering Question 11, principals were asked to state the number of students they thought behaved in ways listed in a series of items. Response categories were "all or nearly all," "most of them," "some of them," and "none or hardly any." The last two categories were collapsed for scaling. The four items associated with the question were used to form the scale *principals' perceptions of students' behavior at school* (CSTUDBEH). This scale had a reliability (Cronbach's alpha) of 0.86 for the pooled ICCS sample. The reliability estimates across participating countries ranged from 0.71 to 0.96 (see Table 12.36). Table 12.37 shows the item wording and the item parameters that were used for scaling. The higher values on this scale come from principals who thought that relatively high numbers of students exhibited the behaviors listed.

Question 12 required principals to indicate the frequency with which students at the school experienced specified social problems ("never," "sometimes," "often," "very often"). The last two categories were collapsed for scaling. The nine items associated with the question were used to form the scale *principals' perceptions of social problems at school* (CSCPROB). The higher scores on this scale indicate perceptions of a relatively high incidence of social problems. This scale had a reliability (Cronbach's alpha) of 0.80 for the pooled ICCS sample, and the country reliabilities ranged from 0.48 to 0.91 (see Table 12.36). The item parameters that were used for scaling are shown in Table 12.37.

The results of the confirmatory factor analysis showed an acceptable model fit (see Figure 12.20) and generally strong factor loadings. Item IC2G15B ("truancy") had somewhat lower item reliability than the other items measuring the scale CSCPROB ("social problems at school"). The two latent factors were negatively correlated at -0.42, a result which indicates that school principals who saw their students exhibiting generally positive behaviors tended to report fewer social problems at their schools.

### Principals' reports on sense of belonging to school

Three scales were derived from questions regarding principals' reports on the local community. The reliabilities of these scales are reported in Table 12.34. The scales, all of which are included in the ICCS school database, were named:

- Principals' perceptions of teachers' sense of belonging to school (TSCSBEL);
- Principals' perceptions of students' sense of belonging to school (SSCSBEL);
- Principals' perceptions of non-teaching staff's sense of belonging to school (NSCSBEL).

Question 12 of the school questionnaire asked principals to rate the extent to which a series of statements regarding sense of belonging to school applied to the teachers, the students, and the non-teaching staff at their school ("to a large extent," "to a moderate extent," "to a small extent," "not at all"). The last two categories were collapsed for scaling.

The first four items in the question related to teachers' sense of belonging, and all four were used to derive the scale *principals' perceptions of teachers' sense of belonging to school* (TSCSBEL). This scale had a reliability (Cronbach's alpha) of 0.83 for the pooled sample and reliabilities that ranged from 0.56 to 0.93 across the participating countries (see Table 12.38). Table 12.39



*Table 12.36: Reliabilities for scales reflecting principals' reports on school climate*

Country	Students' Behavior	Social Problems
Austria	0.79	0.70
Belgium (Flemish)	0.77	0.58
Bulgaria	0.83	0.78
Chile	0.85	0.87
Chinese Taipei	0.87	0.71
Colombia	0.78	0.78
Cyprus	0.84	0.80
Czech Republic	0.77	0.78
Denmark	0.86	0.81
Dominican Republic	0.87	0.84
England	0.93	0.72
Estonia	0.86	0.66
Finland	0.91	0.71
Greece	0.82	0.80
Guatemala	0.77	0.91
Hong Kong SAR	0.92	0.72
Indonesia	0.78	0.71
Ireland	0.89	0.80
Italy	0.89	0.75
Korea, Republic of	0.83	0.62
Latvia	0.74	0.75
Liechtenstein	0.89	0.48
Lithuania	0.89	0.78
Luxembourg	0.96	0.82
Malta	0.80	0.72
Mexico	0.81	0.82
Netherlands	0.84	0.48
New Zealand	0.86	0.83
Norway	0.86	0.74
Paraguay	0.77	0.81
Poland	0.71	0.79
Russian Federation	0.85	0.68
Slovak Republic	0.82	0.59
Slovenia	0.85	0.73
Spain	0.85	0.81
Sweden	0.91	0.84
Switzerland	0.84	0.63
Thailand	0.86	0.73
ICCS average	0.86	0.80

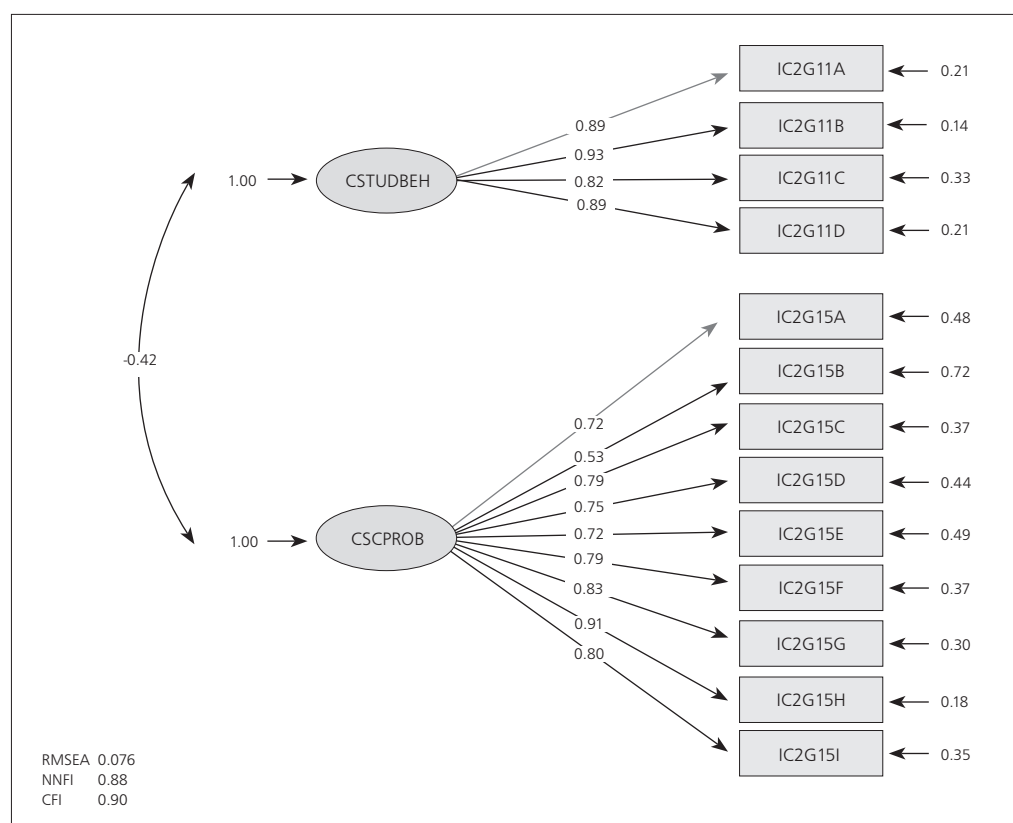
shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect perceptions that teachers had a relatively strong sense of belonging to the school.



Table 12.37: Item parameters for scales reflecting principals' reports on school climate

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Students' Behavior</b>	<b>In your opinion, how many students in this school ...</b>				
IC2G11A	are well behaved on entering and leaving the school premises?	-2.79	-4.85	4.85	
IC2G11B	adhere to school rules?	-2.60	-5.35	5.35	
IC2G11C	show care for school facilities and equipment?	-1.38	-4.57	4.57	
IC2G11D	are well behaved during breaks?	-1.98	-5.04	5.04	
<b>Social Problems</b>	<b>Please indicate how frequently each of the following problems occurs among students at this school</b>				
IC2G15A	Vandalism	1.30	-2.54	2.54	
IC2G15B	Truancy	-0.32	-2.81	2.81	
IC2G15C	Racism	2.71	-2.04	2.04	
IC2G15D	Religious intolerance	3.44	-1.76	1.76	
IC2G15E	Bullying	0.55	-2.44	2.44	
IC2G15F	Violence	1.24	-2.54	2.54	
IC2G15G	Sexual harassment	3.27	-2.02	2.02	
IC2G15H	Drug abuse	2.70	-1.89	1.89	
IC2G15I	Alcohol abuse	1.81	-1.70	1.70	

Figure 12.20: Confirmatory factor analysis of items measuring principals' perceptions of school climate



*Table 12.38: Reliabilities for scales reflecting principals' reports on sense of belonging to school*

Country	Teachers' Sense	Students' Sense	Non-Teaching Staff's Sense
Austria	0.85	0.72	0.81
Belgium (Flemish)	0.87	0.74	0.83
Bulgaria	0.56	0.83	0.81
Chile	0.86	0.87	0.89
Chinese Taipei	0.84	0.80	0.91
Colombia	0.93	0.84	0.93
Cyprus	0.84	0.85	0.90
Czech Republic	0.74	0.69	0.88
Denmark	0.72	0.74	0.85
Dominican Republic	0.79	0.81	0.90
England	0.78	0.73	0.92
Estonia	0.74	0.74	0.81
Finland	0.76	0.76	0.80
Greece	0.83	0.83	0.91
Guatemala	0.87	0.79	0.88
Hong Kong SAR	0.88	0.86	0.88
Indonesia	0.84	0.86	0.92
Ireland	0.86	0.82	0.86
Italy	0.84	0.67	0.92
Korea, Republic of	0.88	0.85	0.86
Latvia	0.74	0.74	0.80
Liechtenstein	0.73	0.72	0.32
Lithuania	0.80	0.80	0.81
Luxembourg	0.87	0.89	0.83
Malta	0.85	0.87	0.95
Mexico	0.86	0.82	0.97
Netherlands	0.83	0.77	0.83
New Zealand	0.82	0.84	0.86
Norway	0.69	0.79	0.40
Paraguay	0.84	0.83	0.93
Poland	0.78	0.73	0.87
Russian Federation	0.68	0.81	0.83
Slovak Republic	0.77	0.82	0.80
Slovenia	0.82	0.76	0.81
Spain	0.82	0.78	0.86
Sweden	0.85	0.86	0.88
Switzerland	0.73	0.75	0.79
Thailand	0.84	0.77	0.96
ICCS average	0.83	0.82	0.88



Table 12.39: Item parameters for scales reflecting principals' reports on sense of belonging to school

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Teachers' Sense</b>	In your opinion, to what extent do the following statements describe the current situation at this school?				
IC2G12A	The teachers have a positive attitude toward the school	-4.98	-2.82	2.82	
IC2G12B	The teachers feel they belong to the school community	-4.35	-2.63	2.63	
IC2G12C	Teachers work with enthusiasm	-3.14	-3.07	3.07	
IC2G12D	Teachers take pride in this school	-3.37	-2.69	2.69	
<b>Students' Sense</b>	In your opinion, to what extent do the following statements describe the current situation at this school?				
IC2G12E	Students enjoy being in school	-3.06	-2.69	2.69	
IC2G12F	Students work with enthusiasm	-0.75	-2.95	2.95	
IC2G12G	Students take pride in this school	-2.19	-2.47	2.47	
IC2G12H	Students feel part of the school community	-2.82	-2.47	2.47	
<b>Non-Teaching Staff's Sense</b>	In your opinion, to what extent do the following statements describe the current situation at this school?				
IC2G12I	Non-teaching staff feel part of the school community	-3.04	-2.31	2.31	
IC2G12J	Non-teaching staff care about how well the school operates	-3.86	-2.44	2.44	
IC2G12K	Non-teaching staff work with enthusiasm	-3.03	-2.84	2.84	

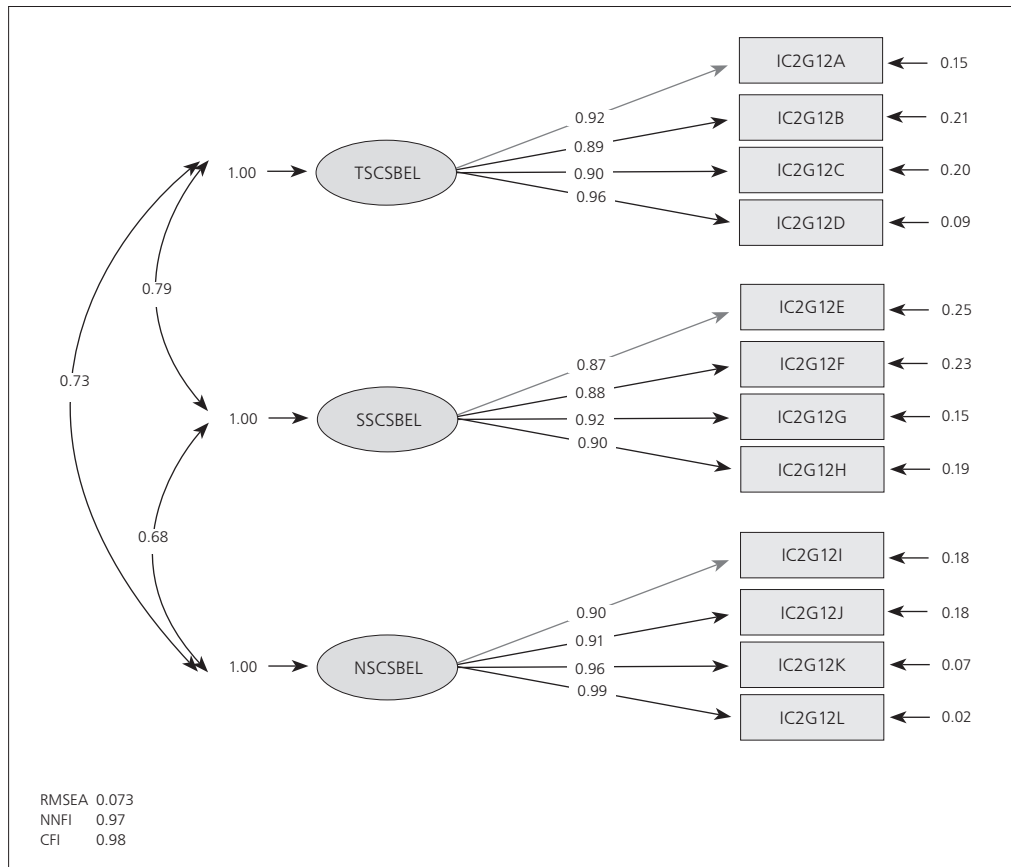
The next four items in Question 12 related to students' sense of belonging. These items were used to derive the scale *principals' perceptions of students' sense of belonging to school* (SSCSBEL). The higher scores correspond to perceptions of a greater sense of belonging. This scale had a reliability (Cronbach's alpha) of 0.82 for the pooled ICCS sample; the national sample reliabilities ranged from 0.67 to 0.89 (see Table 12.38). Table 12.39 shows the item parameters that were used for scaling.

The remaining four items from Question 12 related to non-teaching staff. These were used to form the scale *principals' perceptions of non-teaching staff's sense of belonging to school* (NSCSBEL). This scale had a reliability (Cronbach's alpha) of 0.88 for the pooled ICCS sample; reliabilities ranged from 0.32 to 0.97 across the participating countries (see Table 12.38). Table 12.39 shows the item parameters that were used for scaling.

Figure 12.21 shows the results of the confirmatory factory analysis of this item set. The fit indices suggested a reasonable model fit, while the size of the factor loadings indicated that the items provided a good measure of the underlying latent dimensions. Positive correlations were found between the three latent factors; estimates ranged from 0.68 to 0.79.



Figure 12.21: Confirmatory factor analysis of items measuring principals' perceptions of sense of belonging to school



### European questionnaire

#### Students' perceptions of European identity

Question 1 of the European regional questionnaire asked students to indicate their level of agreement with a series of statements about how they saw themselves in relation to Europe ("strongly agree," "agree," "disagree," "strongly disagree"). Five of the question items were used to construct the scale *students' sense of European identity* (EUIDENT). This scale had a reliability (Cronbach's alpha) of 0.74 for the pooled ICCS sample, and the country reliabilities ranged from 0.68 to 0.80 (see Table 12.40). Table 12.41 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect a relatively strong sense of European identity.

Figure 12.22 shows the results of the confirmatory factor analysis of this item set. The one-factor solution had a good model fit. The factor loadings for Item ES2P01F ("more in common with young people from European countries") were considerably lower than those of the other items, which suggests that this item provided, relative to the others, a less successful measure of the construct.





Table 12.40: Reliabilities for scale reflecting students' perceptions of European identity

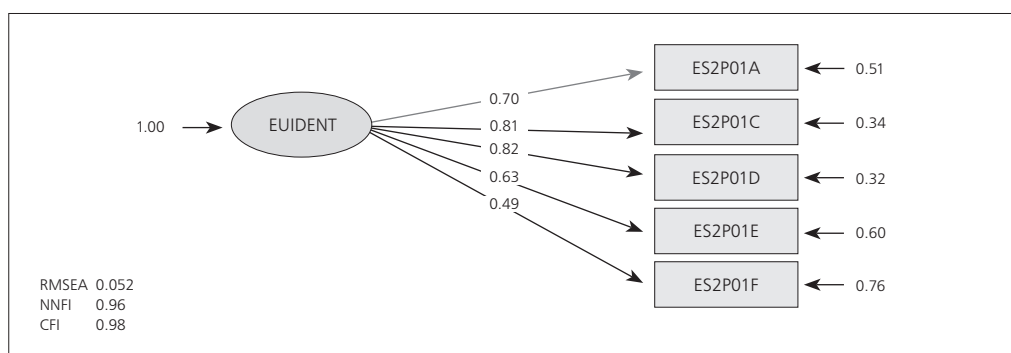
Country	European Identity
Austria	0.72
Belgium (Flemish)	0.72
Bulgaria	0.74
Cyprus	0.77
Czech Republic	0.76
Denmark	0.69
England	0.80
Estonia	0.75
Finland	0.77
Greece	0.69
Ireland	0.75
Italy	0.70
Latvia	0.72
Liechtenstein	0.72
Lithuania	0.71
Luxembourg	0.68
Malta	0.77
Netherlands	0.71
Poland	0.72
Slovak Republic	0.73
Slovenia	0.74
Spain	0.75
Sweden	0.77
Switzerland	0.71
ICCS average	0.74

Table 12.41: Item parameters for scale reflecting students' perceptions of European identity

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
	We would like to find out about how you see yourself.				
<b>Students' Perceptions</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P01A	I see myself as European.	-1.90	-1.13	-0.80	1.92
ES2P01C	I am proud to live in Europe.	-1.73	-1.44	-1.10	2.53
ES2P01D	I feel part of Europe.	-1.09	-1.97	-0.50	2.46
ES2P01E	I see myself first as a citizen of Europe and then as a citizen of the world.	-0.73	-1.92	-0.24	2.16
ES2P01F	I have more in common with young people from European countries than with those from countries outside Europe.	-0.52	-1.91	-0.25	2.15



Figure 12.22: Confirmatory factor analysis of items measuring students' sense of European identity



### Students' reports on activities related to Europe

Three scales were derived from questions regarding students' reports on activities related to Europe. The reliabilities of these scales, which are included in the European ICCS student database and listed as follows, are reported in Table 12.38.

- Students' participation in activities or groups at the European level (EUPART);
- Students' reports on opportunities for learning about Europe at school (EUROPP);
- Students' participation in communication about Europe (EUROCOM).

Question 2 of the European regional questionnaire required students to indicate whether or not they had participated in a range of activities that involved another European country. The response categories were "Yes, I have done this within the last 12 months," "Yes, I have done this but more than a year ago," and "No, I have never done this." The question's eight items were used to form the scale *students' participation in activities or groups at the European level* (EUPART). The higher scores on the scale correspond to greater participation levels, and the scale's reliability was 0.73 for the pooled ICCS sample. The national reliabilities ranged from 0.63 to 0.83 (see Table 12.42). Table 12.43 shows the item parameters that were used for scaling.

Question 3 asked students to indicate their level of agreement with a series of statements about opportunities for learning about Europe at school ("strongly agree," "agree," "disagree," "strongly disagree"). The nine items associated with the question were used to derive the scale *students' reports on opportunities for learning about Europe at school* (EUROPP). Higher scores on this scale correspond to greater amounts of opportunity. The scale reliability (Cronbach's alpha) was 0.83 for the European sample; country reliabilities ranged from 0.78 to 0.86 (see Table 12.42). The item parameters that were used for scaling are shown in Table 12.43.

The items relating to Question 4 cited a range of activities relating to communications (e.g., media-based) about Europe. Students were asked to indicate how often they were involved in each of these activities; response categories were "never or hardly ever," "yearly (at least once a year)," "monthly (at least once a month)," and "weekly (at least once a week)." These items were used to construct the scale *students' participation in communication about Europe* (EUROCOM). The scale reliability (Cronbach's alpha) was 0.85 for the European sample; the national reliabilities ranged from 0.81 to 0.87 (see Table 12.42). Table 12.43 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect greater communication concerning European issues.



Table 12.42: Reliabilities for scales reflecting students' reports on activities related to Europe

Country	Particip. in Groups/Activities	Learning Opps.	Particip. in Communication
Austria	0.72	0.83	0.85
Belgium (Flemish)	0.68	0.83	0.84
Bulgaria	0.83	N/A	0.84
Cyprus	0.80	0.85	0.85
Czech Republic	0.71	0.82	0.84
Denmark	0.64	0.78	0.85
England	0.73	0.83	0.87
Estonia	0.71	0.84	0.85
Finland	0.67	0.85	0.87
Greece	0.82	0.82	0.81
Ireland	0.70	0.84	0.84
Italy	0.69	0.79	0.82
Latvia	0.74	0.78	0.83
Liechtenstein	0.69	0.86	0.85
Lithuania	0.73	0.81	0.83
Luxembourg	0.71	0.84	0.85
Malta	0.75	0.84	0.82
Netherlands	0.63	0.78	0.83
Poland	0.77	0.85	0.86
Slovak Republic	0.72	0.82	0.85
Slovenia	0.73	0.85	0.85
Spain	0.77	0.85	0.82
Sweden	0.73	0.85	0.86
Switzerland	0.65	0.80	0.83
ICCS average	0.73	0.83	0.85

Figure 12.23 presents the results of the confirmatory factor analysis of these three item sets. The RMSEA suggested a good fit for the model, but the NNFI and CFI indicated a considerable lack of fit. Most factor loadings indicated high measurement reliability at the item level. The correlations between the three latent factors were positive and ranged from 0.29 to 0.41.

### Students' attitudes toward learning of European languages

Students were asked in Question 7 to indicate their level of agreement with a series of statements concerning the learning of languages spoken in other European countries. The response categories were "strongly agree," "agree," "disagree," "strongly disagree." The six items associated with the question were used to construct the scale *students' attitudes toward European language learning* (EUATLANG). Higher EUATLANG scores on the scale denote greater agreement with the notion that learning other European languages is important. This scale had a reliability (Cronbach's alpha) of 0.82 for the pooled ICCS sample, and the national reliabilities ranged from 0.77 to 0.85 (see Table 12.44). Table 12.45 shows the item parameters that were used for scaling.



Table 12.43: Item parameters for scales reflecting students' reports on activities related to Europe

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Particip. in Groups/Activities</b>	<b>Have you ever participated in any of the following activities?</b>				
ES2P02A	Activities organised in my local area that involve meeting people from other European countries	1.14	0.02	-0.02	
ES2P02B	Activities related to friendship agreements (twinning) between my local town/city and other European towns/cities	1.48	-0.16	0.16	
ES2P02C	Music, dance or film festival(s) in another European country	1.13	0.00	0.00	
ES2P02D	Sports event(s) in another European country	1.01	0.13	-0.13	
ES2P02F	Exchange programs with students from other European countries (going abroad or others coming to your country)	1.41	0.36	-0.36	
ES2P02G	School trip(s) to another European country	0.92	0.23	-0.23	
ES2P02H	Visits to other European countries for leisure/holidays	-0.41	-0.08	0.08	
ES2P02I	Exhibitions, festivals, or other events about the art and culture (e.g. music, films) of other European countries	0.70	-0.12	0.12	
	How much do you agree or disagree with the following statements?				
<b>Learning Opps.</b>	<b>My school gives me opportunities to ...</b>				
ES2P03A	visit other European countries.	-0.18	-1.07	-0.26	1.33
ES2P03B	meet young people from other European countries.	-0.03	-1.61	-0.20	1.82
ES2P03C	learn about political and economic issues in other European countries.	0.24	-1.93	-0.60	2.53
ES2P03D	find out what is happening in other European countries.	-0.55	-1.80	-0.82	2.63
ES2P03E	find out about other European countries through the internet or the media (press, TV, or radio).	-0.56	-1.71	-0.57	2.27
ES2P03F	learn about arts and culture (e.g., music, films) in other European countries.	-0.65	-1.66	-0.71	2.36
ES2P03G	learn about sport in other European countries.	-0.50	-1.85	-0.17	2.01
ES2P03H	find out what it is like to live in other European countries.	-0.46	-1.78	-0.46	2.24
ES2P03I	learn about how I could work in other European countries.	0.00	-1.89	-0.08	1.96
<b>Particip. in Communication</b>	<b>How often are you involved in each of the following activities?</b>				
ES2P04A	Watching television to inform yourself about European news	-0.39	0.13	-0.48	0.35
ES2P04B	Reading the newspapers to inform yourself about European news	0.08	-0.08	-0.61	0.69
ES2P04C	Discussing the political or economic situation in other European countries with your friends or family	0.65	-0.39	-0.55	0.95
ES2P04D	Discussing European sports events with your friends or family	-0.10	-0.25	-0.23	0.49
ES2P04E	Discussing arts and culture (e.g. music, films) from other European countries with your friends or family	0.22	-0.37	-0.38	0.74
ES2P04F	Discussing the European Union with your friends or family	0.95	-0.64	-0.36	1.01
ES2P04G	Discussing issues raised in the European Parliament with your friends or family	1.27	-0.44	-0.32	0.76
ES2P04H	Talking about what life is like in other European countries with your friends or family	0.42	-0.83	-0.25	1.09
ES2P04I	Talking, with your friends and family, about what it might be like to work in other European countries	0.54	-0.66	-0.26	0.92

Figure 12.23: Confirmatory factor analysis of items reflecting students' reports on activities related to Europe

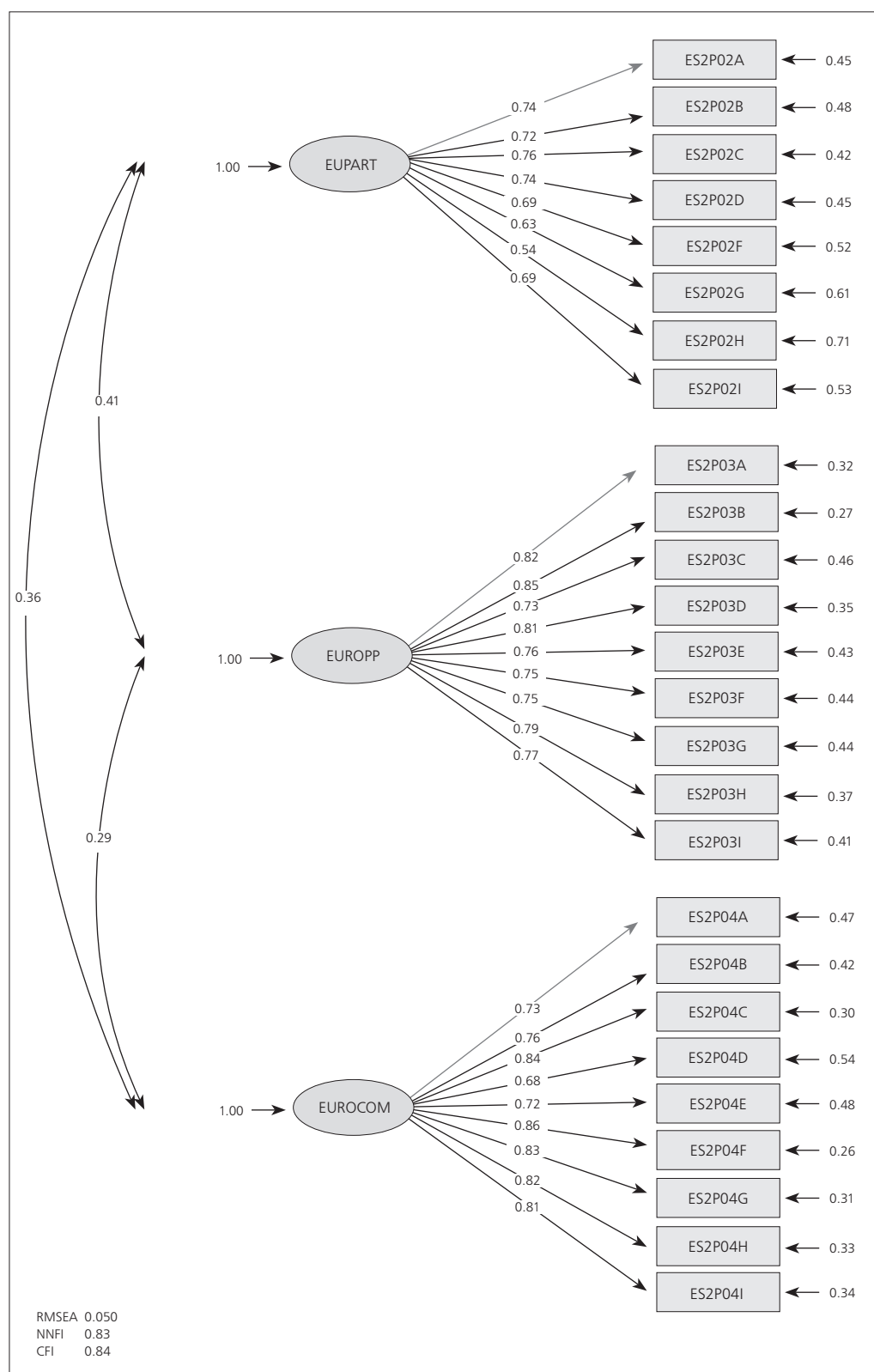


Table 12.44: Reliabilities for scale reflecting students' attitudes toward learning other European languages

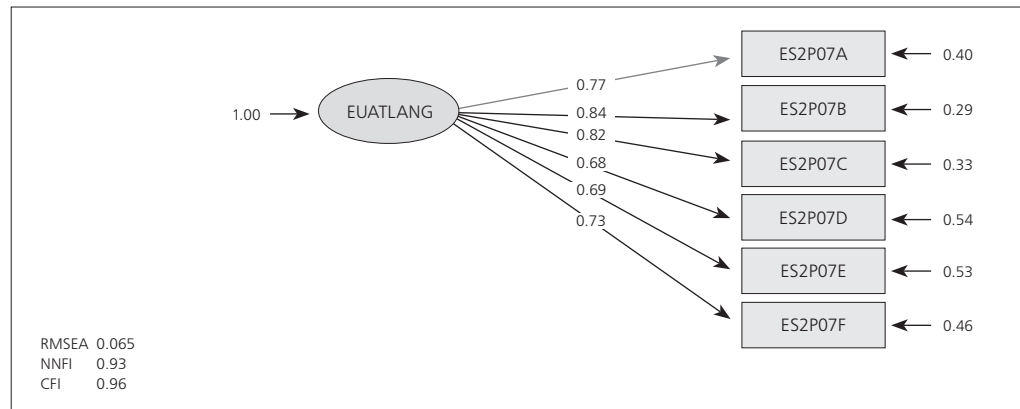
Country	Students' Attitudes
Austria	0.78
Belgium (Flemish)	0.80
Bulgaria	0.85
Cyprus	0.83
Czech Republic	0.80
Denmark	0.78
England	0.85
Estonia	0.78
Finland	0.81
Greece	0.77
Ireland	0.82
Italy	0.79
Latvia	0.78
Liechtenstein	0.83
Lithuania	0.79
Luxembourg	0.80
Malta	0.80
Netherlands	0.79
Poland	0.83
Slovak Republic	0.80
Slovenia	0.79
Spain	0.82
Sweden	0.84
Switzerland	0.81
ICCS average	0.82

Table 12.45: Item parameters for scale reflecting students' attitudes toward learning other European languages

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
	We would like to know what you think about learning languages spoken in other European countries.				
<b>Students' Attitudes</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P07A	Learning a foreign European language is important for travelling/going on holidays in Europe.	-2.42	-1.27	-0.90	2.17
ES2P07B	Learning a foreign European language can make it easier to find a job.	-2.46	-1.55	-0.84	2.39
ES2P07C	Learning a foreign European language is important for working or studying in another European country.	-2.49	-1.26	-0.86	2.13
ES2P07D	Learning a foreign European language helps people understand other European cultures better.	-1.56	-2.14	-0.51	2.65
ES2P07E	All young people in Europe should learn at least two foreign European languages.	-1.22	-2.06	-0.19	2.25
ES2P07F	Schools should give young people more opportunity to learn foreign languages used in other European countries.	-1.63	-1.64	-0.59	2.23

Figure 12.24 shows the results of the confirmatory factory analysis for this item set. The one-factor solution had an acceptable model fit, but there was some variation in the strength of factor loadings across the six items.

Figure 12.24: Confirmatory factor analysis of items measuring students' attitudes toward learning European languages



### Students' attitudes toward migration within Europe

The three scales that were derived from questions regarding students' attitudes toward migration within Europe and that are included in the European ICCS student database are as follows. The scales' reliabilities are reported in Table 12.40.

- Students' attitudes toward freedom of migration within Europe (EUMOVE);
- Students' attitudes toward restricting migration within Europe (EURESTR);
- Students' attitudes toward equal opportunities for other European citizens (EUCITOPP).

Question 8 of the European regional student questionnaire asked students to indicate their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with a range of statements about citizens of European countries travelling in Europe or moving to live in another European country. Four of the statements concerned freedom of individuals to live and work in their choice of European countries. These were used to construct the scale *students' attitudes toward freedom of migration within Europe* (EUMOVE). The higher scores on the EUMOVE scale correspond to more positive attitudes toward freedom of movement within Europe. The scale reliability was 0.63 for the pooled ICCS sample, and the national reliabilities ranged from 0.51 to 0.71 (see Table 12.46). Table 12.47 shows the item parameters used for scaling.

Five items in Question 8 contained statements in favor of restricting freedom of movement within European countries. These were used to derive the scale *students' attitudes toward restricting migration within Europe* (EURESTR). Higher scores on this scale indicate favorable attitudes toward restricting migration within Europe. The scale reliability was 0.68 for the European sample; national reliabilities ranged from 0.59 to 0.74 (see Table 12.46). The item parameters used for scaling are shown in Table 12.47.

Question 9 contained a series of statements about the opportunities that citizens of European countries should have in the country where the survey was undertaken. Students were asked to state their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with each statement. The five items associated with the question were used to derive a scale reflecting *students' attitudes toward equal opportunities for other European citizens* (EUCITOPP). The



Table 12.46: Reliabilities for scales reflecting students' attitudes toward migration within Europe

Country	Freedom of Movement	Restricting Migration	Equal Opportunities
Austria	0.62	0.66	0.85
Belgium (Flemish)	0.62	0.66	0.85
Bulgaria	0.63	0.70	0.81
Cyprus	0.58	0.64	0.82
Czech Republic	0.60	0.66	0.80
Denmark	0.65	0.71	0.87
England	0.71	0.72	0.89
Estonia	0.59	0.67	0.77
Finland	0.71	0.73	0.89
Greece	0.51	0.66	0.82
Ireland	0.69	0.66	0.86
Italy	0.64	0.66	0.89
Latvia	0.53	0.59	0.72
Liechtenstein	0.70	0.67	0.88
Lithuania	0.54	0.67	0.82
Luxembourg	0.59	0.64	0.83
Malta	0.57	0.66	0.79
Netherlands	0.64	0.66	0.80
Poland	0.59	0.73	0.85
Slovak Republic	0.58	0.69	0.85
Slovenia	0.60	0.70	0.83
Spain	0.60	0.68	0.86
Sweden	0.71	0.74	0.91
Switzerland	0.68	0.67	0.88
ICCS average	0.63	0.68	0.85

scale reliability (Cronbach's alpha) was 0.85 for the pooled ICCS sample, and the national reliabilities ranged from 0.72 to 0.91 (see Table 12.46). Table 12.47 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect more favorable attitudes toward equal opportunities for all European citizens.

Figure 12.25 shows the results of the confirmatory factor analysis for these three item sets. The three-factor solution had a good model fit, but the factor loadings for EUMOVE (reflecting positive attitudes toward freedom of movement) and EURESTR (reflecting attitudes in favor of restricting migration within Europe) tended to be rather low, an outcome that coincided with the relatively low reliabilities for these scales. EUMOVE and EUCITOPP (reflecting positive attitudes toward equal opportunities for other European citizens in the country) were positively correlated at 0.71. These two latent factors (in favor of freedom of movement) had weak negative correlations with EURESTR.



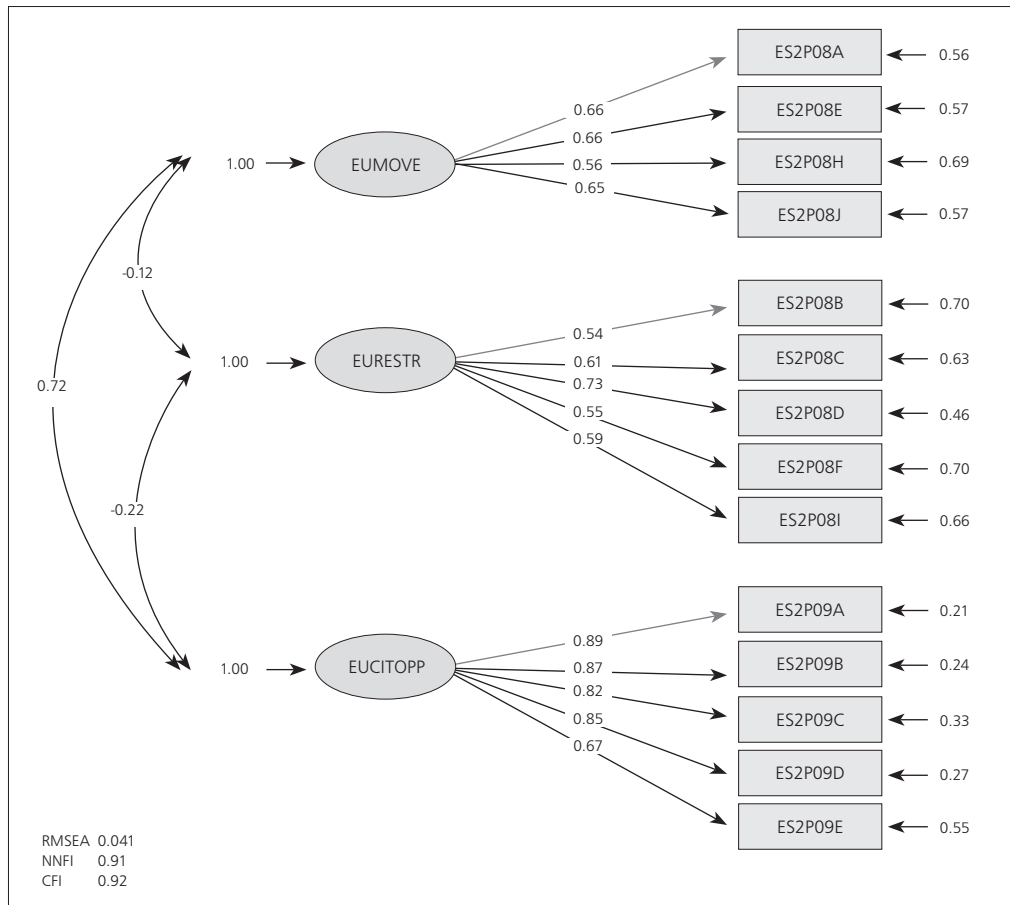


Table 12.47: Item parameters for scales reflecting students' attitudes toward migration within Europe

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
	Here are some statements about citizens of European countries travelling in Europe or moving home to another European country (i.e. becoming <immigrants> there).				
<b>Freedom of Movement</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P08A	Citizens of European countries should be allowed to live and work anywhere in Europe.	-1.66	-1.48	-0.77	2.25
ES2P08E	Other Europeans being allowed to live in <country of test> is good because they bring different cultures with them.	-0.84	-1.66	-0.64	2.30
ES2P08H	Allowing citizens from other European countries to work here is good for the economy of <country of test>.	-0.73	-2.16	-0.27	2.44
ES2P08J	European citizens should be free to travel anywhere in Europe, so they get to understand other European cultures better.	-1.45	-1.26	-0.84	2.09
<b>Restricting Migration</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P08B	The travel of European citizens in Europe should be more restricted to help fight terrorism.	-0.11	-2.00	0.01	2.00
ES2P08C	Other Europeans living in <country of test> leads to conflict and hostility between people of different nationalities.	-0.09	-2.20	0.22	1.99
ES2P08D	Citizens of <country of test> will be safer from crime if they close their borders to <immigrants> from other European countries.	0.04	-1.85	0.22	1.62
ES2P08F	Allowing citizens of other European countries to come and work here leads to more unemployment for citizens of <country of test>.	-0.59	-1.91	-0.11	2.03
ES2P08I	The movement of workers between European countries should be restricted, otherwise some countries will be full of <immigrants>.	-0.55	-2.15	-0.16	2.32
	Below are some statements about the opportunities which citizens from European countries should have in <country of test>.				
	<b>How much do you agree or disagree with the following statements?</b>				
<b>Equal Opportunities</b>	<b>Citizens of European countries who come to &lt;country of test&gt; should have the same opportunities as people from &lt;country of test&gt; ...</b>				
ES2P09A	whatever their ethnic or racial background.	-1.89	-2.17	-0.70	2.86
ES2P09B	whatever their religion or beliefs.	-1.87	-2.26	-0.75	3.00
ES2P09C	whatever language they speak.	-1.50	-2.55	-0.36	2.90
ES2P09D	whether they come from a rich country or a poor one.	-1.98	-1.71	-1.14	2.84
ES2P09E	whatever their level of education.	-1.01	-2.53	-0.36	2.88



Figure 12.25: Confirmatory factor analysis of items measuring students' attitudes toward migration within Europe



### Students' attitudes toward European integration

Four scales were derived from questions regarding students' attitudes toward European integration. Table 12.48 reports the reliabilities of the scales, all four of which are included in the European ICCS student database.

- Students' attitudes toward common policies in Europe (EUCOMPOL);
- Students' attitudes toward European unification (EURUNION);
- Students' attitudes toward common European currency (EUCURR);
- Students' attitudes toward further expansion of the European Union (EUSIZE).

Question 10 asked students to state how much they agreed ("strongly agree," "agree," "disagree," "strongly disagree") with a range of statements about how European countries should be organized. Four of these statements, each of which related to common policies across European countries, were used to derive the scale *students' attitudes toward common policies in Europe* (EUCOMPOL). The scale also included one item (ES2P11A) from Question 11 ("all European countries should have the same economic policies"). The reliability of this scale was 0.63 for the pooled ICCS sample. The national reliabilities ranged from 0.56 to 0.67 (see Table 12.48). Table 12.49 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale denote greater agreement with the idea that European countries should have common policies.



Table 12.48: Reliabilities for scales reflecting students' attitudes toward European integration

Country	Common Policies	Unification	Common Currency	EU
Austria	0.63	0.65	0.67	0.75
Belgium (Flemish)	0.61	0.65	0.66	0.80
Bulgaria	0.67	0.72	0.69	0.80
Cyprus	0.61	0.63	0.62	0.74
Czech Republic	0.60	0.76	0.75	0.80
Denmark	0.60	0.72	0.76	0.76
England	0.66	0.78	0.72	0.80
Estonia	0.62	0.79	0.74	0.76
Finland	0.64	0.80	0.73	0.81
Greece	0.56	0.64	0.59	0.69
Ireland	0.63	0.75	0.67	0.79
Italy	0.59	0.67	0.64	0.74
Latvia	0.57	0.63	0.70	0.73
Liechtenstein	0.66	0.73	0.79	0.82
Lithuania	0.66	0.68	0.70	0.77
Luxembourg	0.65	0.66	0.64	0.75
Malta	0.63	0.67	0.68	0.80
Netherlands	0.59	0.66	0.68	0.72
Poland	0.63	0.77	0.76	0.78
Slovak Republic	0.62	0.75	0.73	0.79
Slovenia	0.59	0.68	0.65	0.77
Spain	0.62	0.70	0.68	0.78
Sweden	0.66	0.75	0.80	0.82
Switzerland	0.63	0.76	0.75	0.79
ICCS average	0.63	0.73	0.72	0.78

Three items in Question 10 were designed to measure students' perceptions of European unification. These items were used to form the scale *students' attitudes toward European unification* (EURUNION). The scale reliability was 0.73 for the pooled ICCS sample. Reliabilities ranged in size from 0.63 to 0.80 across the participating countries (see Table 12.48). The item parameters that were used for scaling are shown in Table 12.49.

Question 11 included statements about economies and currencies within European countries. Students were asked to indicate their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with each one. Three of the items consisted of statements related to common currencies and the euro. These were used to derive the scale *students' attitudes toward common European currency* (EUCURR). The scale reliability was 0.72 for the combined European ICCS sample. The national reliabilities ranged from 0.59 to 0.80 (see Table 12.48). Table 12.49 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect greater agreement with the notion that European countries should have a shared currency.



Table 12.49: Item parameters for scales reflecting students' attitudes toward European integration

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
	Here are some statements about European countries and how they should be organized.				
<b>Common Policies</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P10B	All European countries should have the same approach to their relationships with countries outside Europe.	-0.85	-2.24	-0.61	2.85
ES2P10C	European countries should try to have a common set of policies regarding the environment.	-1.31	-1.55	-0.82	2.36
ES2P10D	European countries should try to have similar education systems.	-0.94	-1.68	-0.61	2.29
ES2P10F	It would be good if European countries had more similar rules and laws.	-0.76	-1.75	-0.66	2.41
ES2P11A	All European countries should have the same economic policies.	-0.81	-2.39	-0.04	2.44
<b>Unification</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P10E	The heads of state of European countries (<presidents, kings, queens, etc.>) should one day be replaced by a "president" of all Europe.	0.64	-2.00	0.36	1.64
ES2P10H	When countries join the European Union, they should give up their individual governments.	0.60	-2.58	0.43	2.14
ES2P10I	The European Parliament should one day replace the parliaments of all European countries.	0.46	-2.39	0.34	2.06
	Below are some statements about the opportunities which citizens from European countries should have in <country of test>.				
	Here are some more statements about European countries, their economies and their currencies (<money they use>). Some statements refer to the euro, which is used in several European Union countries and is therefore known as a "common currency."				
<b>Common Currency</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P11B	If all European countries had the same currency, they would be economically stronger.	-1.36	-2.85	-0.27	3.12
ES2P11C	There are more advantages to joining a common currency, such as the euro, than there are disadvantages.	-1.32	-2.62	-0.43	3.04
ES2P11D	All countries in Europe should join the euro.	-0.79	-1.96	-0.26	2.21
	Here are some more statements about the European Union and its enlargement (the increase in the number of countries that are members of the European Union).				
<b>Restrict Expansion</b>	<b>How much do you agree or disagree with the following statements?</b>				
ES2P12A	The European Union should continue to enlarge until it includes all European countries.	-0.89	-2.25	-0.16	2.41
ES2P12B	The European Union should be enlarged so more countries can benefit from the economic advantages it brings.	-1.00	-2.34	-0.60	2.95
ES2P12C	All countries in Europe should aspire to become members of the European Union.	-0.70	-2.34	-0.15	2.48
ES2P12D	The advantage of European Union enlargement is that it encourages countries that want to join to be democratic.	-0.88	-2.13	-0.64	2.76
ES2P12E	The European Union will have greater influence in the world if more countries join it.	-1.15	-1.99	-0.61	2.60
ES2P12F	The European Union needs to include all European countries to be a worthwhile organization.	-0.43	-2.39	-0.01	2.40
ES2P12G	The advantage of European Union enlargement is that it encourages countries that want to join to respect human rights.	-1.18	-1.52	-0.94	2.45

All of the items associated with Question 12 were used to form the scale *students' attitudes toward further expansion of the European Union* (EUSIZE). The items contained a series of statements about the European Union and its enlargement, and students were asked to indicate their level of agreement with them (response categories were “strongly agree,” “agree,” “disagree,” “strongly disagree”). The higher scores on the scale signify agreement with further expansion. The scale reliability was 0.78 for the pooled ICCS sample. The national reliabilities ranged from 0.69 to 0.82 (see Table 12.48). Table 12.49 shows the item parameters that were used for scaling.

Figure 12.26 shows the results of the confirmatory factor analysis for these item sets. The RMSEA for the four-factor solution indicated a close model fit. However, both the NNFI and CFI suggested that the data did not entirely fit the model. With the exception of the items for EUCURR (common European currency), factor loadings were not consistently high. The reliabilities for the items measuring EUCOMPOL (common policies in Europe) appeared to be relatively low, an outcome that was also reflected in the overall scale reliability of only 0.63. All latent factors were positively correlated with one another. The coefficients ranged from 0.42 to 0.76. The correlation at the low end of the range (i.e., 0.42) was between EURUNION (attitudes toward European unification) and EUCURR (common European currency). The correlation at the high end of the range (i.e., 0.76) was between EUCOMPOL (common policies in Europe) and EUSIZE (expanding the size of the EU).

### **Students' self-reported knowledge about the European Union**

Question 13 asked students how much they thought they knew about topics related to the European Union (“a lot,” “quite a lot,” “a little,” “nothing”). The four question items were used to form the scale *students' self-reported knowledge about the European Union* (EUKNOW). This scale had a reliability (Cronbach's alpha) of 0.78 for the European sample; the national reliabilities ranged from 0.70 to 0.84 (see Table 12.50). Table 12.51 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect higher levels of self-reported knowledge about topics related to the European Union.

The results of the confirmatory factor analysis of these items (see Figure 12.27) showed that the one-factor solution fitted the data. However, the relatively low factor loading for Item ES2P13D (“knowledge about the euro”) indicated that it did not measure the underlying latent trait as well as the other items did.



Figure 12.26: Confirmatory factor analysis of items measuring students' attitudes toward European integration

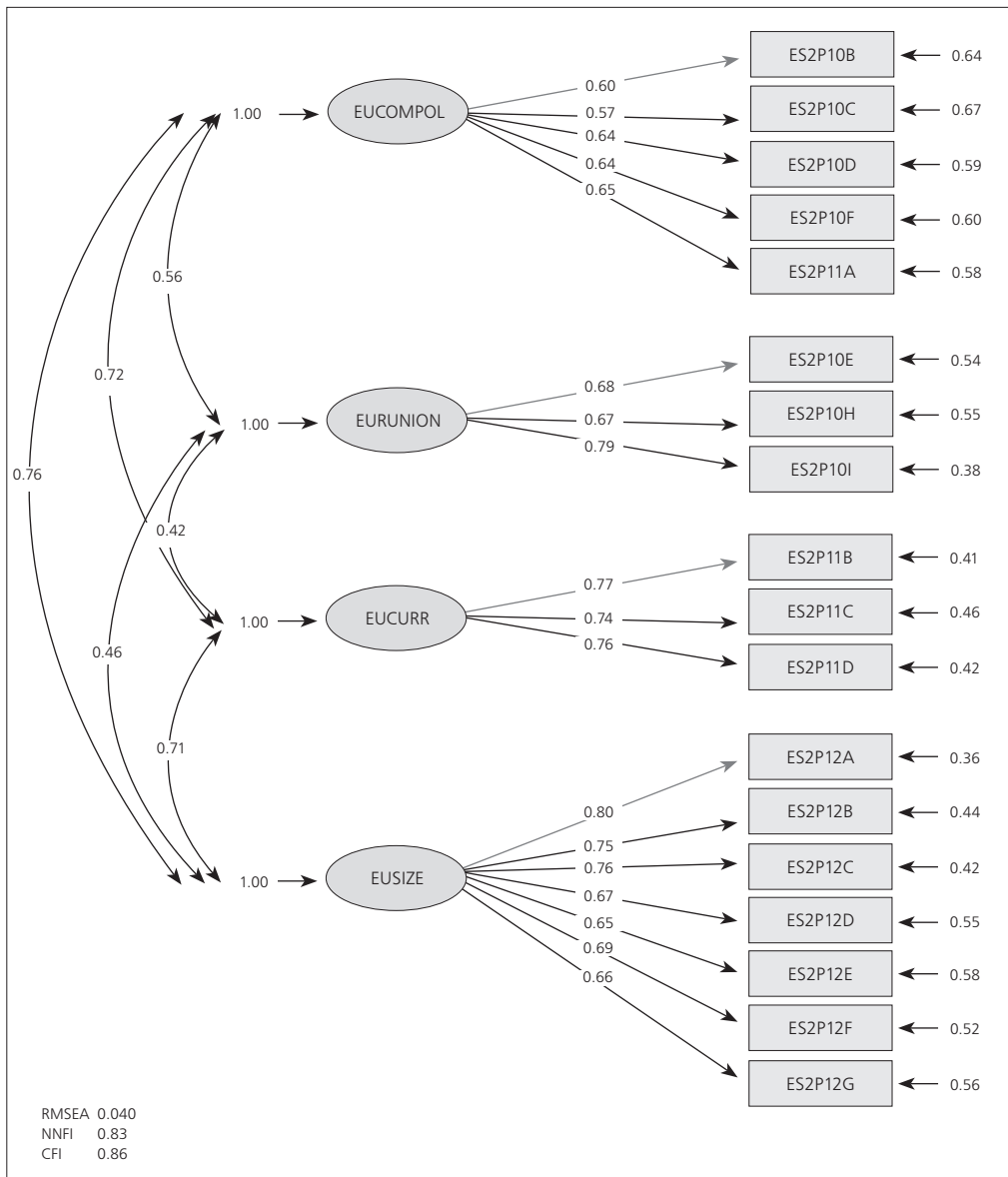


Table 12.50: Reliabilities for scale reflecting students' self-reported knowledge about the European Union

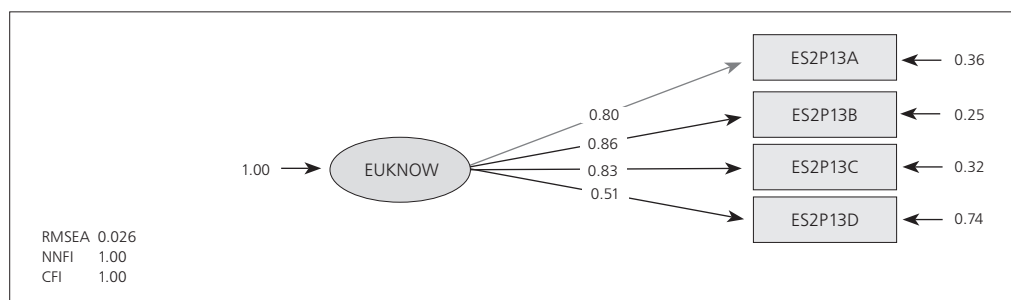
Country	Student's Knowledge
Austria	0.73
Belgium (Flemish)	0.80
Bulgaria	0.78
Cyprus	0.70
Czech Republic	0.76
Denmark	0.81
England	0.81
Estonia	0.74
Finland	0.79
Greece	0.72
Ireland	0.79
Italy	0.74
Latvia	0.72
Liechtenstein	0.78
Lithuania	0.71
Luxembourg	0.79
Malta	0.75
Netherlands	0.75
Poland	0.79
Slovak Republic	0.76
Slovenia	0.75
Spain	0.76
Sweden	0.84
Switzerland	0.79
ICCS average	0.78

Table 12.51: Item parameters for scale reflecting students' self-reported knowledge about the European Union

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
Students' Knowledge	How much do you know about the following topics?				
ES2P13A	Facts about the European Union	0.21	-2.93	0.44	2.50
ES2P13B	Laws and policies of the European Union	0.82	-3.02	0.49	2.54
ES2P13C	Institutions of the European Union (e.g. European Parliament)	1.10	-2.63	0.52	2.10
ES2P13D	The euro (the currency of some European Union countries)	-1.06	-2.30	-0.05	2.35



Figure 12.27: Confirmatory factor analysis of items measuring students' self-reported knowledge about the European Union



### Latin American questionnaire

#### Students' perceptions of Latin American identity

Question 1 of the Latin American regional questionnaire was designed to measure the extent to which students identified with the Latin American region. Students were asked to give their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with statements contained in five items. The five items were used to derive the scale *students' sense of Latin American identity* (LAIDENT). The scale reliability was 0.62 for the pooled ICCS sample. The national reliabilities ranged from 0.52 to 0.68 (see Table 12.44). Table 12.45 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect a greater sense of Latin American identity.

Table 12.52: Reliabilities for scale reflecting students' perceptions of Latin American identity

Country	Latin American Identity
Chile	0.68
Colombia	0.65
Dominican Republic	0.52
Guatemala	0.62
Mexico	0.64
Paraguay	0.57
ICCS average	0.62

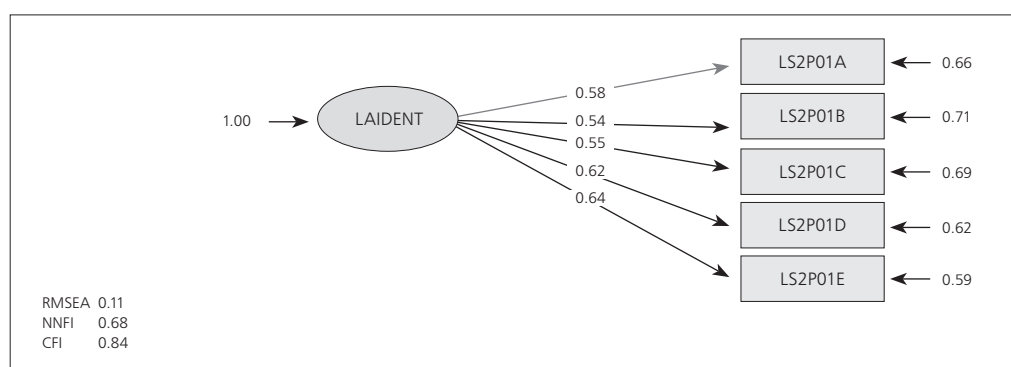
Table 12.53: Item parameters for scale reflecting students' perceptions of Latin American identity

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
Latin American Identity	How much do you agree or disagree with the following statements about Latin America and its people?				
LS2P01A	We Latin Americans have a lot in common even if we come from different countries.	-1.47	-1.25	-0.94	2.18
LS2P01B	In Latin America more things unite us than separate us.	-1.11	-1.95	-0.44	2.38
LS2P01C	I feel I have a lot in common with other Latin American youths.	-0.90	-1.66	-0.31	1.96
LS2P01D	Sometimes I support teams from other Latin American countries during international competitions.	-0.66	-0.98	-0.36	1.34
LS2P01E	I often support teams from other Latin American countries when my country has been eliminated from a competition.	-0.49	-0.79	-0.39	1.19



Figure 12.28 shows the results of the confirmatory factor analysis of this item set. The one-factor solution had only a poor model fit, and the factor loadings indicated that the item's ability to measure the underlying latent trait was weak; for all items, the latent factor failed to explain over 50 percent of the variance.

Figure 12.28: Confirmatory factor analysis of items measuring students' sense of Latin American identity



### Students' perceptions of government and law

Three scales were derived from questions regarding students' perceptions of government and law. Table 12.46 reports the reliabilities of these scales, each of which is included in the Latin American ICCS student database.

- Students' attitudes toward authoritarianism in government (AUTGOV);
- Students' attitudes toward corrupt practices in government (ATTCORR);
- Students' attitudes toward disobeying the law (DISLAW).

Questions 2 and 3 contained statements about government and its leaders, or the government's power. Students were asked to indicate their level of agreement with the statement contained in each item ("strongly agree," "agree," "disagree," or "strongly disagree"). All the items from Question 2 and the first three items from Question 3 were used to construct the scale *students' attitudes toward authoritarianism in government* (AUTGOV). Higher AUTGOV scores correspond to greater acceptance of governments engaging in authoritarian practices. The scale reliability was 0.83 for the Latin American sample; the national reliabilities ranged from 0.80 to 0.86 (see Table 12.54). The item parameters that were used for scaling are shown in Table 12.55.

Table 12.54: Reliabilities for scales reflecting students' perceptions of government and law in Latin America

Country	Authoritarian Govt.	Corruption	Disobeying Law
Chile	0.84	0.83	0.85
Colombia	0.81	0.80	0.84
Dominican Republic	0.82	0.77	0.80
Guatemala	0.80	0.82	0.80
Mexico	0.86	0.84	0.82
Paraguay	0.80	0.81	0.81
ICCS average	0.83	0.82	0.83



Table 12.55: Item parameters for scales reflecting students' perceptions of government and law in Latin America

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Authoritarian Govt.</b>	<b>How much do you agree or disagree with the following statements about the government and its leaders?</b>				
LS2P02A	It is better for government leaders to make decisions without consulting anybody.	0.82	-1.24	0.93	0.31
LS2P02B	People in government must enforce their authority even if it means violating the rights of some citizens.	0.70	-1.28	0.33	0.95
LS2P02C	People in government lose part of their authority when they admit their mistakes.	0.01	-1.52	0.06	1.45
LS2P02D	People whose opinions are different than those of the government must be considered its enemies.	0.91	-1.53	0.86	0.68
LS2P02E	The most important opinion of a country should be that of the president.	-0.12	-1.15	0.20	0.95
LS2P02F	It is fair that the government does not comply with the law when it thinks it is not necessary.	0.59	-1.11	0.35	0.75
LS2P03A	Concentration of power in one person guarantees order.	-0.37	-1.68	0.11	1.58
LS2P03B	The government should close communication media that are critical.	0.80	-1.60	0.61	0.98
LS2P03C	If the president does not agree with <Congress>, he/she should dissolve it.	0.26	-1.60	0.36	1.24
<b>Corruption</b>	<b>How much do you agree or disagree with the following statements about the civic service and government?</b>				
LS2P04A	It is acceptable for a civil servant to accept bribes if his/her salary is too low.	0.72	-1.33	0.69	0.63
LS2P04B	It is acceptable for a civil servant to use the resources of the institution in which he/she works for personal benefit	0.74	-1.43	0.19	1.24
LS2P04C	Good candidates grant personal benefits to voters in return for their votes.	0.48	-1.37	0.31	1.06
LS2P04D	Paying an additional amount to a civil servant in order to to obtain a personal benefit is acceptable.	0.74	-1.57	0.31	1.26
LS2P04E	It is acceptable that a civil servant helps his/her friends by giving them employment in his/her office.	-0.01	-1.34	-0.19	1.54
LS2P04F	Since public resources belong to everyone, it is acceptable that those who can keep part of them.	0.60	-1.74	0.45	1.28
	<b>How much do you agree or disagree with the following statements about situations where the law is disobeyed?</b>				
<b>Disobeying Law</b>	<b>A law may be disobeyed ...</b>				
LS2P05A	when it is the only alternative left for achieving important objectives.	-0.50	-1.40	-0.07	1.48
LS2P05B	when it is the only way one has to help one's family.	-0.73	-1.45	-0.25	1.71
LS2P05C	when others who disobeyed it were not punished.	0.27	-1.57	0.44	1.12
LS2P05D	when others do it.	0.41	-1.48	0.48	1.00
LS2P05E	when one distrusts the enacting body.	0.04	-1.69	0.26	1.42
LS2P05F	when one is sure nobody will realize.	0.40	-1.50	0.58	0.93
LS2P05H	when nobody gets hurt.	-0.39	-1.14	-0.24	1.38
LS2P05I	when it is not done with bad intentions.	-0.43	-1.35	-0.15	1.51
LS2P05J	when one is not familiar with the law.	-0.18	-1.46	-0.05	1.52
LS2P05K	when one distrusts the authority executing the law.	-0.03	-1.62	0.25	1.37
LS2P05L	when one can obtain economic benefits.	0.38	-1.22	0.46	0.76

In Question 4, students were asked to rate their level of agreement (“strongly agree,” “agree,” “disagree,” “strongly disagree”) with statements about corruption in the civic service and the government. The first six items, which related to the acceptability of corrupt practices, were used to derive the scale students’ attitudes toward corrupt practices in government (ATTCORR). The higher scores on this scale denote a greater degree of acceptance of corrupt practices. The scale reliability was 0.82 for the pooled ICCS sample, and the reliabilities across countries ranged from 0.77 to 0.84 (see Table 12.54). Table 12.55 shows the item parameters used for scaling.

Question 5 asked students to state the extent to which they agreed (“strongly agree,” “agree,” “disagree,” “strongly disagree”) with statements reflecting the idea that the law can, at times, be disobeyed. Eleven of the 12 items were used to derive the scale *students’ attitudes toward disobeying the law* (DISLAW). The scale reliability was 0.83 for the pooled ICCS sample, and the national reliabilities ranged from 0.80 to 0.85 (see Table 12.54). Table 12.55 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect greater agreement with the notion that it is acceptable to disobey a law under certain circumstances.

Figure 12.29 illustrates the results of the confirmatory factor analysis of these items. The RMSEA for the three-factor solution indicated a close model fit, but the NNFI and CFI suggested some lack of fit. Factor loadings generally tended to be quite large. This was not the case with Item LS2P05J (“disobeying the law is acceptable when one is not familiar with it”), which did not measure DISLAW as well as the other items did; only 25 percent of the item’s variance was explained by the latent factor. The estimated (positive) correlations between the three latent factors were very high, ranging from 0.81 to 0.91.

### **Students’ perceptions regarding peaceful coexistence**

Four scales were derived from questions regarding students’ perceptions related to peaceful coexistence. Their reliabilities are reported in Table 12.48. The following scales are included in the Latin American ICCS student database:

- Students’ attitudes toward neighborhood diversity (ATTDIFF);
- Students’ attitudes toward the use of violence (ATTVIOL);
- Students’ feelings of empathy toward classmates (EMPATH);
- Students’ personal experience of physical and verbal abuse at school (EXPAGG).

Question 6 asked students how they would react (“I would like it,” “I wouldn’t care,” “I would dislike it”) to having neighbors from diverse populations—racial, national, and religious—as well as neighbors who had made particular lifestyle choices or had disabilities or medical conditions. The 10 items associated with the question were used to construct the scale *students’ attitudes toward neighborhood diversity* (ATTDIFF), the higher scores on which correspond to increased acceptance of diversity. The scale reliability was 0.82 for the Latin American sample, and the reliabilities across the participating countries ranged from 0.78 to 0.84 (see Table 12.56). The item parameters that were used for scaling appear in Table 12.57.

Question 8 of the student questionnaire asked students to rate their level of agreement (ranging from “strongly agree” to “strongly disagree”) with four statements relating to the use of violence. The question’s four items were used to derive the scale *students’ attitudes toward the use of violence* (ATTVIOL); the higher scale scores indicate more positive attitudes toward the use of violence. The scale reliability was 0.76 for the Latin American sample. Reliabilities across the six participating countries ranged from 0.71 to 0.79 (see Table 12.56). Table 12.57 shows the item parameters that were used for scaling.



Figure 12.29: Confirmatory factor analysis of items measuring students' attitudes toward government and law in Latin America

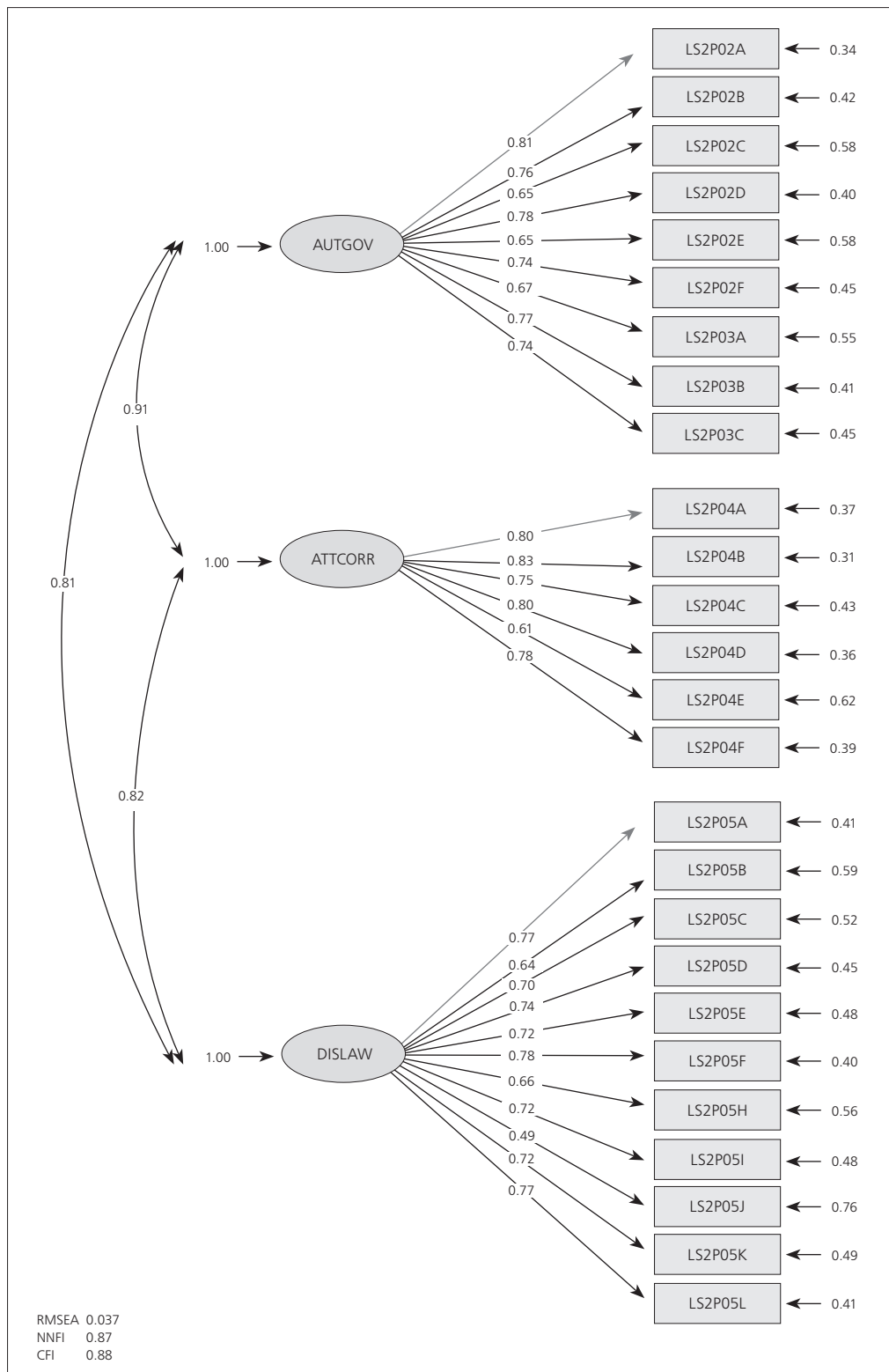


Table 12.56: Reliabilities for scales related to Latin American students' perceptions of peaceful coexistence

Country	Diversity	Violence	Empathy	Abuse
Chile	0.84	0.77	0.86	0.75
Colombia	0.84	0.78	0.84	0.73
Dominican Republic	0.78	0.75	0.93	0.73
Guatemala	0.84	0.76	0.84	0.72
Mexico	0.83	0.79	0.86	0.72
Paraguay	0.79	0.71	0.83	0.74
ICCS average	0.82	0.76	0.87	0.73

Table 12.57: Item parameters for scales related to Latin American students' perceptions of peaceful coexistence

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Diversity</b>	<b>How much would you like or dislike having neighbours belonging to the following groups?</b>				
LS2P06A	People with different skin color than yours.	-2.07	-1.84	1.84	
LS2P06B	People of a different social class than yours.	-1.28	-1.93	1.93	
LS2P06C	People of a different religion than yours.	-1.07	-1.77	1.77	
LS2P06D	Homosexuals or lesbians.	0.92	-1.33	1.33	
LS2P06E	People who come from another region of the country.	-1.73	-1.62	1.62	
LS2P06F	People with physical disabilities.	-1.34	-1.70	1.70	
LS2P06G	People with mental disorders.	-0.40	-1.51	1.51	
LS2P06H	People of a different nationality than yours.	-1.70	-1.57	1.57	
LS2P06I	People with AIDS.	0.55	-1.27	1.27	
LS2P06J	People of indigenous origin.	-1.17	-1.55	1.55	
<b>Violence</b>	<b>How much do you agree or disagree with the following phrases?</b>				
LS2P08A	He who does me harm will have to pay for it.	0.00	-2.11	0.70	1.40
LS2P08B	Watching fights between classmates is fun.	1.04	-2.19	0.74	1.44
LS2P08C	If you cannot do it the easy way, do it the hard way.	0.75	-2.01	0.64	1.37
LS2P08D	You have to fight so people do not think you are a coward.	1.19	-1.76	0.89	0.88
<b>Empathy</b>	<b>How do you feel when you witness the following situations at your school?</b>				
LS2P09A	A classmate falls and gets hurt.	-1.35	-1.08	1.08	
LS2P09B	A classmate gets beaten up.	-2.55	-1.08	1.08	
LS2P09C	A classmate gets unfairly reprimanded.	-2.60	-0.54	0.54	
LS2P09D	A classmate gets unfairly punished.	-2.88	-0.64	0.64	
LS2P09E	A classmate gets something stolen from him/her.	-2.92	-0.90	0.90	
LS2P09F	A classmate gets ridiculed.	-2.22	-0.69	0.69	
LS2P09G	A classmate gets insulted.	-2.68	-1.05	1.05	
LS2P09H	A classmate looks very sad.	-2.43	-1.59	1.59	
LS2P09I	A classmate gets bad grades.	-1.50	-2.46	2.46	
LS2P09J	A classmate has nobody to play with.	-2.30	-1.29	1.29	
<b>Abuse</b>	<b>Last month, how often did the following happen to you at your school?</b>				
LS2P10A	Someone in your school hit, slapped, kicked, pushed, or pinched you.	-1.39	-0.16	-0.14	0.30
LS2P10B	Someone in your school insulted you.	-0.69	-0.37	-0.36	0.74
LS2P10C	Someone threatened to hit you.	-1.48	-0.33	0.04	0.30
LS2P10D	Someone rejected you and did not allow you to join their group.	-1.61	-0.16	-0.05	0.20
LS2P10E	A classmate called you an offensive nickname.	-0.67	0.16	-0.46	0.30

Question 9 presented students with 10 items conveying a series of situations involving classmates that they might witness at school. Students were asked to express how they felt about the events depicted (“I think it’s fun,” “I don’t care,” “It bothers me”). All 10 items were used to derive the scale *students’ feelings of empathy toward classmates* (EMPATH). The scale reliability was 0.87 for the pooled ICCS sample, and the national reliabilities ranged from 0.83 to 0.93 (see Table 12.56). Table 12.57 shows the item wording and the item parameters that were used for scaling. Students who scored highly on this scale were students who expressed a greater degree of empathy toward their classmates.

Question 10 asked students to indicate how often they themselves had experienced acts of physical and/or verbal abuse in the past month (“never,” “only once,” “two to four times,” “five times or more”). The five question items were used to form the scale *students’ personal experience of physical and verbal abuse at school* (EXPAGG). The scale had a reliability (Cronbach’s alpha) of 0.73 for the pooled ICCS sample; national reliabilities ranged from 0.72 to 0.75 (see Table 12.56). Table 12.57 shows the item wording as well as the item parameters that were used for scaling. Higher values on this scale reflect more frequent experiences with physical and verbal aggression at school.

The results of the confirmatory factor analysis for these item sets appear in Figure 12.30. The four-factor solution had a close model fit, while the size of the factor loadings indicated high measurement reliability at the item level. Correlations between the latent factors tended to be weak. Only ATTVIOL (attitudes toward the use of violence) and EMPATH (feelings of empathy) had a strong negative correlation, indicating that students who felt more empathetic were also more likely to reject the use of violence.

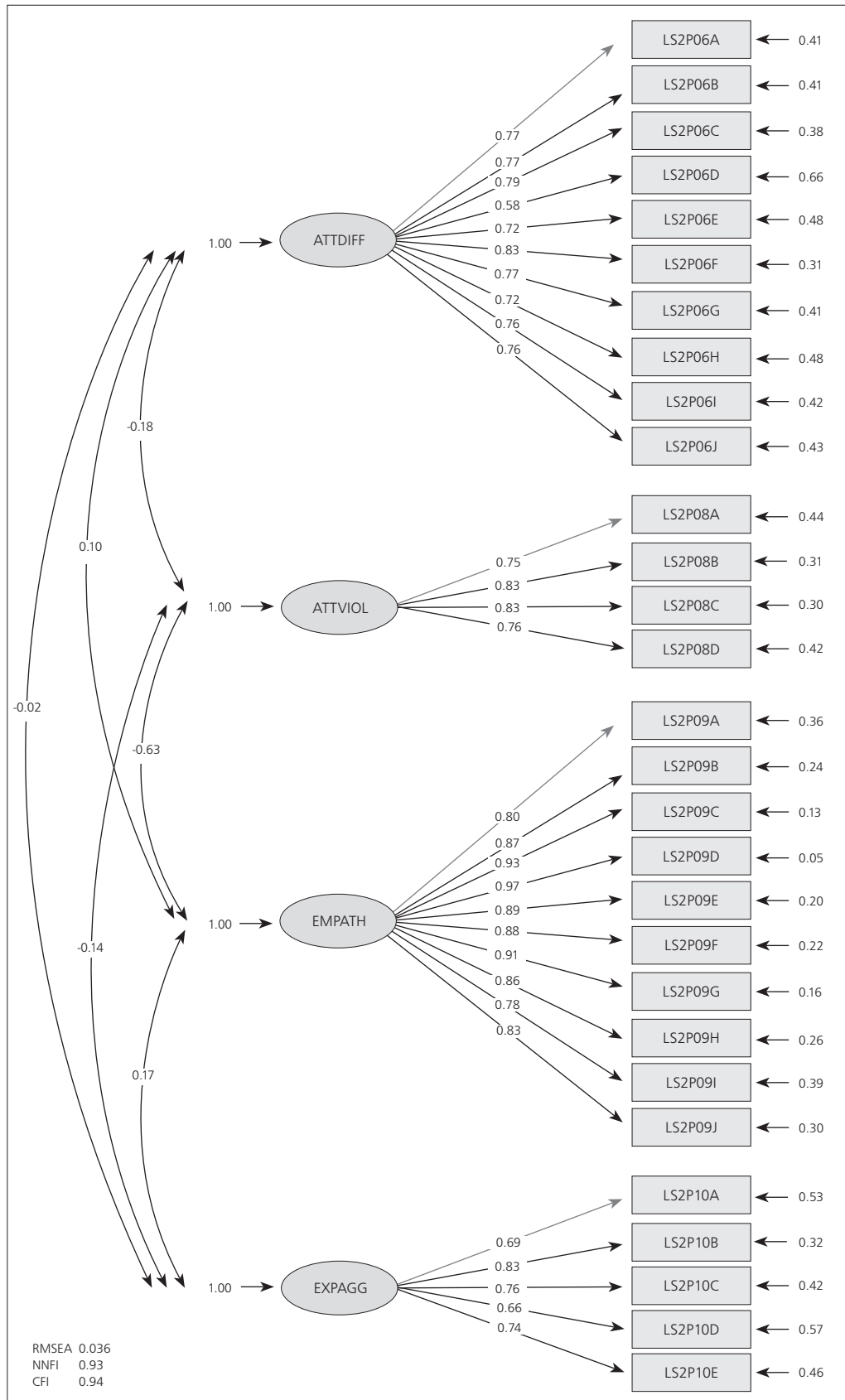
#### **Students’ reports on discussion of civic issues at school**

Question 12 asked students to indicate how often (“not at all,” “a little,” “sometimes,” “often”) a series of civic issues were discussed at their school. The nine items stating these issues were used to form the scale *students’ reports on frequency of discussions about civic issues at school* (SCHDISC). The scale reliability was 0.84 for the pooled ICCS sample, and the national reliabilities ranged from 0.82 to 0.85 (see Table 12.58). The item parameters that were used for scaling are shown in Table 12.59.

Figure 12.31 shows the results from the confirmatory factor analysis of this item set. The one-factor solution had a good model fit. The factor loading for Item LS2P12A (“rights and duties of citizens”) indicated a somewhat lower reliability for this item than for others, with only 35 percent of its variance explained by the latent factor.



Figure 12.30: Confirmatory factor analysis of items measuring Latin American students' perceptions of peaceful coexistence



*Table 12.58: Reliabilities for scale reflecting Latin American students' reports on frequency of discussions at school about civic issues*

Country	Discussion Civic Issues
Chile	0.85
Colombia	0.85
Dominican Republic	0.83
Guatemala	0.85
Mexico	0.85
Paraguay	0.82
ICCS average	<b>0.84</b>

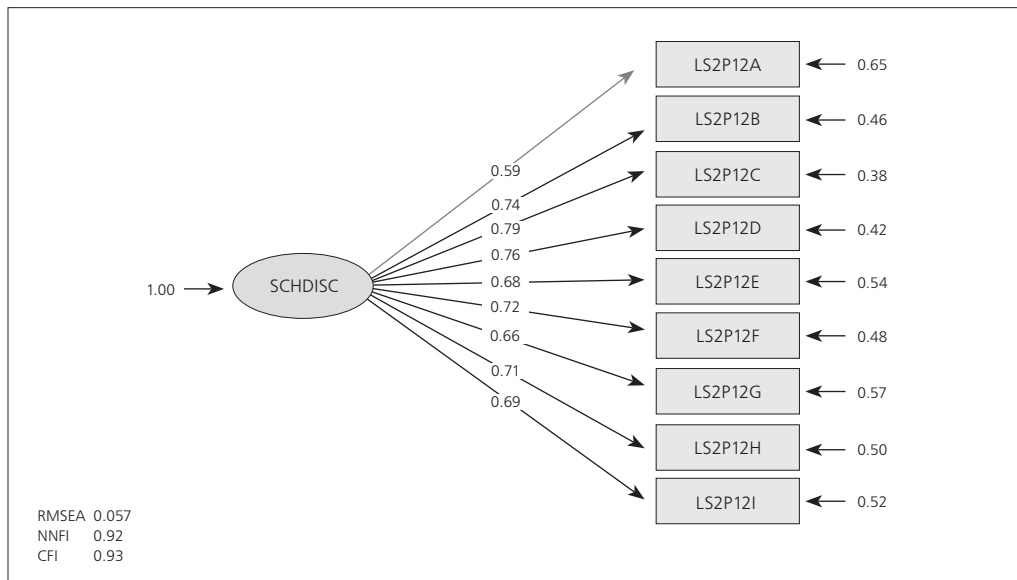
*Table 12.59: Item parameters for scale reflecting Latin American students' reports on frequency of discussions at school about civic issues*

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
Discussion Civic Issues	At your school, how much have the following issues been discussed?				
LS2P12A	Rights and duties you assume as a citizen when you become an adult	-0.58	-0.61	-0.15	0.76
LS2P12B	Consequences of consuming illegal drugs	-0.56	0.25	-0.14	-0.12
LS2P12C	Integration of people with different sexual tendencies and orientations in the community	-0.17	-0.85	-0.24	1.08
LS2P12D	Discrimination against people with different sexual orientation	-0.15	-0.63	-0.13	0.76
LS2P12E	Advantages and disadvantages of non-governmental organizations operating in a democratic country	-0.01	-1.03	-0.13	1.16
LS2P12F	Integration of people with different cultural backgrounds in the school, neighborhood, or community	-0.30	-0.97	-0.14	1.12
LS2P12G	Respect for different religious rites	-0.51	-0.63	0.05	0.57
LS2P12H	Facilities that people with physical and mental disabilities should have in different environments (school, street, workplace, etc.	-0.42	-0.87	-0.01	0.88
LS2P12I	Difficulties encountered by people with AIDS in being accepted by society	-0.32	-0.55	-0.12	0.68





Figure 12.31: Confirmatory factor analysis of items measuring Latin American students' reports on frequency of discussions at school about civic issues



#### Asian questionnaire

##### Students' perceptions of government and law in Asia

The three scales derived from questions regarding students' perceptions of government and law in Asia and included in the Asian ICCS student database are listed immediately below. The scales' reliabilities are reported in Table 12.60.

- Students' acceptance of authoritarian government practice (UNDEMGOV);
- Students' attitudes toward obedience to authority (OBAUTH);
- Students' perceptions of the integrity of the legal system (LEGSYS).

The first question of the Asian regional questionnaire asked students to indicate their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with five statements denoting attitudes toward democratic government. The items were used to derive the scale *students' acceptance of authoritarian government practice* (UNDEMGOV). The scale reliability was 0.81 for the Asian sample, and reliabilities across the countries participating in the Asian regional module ranged from 0.59 to 0.81 (see Table 12.60). Table 12.61 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect agreement with statements denoting acceptance of undemocratic government.

Question 2 in the Asian module required students to indicate their level of agreement (ranging from "strongly agree" to "strongly disagree") with statements about how best to behave in society. Four of these items were used to derive the scale *students' attitudes toward obedience to authority* (OBAUTH). The higher scores on this scale denote agreement with the notion that people need to be obedient. The scale reliability was 0.84 for the pooled ICCS sample, and the national reliabilities ranged from 0.61 to 0.82 (see Table 12.60). Table 12.61 shows the item parameters that were used for scaling.



Table 12.60: Reliabilities for scales reflecting students' perceptions of government and law in Asia

Country	Authoritarian Govt.	Obedience	Integrity Legal System
Chinese Taipei	0.81	0.80	0.72
Hong Kong SAR	N/A	0.80	0.70
Indonesia	0.59	0.69	0.42
Korea, Republic of	0.78	0.82	0.73
Thailand	0.71	0.61	0.51
ICCS average	0.81	0.84	0.68

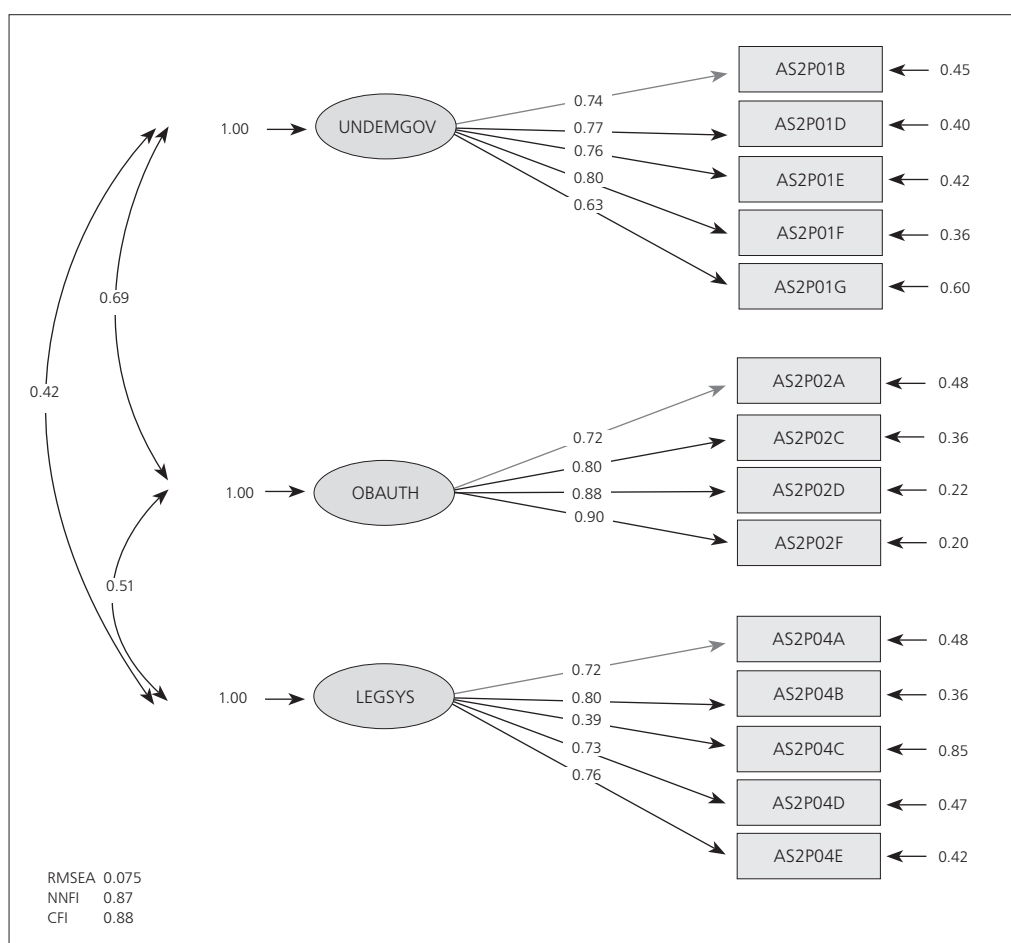
Table 12.61: Item parameters for scales reflecting students' perceptions of government and law in Asia

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Authoritarian Govt.</b>	<b>How much do you agree or disagree with each of the following statements about the government of your country?</b>				
AS2P01B	As long as everyone can enjoy prosperity, it does not matter whether the government is democratic or not.	0.52	-2.34	0.28	2.05
AS2P01D	As long as the government represents citizens' ideas, it does not matter whether the government is democratic or not.	0.50	-2.44	0.37	2.07
AS2P01E	It is acceptable for the government to act undemocratically in order to do its job more efficiently.	0.73	-2.13	0.25	1.89
AS2P01F	The more power the government has, the more likely it is to solve its people's problems.	0.05	-1.84	-0.10	1.94
AS2P01G	It is acceptable for the government to break the law when it considers it necessary.	0.90	-1.54	0.04	1.51
<b>Obedience</b>	<b>How much do you agree or disagree with each of the following statements about how to best behave in society?</b>				
AS2P02A	Even if you have a different opinion, you should always follow the advice of elders when making important decisions.	-0.86	-3.01	-0.31	3.32
AS2P02C	Even if you have a different opinion, you should always follow the advice of the people with the highest status position when making important decisions.	0.16	-2.82	0.05	2.78
AS2P02D	Even if you have a different opinion, you should always obey your teachers.	-0.40	-2.57	-0.13	2.71
AS2P02F	Even if you have a different opinion, you should always obey your parents.	-1.01	-2.50	0.05	2.45
<b>Integrity Legal System</b>	<b>How much do you agree or disagree with each of the following statements about the law and judiciary in your country?</b>				
	In <country of test> ...				
AS2P04A	the law favors those who have money and power.	-0.27	-1.35	-0.15	1.50
AS2P04B	everyone is equally treated by the law.	-1.24	-1.55	-0.11	1.65
AS2P04C	the government often intervenes in decisions made by the courts.	0.11	-2.05	0.00	2.05
AS2P04D	the courts are able to apply the law fairly.	-1.33	-1.35	-0.53	1.89
AS2P04E	there is no corruption in the legal system.	-0.39	-1.42	0.15	1.27

Question 4 asked students to rate their level of agreement (again ranging from “strongly agree” to “strongly disagree”) with five statements concerning the law and the judiciary in their country. These five items were used to derive the scale *students’ perceptions of the integrity of the legal system* (LEGSYS); students with higher scale scores were those who expressed greater faith in the integrity of the legal system. The scale reliability was 0.68 for the pooled ICCS sample, and the national reliabilities ranged from 0.42 to 0.73 (see Table 12.50). Table 12.51 shows the item parameters that were used for scaling.

The results of the confirmatory factor analysis of these three items sets are shown in Figure 12.32. The model fit was not close. However, as judged by the RMSEA, it was still in an acceptable range. The low factor loading for Item AS2P04C (frequent intervention of government in legal decisions) suggests that this item did not measure LEGSYS (integrity of legal system) as well as the other items did. The (positive) correlations between the three latent factors ranged from 0.42 (between UNDEMGOV, i.e., authoritarian government practices, and LEGSYS, integrity of the legal system) to 0.69 (between UNDEMGOV and OBAUTH, obedience to authority).

Figure 12.32: Confirmatory factor analysis of items measuring students’ perceptions of government and law in Asia



### Students' perceptions of identity, citizenship, and culture in Asia

Three scales were derived from questions that asked students to give their perceptions of traditional culture, good citizenship, and Asian identity. The reliabilities of these scales, each of which is included in the Asian ICCS student database, are reported in Table 12.62.

- Students' attitudes toward the preservation of traditional culture (TRADCL);
- Students' sense of Asian identity (ASIAID);
- Students' perceptions of good citizenship (ASIACIT).

Table 12.62: Scale reliabilities for scales reflecting students' perceptions of identity, citizenship, and culture in Asia

Country	Traditional Culture	Asian Identity	Good Citizenship
Chinese Taipei	0.77	0.90	0.80
Hong Kong SAR	0.76	0.86	0.79
Indonesia	0.65	0.84	0.60
Korea, Republic of	0.69	0.87	0.76
Thailand	0.69	0.78	0.69
ICCS average	0.75	0.85	0.73

Question 3 of the questionnaire for students participating in the Asian ICCS regional module asked these students to signal their level of agreement (“strongly agree,” “agree,” “disagree,” “strongly disagree”) with four statements about preserving the traditional culture of their respective countries. All four statement items were used to derive the scale *students' attitudes toward the preservation of traditional culture* (TRADCL). The scale reliability was 0.75 for the Asian sample, and the national reliabilities ranged from 0.65 to 0.77 (see Table 12.62). Table 12.63 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect positive attitudes toward preservation of traditional culture.

Question 6 asked students to indicate their level of agreement (“strongly agree” to “strongly disagree”) with statements about the Asian region and their sense of identity as Asians. Seven of the question items were used to construct the scale *students' sense of Asian identity* (ASIAID); the higher scale scores corresponded to a greater sense of Asian identity. The scale reliability was 0.85 for the pooled ICCS sample, and the national reliabilities ranged from 0.78 to 0.90 (see Table 12.62). Table 12.63 presents the item parameters that were used for scaling.

In Question 7, students were presented with seven statements indicating possible characteristics of good citizenship. Students were asked to rate their level of agreement with each of the statements (“strongly agree” to “strongly disagree”). The seven items were used to derive the scale *students' perceptions of good citizenship* (ASIACIT), which had a scale reliability of 0.73 for the Asian sample and national reliabilities ranging from 0.60 to 0.80 (see Table 12.52). Table 12.53 shows the item wording and the item parameters that were used for scaling. Higher values on this scale reflect stronger agreement with the listed requirements for being a good citizen.

Figure 12.33 shows the results of the confirmatory factor analysis for these item sets. The RMSEA suggested a close item fit, but both the NNFI and CFI indicated some lack of fit for the three-factor solution. The relatively low factor loading for Item AS2P07B (“a person who obeys the law but does not behave morally is not a good citizen”), for which only 25 percent of its variance was explained by the underlying construct, suggests that the item did not measure the latent factor ASIACIT (characteristics of a good citizen) to the same extent as the other items in the scale did. Strong positive correlations were observed between the three latent factors; these ranged from 0.67 to 0.70.

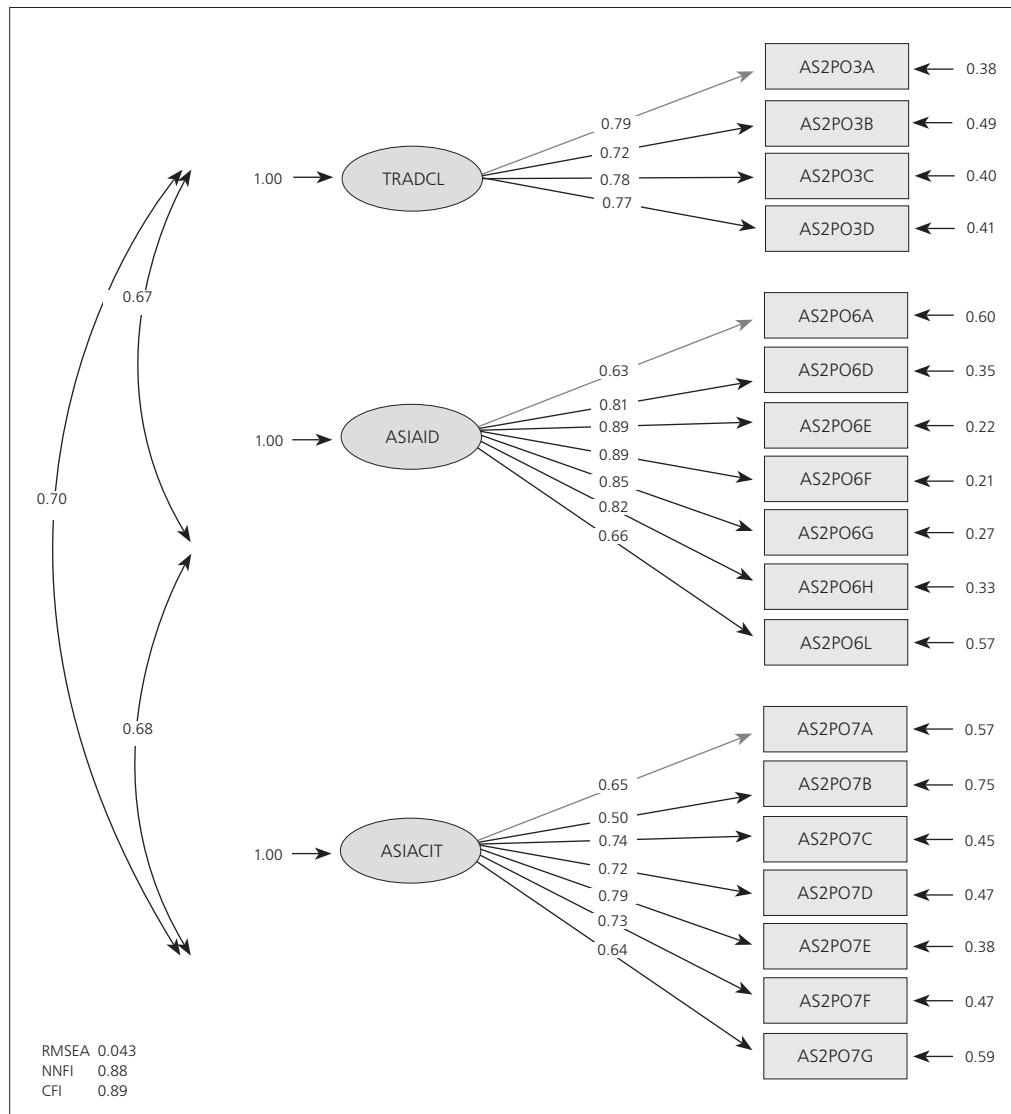


Table 12.63: Item parameters for scales reflecting students' perceptions of identity, citizenship, and culture in Asia

Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Traditional Culture</b>	<b>How much do you agree or disagree with each of the following statements about traditional culture in your own country?</b>				
AS2P03A	I would like to have more opportunities to learn about <country of test>'s traditional culture.	-2.11	-1.70	-1.13	2.84
AS2P03B	<country of test> needs to maintain its unique cultural identity against the influence of other cultures.	-1.91	-2.32	-0.33	2.65
AS2P03C	Because <country of test>'s traditional culture represents our cultural heritage, all parts of our traditional culture should be preserved.	-2.35	-1.67	-0.91	2.58
AS2P03D	I feel responsible for preserving <country of test>'s traditional culture.	-1.99	-1.98	-0.95	2.92
<b>Asian Identity</b>	<b>How much do you agree or disagree with each of the following statements about the Asian region and Asian identity?</b>				
AS2P06A	I think of myself as an Asian citizen.	-1.66	-1.75	-0.85	2.60
AS2P06D	I am proud of the economic progress that has been made across Asia as a whole.	-1.81	-2.23	-0.82	3.04
AS2P06E	I am proud of being Asian.	-1.69	-2.22	-0.72	2.95
AS2P06F	I am proud of Asian cultural traditions.	-1.89	-2.29	-0.73	3.01
AS2P06G	I am proud of the progress of democracy that has been made across Asia as a whole.	-1.87	-2.35	-0.87	3.22
AS2P06H	I am proud of the progress that has been made in human rights across Asia as a whole.	-1.98	-2.45	-0.85	3.29
AS2P06L	I feel I have a lot in common with other young people in Asia.	-1.33	-2.89	-0.54	3.44
<b>Good Citizenship</b>	<b>How much do you agree or disagree with each of the following statements about being a good citizen?</b>				
AS2P07A	A person who obeys the law is a good citizen.	-1.71	-1.52	-0.43	1.94
AS2P07B	A person who obeys the law but does not behave morally is not a good citizen.	-0.97	-1.65	-0.62	2.28
AS2P07C	One can only be a good citizen if one is a good moral person.	-1.42	-2.09	-0.16	2.24
AS2P07D	Having good morality is more important than having good knowledge for one to be a good citizen.	-1.46	-1.64	-0.66	2.30
AS2P07E	Self-cultivation is an important process of becoming a good citizen.	-1.76	-1.45	-1.09	2.55
AS2P07F	For one to become a good citizen, one must have a high quality of spirituality.	-1.68	-1.89	-0.60	2.49
AS2P07G	Even if a person behaves properly, they cannot be a good citizen without a high quality of spirituality.	-0.94	-2.03	-0.36	2.39



Figure 12.33: Confirmatory factor analysis of items measuring students' perceptions of identity, citizenship, and culture in Asia



### Students' perceptions of public service

Three scales, the reliabilities of which are reported in Table 12.64, were derived from questions regarding students' perceptions of public service. The scales, all of which are included in the Asian ICCS student database, were named as follows:

- Students' attitudes toward personal morality of politicians (MORALPOL);
- Students' attitudes toward corruption in public service (CORRPUB);
- Students' attitudes toward the use of connections to hold public office (GUANXI).

Table 12.64: Reliabilities for scales reflecting students' perceptions of public service in Asia

Country	Morality Politicians	Corruption Pub. Service	Guanxi
Chinese Taipei	0.72	0.71	0.87
Hong Kong SAR	0.71	0.72	0.86
Indonesia	0.62	0.58	0.75
Korea, Republic of	0.74	0.63	0.83
Thailand	0.64	0.59	0.78
ICCS average	0.68	0.72	0.85

Question 5 of the Asian regional instrument contained statements about public officials and politicians for which students indicated their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree"). Two scales were derived from the items within this question.

The scale *students' attitudes toward personal morality of politicians* (MORALPOL) was constructed from five items that addressed morality and principles displayed by politicians along with the responsibility that these people have to ensure that their families behave morally. The scale reliability (Cronbach's alpha) of MORALPOL was 0.68 for the pooled ICCS sample, and the national reliabilities ranged from 0.62 to 0.74 (see Table 12.64). Table 12.65 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect stronger agreement with statements about the importance of politicians behaving morally.

The three remaining items in Question 5 were used to derive the scale *students' attitudes toward corruption in public service* (CORRPUB). These items featured statements regarding the acceptability of corruption in public office (guanxi in the Chinese context). The scale reliability (Cronbach's alpha) was 0.72 for the pooled ICCS sample, and reliabilities ranged from 0.58 to 0.72 across the five national samples (see Table 12.64). The item parameters that were used for scaling are shown in Table 12.65.

The final question in the Asian regional instrument (Question 8) asked students to rate their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with statements about the role of connections in elections or public office. The five items associated with the question were used to derive the scale *students' attitudes toward the use of connections to hold public office* (GUANXI), which had a scale reliability (Cronbach's alpha) of 0.85 for the Asian sample and reliabilities ranging from 0.75 to 0.87 across the national samples (see Table 12.64). Table 12.65 shows the item wording as well as the item parameters that were used for scaling. The higher values on this scale reflect positive attitudes toward using connections in order to secure public office.



Table 12.65: Item parameters for scales reflecting students' perceptions of public service in Asia

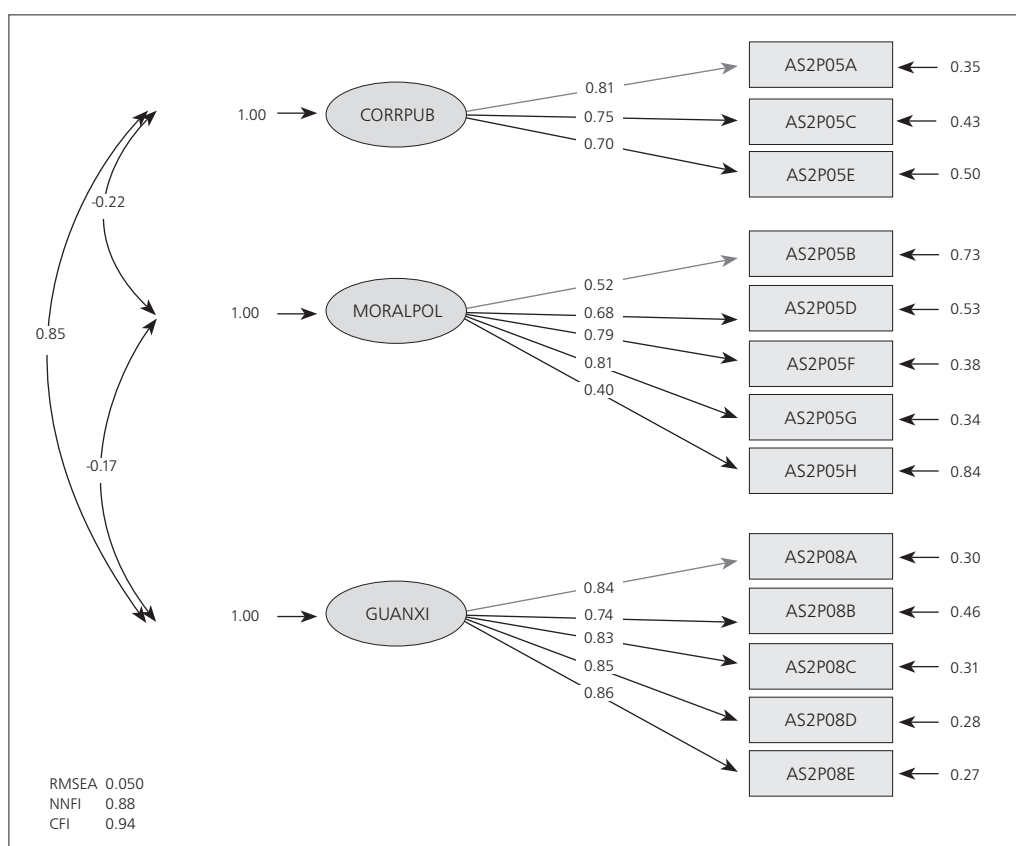
Scale or Item	Question/Item Wording	Delta	Tau(1)	Tau(2)	Tau(3)
<b>Morality Politicians</b>	<b>How much do you agree or disagree with each of the following statements about public officials and politicians?</b>				
AS2P05B	The honesty and morality of a politician is more important than his/her abilities.	-1.26	-1.11	-0.77	1.87
AS2P05D	Political leaders should be role models of morality.	-1.63	-0.66	-1.12	1.79
AS2P05F	Politicians have the responsibility to make sure that their family obeys the law.	-1.43	-1.05	-1.05	2.11
AS2P05G	Politicians have the responsibility to make sure that their family behaves morally.	-1.48	-1.09	-0.99	2.08
AS2P05H	Politicians should be accountable if a member of their family breaks the law or behaves immorally.	-0.97	-1.45	-0.22	1.68
<b>Corruption Pub. Service</b>	<b>How much do you agree or disagree with each of the following statements about public officials and politicians?</b>				
AS2P05A	It is acceptable to bribe government officials to get things done effectively.	1.25	-1.17	-0.17	1.35
AS2P05C	It doesn't matter if a public official uses resources from the institution where he/she works for his/her personal benefit.	1.03	-1.96	0.23	1.74
AS2P05E	Preventing corruption is adults' business; it has nothing to do with me.	0.92	-2.07	0.47	1.61
<b>Guanxi</b>	<b>How much do you agree or disagree with each of the following statements about the role of &lt;connections&gt; in elections or public office?</b>				
AS2P08A	If there are many candidates in an election, we should only vote for the people from our <hometown/local area>.	0.83	-2.54	0.66	1.87
AS2P08B	Only the candidates we have <connections> with would truly serve us after they get elected.	0.46	-2.71	0.30	2.42
AS2P08C	If a candidate is a friend or relative, then we should vote for him/her even if he/she is not the best candidate for the job.	1.09	-2.43	0.76	1.66
AS2P08D	It is acceptable for public officials to give preference to family and friends when hiring people for public office.	0.84	-2.05	-0.05	2.09
AS2P08E	It is acceptable for a public official to give government contracts to people they have <connections> with even if they are not the best qualified to do the contract work.	1.08	-2.04	0.25	1.78

The results of a confirmatory factor analysis of these item sets (see Figure 12.34) suggested a good model fit for the three-factor solution. Two of the five items measuring MORALPOL (personal morality of politicians), AS2P05B ("the honesty and morality of a politician is more important than his/her abilities") and AS2P05H ("politicians should be accountable if a member of their family breaks the law or behaves immorally"), had much lower factor loadings than the other items forming this scale, and only a low percentage of their variance was explained by the underlying latent construct. There was a high positive correlation (0.85) between the two latent factors CORRPUB (attitudes toward corruption in public service) and GUANXI (attitudes toward use of connections).





Figure 12.34: Confirmatory factor analysis of items measuring students' perceptions of public service in Asia



## Summary

ICCS derived two different types of indices from the different questionnaires administered to students, teachers, and schools. A number of indices were constructed through recoding of values, combination of separate variables, and arithmetic calculations. Another type of index was derived through scaling procedures.

Item response modeling (mainly applying the Rasch partial credit model) was used to derive 24 international student questionnaire scales, 12 teacher questionnaire scales, and 12 school questionnaire scales. In addition, regional questionnaire data provided further scale indices for ICCS: 13 for the European regional database and nine for the Asian and Latin American regional databases. A composite index reflecting socioeconomic background was derived using principal component analysis of three home background indicators, namely, parental occupation, parental education, and home literacy resources.

Generally, the scales used in ICCS had sound psychometric properties, such as high reliabilities. Confirmatory factor analyses showed satisfactory model fit for the sets of items used to derive latent variables.



## References

- Andersen, E. B. (1997). The rating scale model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 67–84). New York/Berlin/Heidelberg: Springer.
- Bollen, K. A., & Long, S. J. (1993). (Eds.). *Testing structural equation models*. Newbury Park, CA: Sage Publications.
- Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A standard international socioeconomic index of occupational status. *Social Science Research*, 21, 1–56.
- International Labour Organization (ILO). (1990). *International Standard Classification of Occupations: ISCO-88*. Geneva, Switzerland: Author.
- Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality & Quantity*, 24, 387–404.
- Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, 59(3), 381–389.
- Jöreskog, K. G., & Sörbom, D. (2004). *LISREL 8.7 for Windows* [computer software]. Lincolnwood, IL: Scientific Software International.
- Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks, CA: Sage Publications.
- 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.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.
- Schulz, W. (2004). Mapping student scores to item responses. In W. Schulz & H. Sibberns (Eds.), *IEA Civic Education Study technical report* (pp. 127–132). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study. *IERI Monograph Series Volume 2*, 113–135.
- UNESCO. (2006). *International Standard Classification of Education: ISCED 1997*. Montreal, Quebec, Canada: UNESCO-UIS.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ACER ConQuest: General item response modelling software* [computer program]. Camberwell, Victoria, Australia: Australian Council for Educational Research.





# The reporting of ICCS results

*Wolfram Schulz*

## Introduction

This chapter describes the procedures that were used to report results in the ICCS publications. The chapter begins with a description of the replication methodology used to estimate sampling variance and then provides an outline of how the imputation variance of the civic knowledge scores was computed. The subsequent section describes how the significance tests for differences between country and subsample means or percentages were conducted.

This chapter also includes descriptions of how the multiple regression analyses were conducted and how the hierarchical models explaining civic knowledge were estimated. The final section of the chapter outlines how missing data were treated during multivariate analyses of the ICCS data.

## Estimation of sampling variance

ICCS employed two-stage cluster sampling procedures to obtain the student as well as the teacher samples. During the first stage, schools were sampled from a sampling frame with a probability proportional to their size. During the second stage, intact classrooms were randomly sampled within schools. Cluster sampling techniques permit an efficient and economic data collection. However, because these samples are not simple random samples, the usual formulae used to obtain standard errors for population estimates are not appropriate.

Replication techniques provide tools with which to estimate the correct sampling variance on population estimates (Gonzalez & Foy, 2000; Wolter, 1985). ICCS used the jackknife repeated replication technique (JRR) to compute standard errors for population means, percentages, regression coefficients, and any other population statistic.

Generally, the JRR method for stratified samples requires pairing primary sampling units (PSUs)—in this survey, schools—into pseudo-strata. Because assignment of schools to these “sampling zones” needed to be consistent with the sampling frame from which they were sampled, sampling zones were constructed within explicit strata. Occurrences of an odd number of schools within an explicit stratum or the sampling frame saw the remaining school randomly divided into two halves, thereby forming a sampling zone of two “quasi-schools.”

Each of the countries participating in ICCS had up to 75 sampling zones. In countries with a large number of participating schools, some schools were combined into bigger “pseudo-schools” in order to keep the total number to 75. A three-stage sample design was applied to the Russian Federation. The first stage of this process consisted of a sample of regions. If a selected region was large enough to be selected with certainty, schools were paired. If this was not the case, regions in the sampling zones were paired. In countries where all schools were tested and two classrooms within each school had been sampled (i.e., Cyprus and Malta), schools were defined as sampling zones and classrooms as PSUs. In countries with census surveys (Liechtenstein and Luxembourg), students were randomly assigned to sampling zones and quasi-schools. The same procedure was applied to teachers in countries in which all schools were selected. Table 13.1 shows the range of sampling zones for the student, school, and teacher samples used in each participating country.

Within each of the sampling zones, one school was randomly assigned a value of 2 and the other school a value of 0. Replicate weights were computed for each of the sampling zones. This meant that one of the paired schools had a contribution of zero, the second a double contribution, and all other schools remained the same. The replicate weights procedure is achieved by simply multiplying student or teacher weights with the jackknife indicators once only for each sampling zone.



Table 13.1: Numbers of sampling zones in national samples

Country	Student Data	School Data	Teacher Data
Austria	68	55	38
Belgium (Flemish)	75	75	68
Bulgaria	75	75	75
Chile	75	75	75
Chinese Taipei	75	75	73
Colombia	75	75	75
Cyprus	75	30	75
Czech Republic	73	68	74
Denmark	75	75	57
Dominican Republic	73	73	73
England	62	56	59
Estonia	71	64	67
Finland	75	75	75
Greece	75	65	N/A
Guatemala	74	73	74
Hong Kong SAR	39	43	52
Indonesia	71	70	71
Ireland	73	68	69
Italy	75	75	75
Korea, Republic of	75	75	75
Latvia	75	69	74
Liechtenstein	75	5	59
Lithuania	75	75	75
Luxembourg	75	11	75
Malta	75	27	75
Mexico	75	75	75
Netherlands	34	24	N/A
New Zealand	75	62	75
Norway	65	60	37
Paraguay	75	75	71
Poland	75	75	75
Russian Federation	62	62	62
Slovak Republic	71	71	71
Slovenia	75	75	75
Spain	75	75	75
Sweden	75	75	75
Switzerland	75	74	74
Thailand	75	75	75



This process results in a weight being added to the data file for each jackknife replicate. Thus, within one sampling zone at a time, each element of one PSU receives a double weight and each element of the other PSU receives a zero weight. Table 13.2 illustrates this procedure through a simple example featuring 24 students from six different schools (A–F) paired into three sampling zones.

For each country sample, 75 replicate weights were computed regardless of the number of sampling zones. In countries with fewer sampling zones, the remaining replicate weights were equal to the original sampling weight and therefore did not contribute to the sampling variance estimate.

Estimating the sampling variance for a statistic,  $\mu$ , involves computing it once with the sampling weights for the original sample and then with each of the 75 replication weights separately. The sampling variance  $SV_\mu$  estimate is computed using the formula

$$SV_\mu = \sum_{i=1}^{75} [\mu_i - \mu_s]^2,$$

where  $\mu_s$  is the statistic  $\mu$  estimated for the population through use of the original sampling weights and  $\mu_i$  is the same statistic estimated by using the weights for the  $i^{th}$  of 75 jackknife replicates. The standard error  $SE_\mu$  for statistic  $\mu$  is computed as:

$$SE_\mu = \sqrt{SV_\mu}.$$

The computation of sampling variance using jackknife replication can be obtained for any statistic, including means, percentages, standard deviations, correlations, regression coefficients, and mean differences. Standard statistical software rarely includes procedures for replication techniques.

Tailored SPSS software macros were used for the jackknife replication of ICCS data. Most of these results can be replicated by using the IEA IDB Analyzer, which is generally recommended as a tool for analyzing IEA data.<sup>1</sup> Alternatively, analysts can use other specialized software, such as WESVAR (Westat, 2007), or tailored applications, such as the SPSS Replicates Module developed by the Australian Council for Educational Research (ACER).<sup>2</sup>

### Estimation of imputation variance for civic knowledge scores

When estimating standard errors for test scores reflecting ICCS civic knowledge, it is important to take the imputation variance into account (see Chapter 11 for a description of the scaling methodology for ICCS test items). Therefore, population statistics for ICCS civic knowledge scores should always be estimated through use of all five plausible values.

If  $\theta$  is the international civic knowledge and  $\mu_\theta^p$  is the statistic of interest computed based on each of the  $P$  plausible values, then the statistic  $\mu_\theta$  based on all plausible values can be computed as follows:

$$\mu_\theta = \frac{1}{P} \sum_{p=1}^P \mu_\theta^p.$$

1 The IDB Analyzer is a plug-in for the Statistical Package for the Social Sciences (SPSS) that allows the user to combine and analyze data from IEA's large-scale assessments such as TIMSS, PIRLS, and SITES. The application can be downloaded at [http://www.iea.nl/iea\\_studies\\_datasets.html](http://www.iea.nl/iea_studies_datasets.html)

2 The module is an add-in component running under SPSS. It offers some features for applying different replication methods when estimating sampling and imputation variance. The application can be downloaded at <https://mypisa.acer.edu.au/>





Table 13.2: Example of computation of replicate weights

ID	Student Weight	School	Sampling Zone	Jackknife Indicator	Replicate Weight 1	Replicate Weight 2	Replicate Weight 3
1	5.2	A	1	0	0	5.2	5.2
2	5.2	A	1	0	0	5.2	5.2
3	5.2	A	1	0	0	5.2	5.2
4	5.2	A	1	0	0	5.2	5.2
5	9.8	B	1	2	19.6	9.8	9.8
6	9.8	B	1	2	19.6	9.8	9.8
7	9.8	B	1	2	19.6	9.8	9.8
8	9.8	B	1	2	19.6	9.8	9.8
9	6.6	C	2	2	6.6	13.2	6.6
10	6.6	C	2	2	6.6	13.2	6.6
11	6.6	C	2	2	6.6	13.2	6.6
12	6.6	C	2	2	6.6	13.2	6.6
13	7.2	D	2	0	7.2	0	7.2
14	7.2	D	2	0	7.2	0	7.2
15	7.2	D	2	0	7.2	0	7.2
16	7.2	D	2	0	7.2	0	7.2
17	4.9	E	3	2	4.9	4.9	9.8
18	4.9	E	3	2	4.9	4.9	9.8
19	4.9	E	3	2	4.9	4.9	9.8
20	4.9	E	3	2	4.9	4.9	9.8
21	8.2	F	3	0	8.2	8.2	0
22	8.2	F	3	0	8.2	8.2	0
23	8.2	F	3	0	8.2	8.2	0
24	8.2	F	3	0	8.2	8.2	0

The sampling variance  $SV_{\mu}$  is calculated as the average of the sampling variance for each plausible value  $SV_{\mu}^p$ :

$$SV_{\mu} = \frac{1}{P} \sum_{p=1}^P SV_{\mu}^p .$$

Use of the  $P$  plausible values for data analysis also allows one to estimate the amount of error associated with the measurement of civic knowledge. The measurement variance or imputation variance  $IV_p$  is computed as

$$IV_p = \frac{1}{P-1} \sum_{p=1}^P (\mu_{\theta}^p - \mu_{\theta})^2 .$$

Here,  $\mu_{\theta}^p$  is the statistic of interest computed on each plausible value  $p$  and  $\mu_{\theta}$  is the mean statistic based on all  $P$  plausible values.

The estimate of the total variance  $TV_{\mu}$ , consisting of sampling variance and imputation variance, can be computed as

$$TV_{\mu} = SV_{\mu} + (1 + \frac{1}{P}) IV_{\mu} .$$

The estimate of the final standard error  $SE_{\mu}$  is equal to

$$SE_{\mu} = \sqrt{TV_{\mu}} .$$

Table 13.3 shows the average scale scores as well as their sampling and overall standard errors. It also records the number of students that were assessed in each country. The comparison between sampling and combined standard error shows that most of the error was due to sampling and that only a small proportion could be attributed to measurement error.

Use of the IEA IDB Analyzer for estimating civic knowledge standard errors allowed both sampling and imputation errors to be automatically calculated and combined. Alternative applications, such as the software package WESVAR (Westat, 2007) or the ACER SPSS Replicates Module, have similar features that make it possible to take the imputation error into account when analyzing the plausible values.

## Reporting of differences

Significance tests were conducted for:

- Differences in population estimates between countries (multiple comparisons);
- Differences between a country and the international average;
- Differences in population estimates between subgroups within countries; and
- Differences between population estimates in ICCS and in CIVED (estimation of changes).

Differences between two score averages (or percentages)  $a$  and  $b$  were considered significant ( $p < 0.05$ ) when the test statistic  $t$  was greater than the critical value, 1.96. The calculation of  $t$  was conducted by dividing the difference by its standard error,  $SE_{dif\_ab}$ :

$$t = \frac{(a-b)}{SE_{dif\_ab}} .$$

In the case of differences between score averages from independent samples (evident, for example, with respect to comparisons of country averages), the standard error of the difference  $SE_{dif\_ab}$  can be computed as:

$$SE_{dif\_ab} = \sqrt{SE_a^2 + SE_b^2} .$$

Here,  $SE_a$  and  $SE_b$  are the standard errors of the means from the two independent samples  $a$  and  $b$ .





Table 13.3: National averages for civic knowledge with standard deviations, sampling, and overall errors

Country	Civic Knowledge Score	Sampling Error	Combined Standard Error	Number of Assessed Students
Austria	503	3.90	3.98	3,385
Belgium (Flemish)	514	4.65	4.67	2,968
Bulgaria	466	5.03	5.03	3,257
Chile	483	3.48	3.54	5,192
Chinese Taipei	559	2.40	2.44	5,167
Colombia	462	2.92	2.95	6,206
Cyprus	453	2.38	2.41	3,194
Czech Republic	510	2.36	2.38	4,630
Denmark	576	3.51	3.59	4,509
Dominican Republic	380	2.41	2.41	4,589
England	519	4.39	4.39	2,921
Estonia	525	4.49	4.54	2,743
Finland	576	2.33	2.39	3,307
Greece	476	4.37	4.39	3,153
Guatemala	435	3.77	3.78	4,002
Indonesia	433	3.43	3.43	5,068
Ireland	534	4.51	4.56	3,355
Italy	531	3.25	3.29	3,366
Korea, Republic of	565	1.87	1.92	5,254
Latvia	482	3.97	3.99	2,761
Liechtenstein	531	3.22	3.31	357
Lithuania	505	2.74	2.84	3,902
Luxembourg	473	2.20	2.23	4,852
Malta	490	4.30	4.45	2,143
Mexico	452	2.76	2.79	6,576
New Zealand	517	4.94	4.97	3,979
Norway	515	3.29	3.41	3,013
Paraguay	424	3.41	3.41	3,399
Poland	536	4.64	4.66	3,249
Russian Federation	506	3.76	3.77	4,295
Slovak Republic	529	4.39	4.49	2,970
Slovenia	516	2.64	2.65	3,070
Spain	505	4.11	4.13	3,310
Sweden	537	3.08	3.10	3,464
Switzerland	531	3.74	3.78	2,924
Thailand	452	3.64	3.65	5,263
<b>Countries not meeting sampling requirements</b>				
Hong Kong SAR	554	5.65	5.68	2,902
Netherlands	494	7.54	7.62	1,969



The formula for calculating the standard error provided above is only suitable when the subsamples being compared are independent. Because subgroups (e.g., gender groups) within countries are typically not independent samples, the difference between statistics for subgroups of interest and the standard error of the difference was derived through use of a jackknife replication that involved the following formula:

$$SE_{dif\_ab} = \sqrt{\sum_{i=1}^{75} ((a^i - b^i) - (a - b))^2}.$$

Here,  $a$  and  $b$  represent the averages (or percentages) in each of the two subgroups for the fully weighted sample, and  $a^i$  and  $b^i$  are those for the replicate samples.

In the case of differences in civic knowledge scores between dependent subsamples, the standard error of the differences with ( $P = 5$ ) plausible values was calculated on the basis of this formula:

$$SE_{dif\_ab} = \sqrt{\left[ \sum_{p=1}^P \left( \sum_{i=1}^{75} ((a_p^i - b_p^i) - (a_p - b_p))^2 \right) / P \right] + \left[ \left( 1 + \frac{1}{P} \right) \frac{\sum_{p=1}^P ((a_p - b_p) - (\bar{a}_p - \bar{b}_p))^2}{P-1} \right]}.$$

Here,  $a_p$  and  $b_p$  represent the weighted subgroup averages in Groups  $a$  and  $b$  for each of the  $P$  plausible values,  $a_p^i$  and  $b_p^i$  are the subgroup averages within replicate samples for each of the  $P$  plausible values, and  $\bar{a}_p$  and  $\bar{b}_p$  are the means of the two weighted subgroup averages across the  $P$  plausible values.

Comparison of the country means  $c$  with the overall ICCS average  $i$  necessitated accounting for the fact that the country being considered had contributed to the international standard error. This was done by calculating the standard error  $SE_{dif\_ic}$  of the difference between the overall ICCS average and the country average as

$$SE_{dif\_ic} = \sqrt{\frac{((N-1)^2 - 1)SE_c^2 + \sum_{k=1}^N SE_k^2}{N}},$$

where  $SE_c$  is the sampling standard error for country  $c$ , and where  $SE_k$  is the sampling error for  $k^{th}$  of  $N$  participating countries. This formula was used to determine the statistical significance of differences due to sampling error between countries and the ICCS averages of the questionnaire scales throughout the ICCS reports.

Although the above formula was sufficient for the questionnaire scale scores, it was necessary to also take the imputation component of standard errors for countries into account when comparing the test score averages of a country with the overall ICCS average. The imputation variance component of standard errors  $SE_{i\_dif\_ic}^2$  was computed as

$$SE_{i\_dif\_ic}^2 = \sqrt{\left( 1 + \frac{1}{P} \right) \text{var}(d_1, \dots, d_p, \dots, d_5)},$$

where  $d_p$  is the difference between the overall ICCS and the country mean for the plausible value  $p$ .

The final standard error ( $SE_{a\_dif\_ic}$ ) of the difference between ICCS country test scores and the ICCS average was computed as

$$SE_{a\_dif\_ic} = \sqrt{SE_{dif\_ic}^2 + SE_{i\_dif\_ic}^2}.$$



The ICCS international report also included comparisons of test results between ICCS and the CIVED survey of 1999. Because the process of equating the tests across the cycles introduced some additional error into the calculation of any test statistic, an equating error term was added to the formula for the standard error of the difference between country averages.

During testing of the difference of a statistic between the two assessments, the standard error of the difference was computed as follows:

$$SE(\mu_{ICCS} - \mu_{CIVED}) = \sqrt{SE_i^2 + SE_j^2 + EqErr^2}$$

Here,  $\mu$  can be any statistic in units on the ICCS–CIVED link scale (mean, percentile, gender difference, but not percentages) and  $SE_i$  and  $SE_j$  are the respective standard errors of this statistic from the two surveys.  $EqErr$  denotes the equating error that reflects the uncertainty in the link between both assessments, which was equal to 0.65 score points on the link scale (see Chapter 11 for the calculation of the equating error). Because the link scale scores were maximum likelihood estimates and not plausible values, there was no need to provide for imputation error when computing the standard errors of the differences in civic content knowledge between 1999 and 2009.

## Multiple regression modeling

Both single-level multiple regression models and hierarchical linear modeling were used to report ICCS data. The criterion variables were indicators of civic and citizenship learning outcomes, such as civic knowledge, students' interest in political and social issues, and expected electoral participation as an adult. Predictor variables were student background variables, characteristics, and behaviors, as well as school context factors.

When conducting multiple regression models, analysts regress the criterion variable  $Y_i$  on a set of  $k$  predictors  $X_{1i} \dots X_{ki}$ , with  $\alpha$  being the intercept,  $\varepsilon_i$  the unexplained part of the model (residual), and  $\beta$  the regression coefficients:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

In ICCS, multiple regression modeling was used mainly to review the influence of family background, and the unstandardized regression coefficients and the variance explained by the model were reported in order to show the effects for each predictor and the overall explanatory power of the model. Jackknife replication using tailored SPSS macros was used to estimate the standard errors for the multiple regression model parameters (unstandardized coefficients and estimates of explained variance).

In order to estimate the unique contribution of each set of predictors to the variance explanation of the model as well as the proportion of variance explained by more than one set of predictors, different linear regression models were computed. For each set of  $j$  with  $m$  predictor variables, one predictor variable was left out of the model. The difference in variance explanation for the full model and the model without a certain set of predictors showed the unique contribution this factor made with respect to explaining variance in the criterion variable. The variance uniquely explained for predictor variable set  $j$  ( $r_{u-j}^2$ ) was computed as

$$r_{u-j}^2 = r_n^2 - r_{n-m}^2,$$

where  $r_n^2$  is the  $R$  square for the model and  $r_{n-m}^2$  is the  $R$  square for the regression model without the  $m$  variables in predictor block  $j$ .

The unique contribution of predictor set  $j$  to the explained variance in the predictor variable  $Y_i$  was then expressed in percentages:

$$UVC_j = r_{u-j}^2 \times 100.$$



The joint explained variance contribution reflects the proportion of variance explained by more than one of  $k$  sets of predictors. The proportion of variance explained by more than one set of predictors JVC<sub>j</sub> was computed as

$$JVC_j = (r_n^2 \times 100) - \sum_{k=1}^k UVC_j.$$

### Hierarchical linear modeling

Hierarchical (or multilevel) linear regression models (Raudenbush & Bryk, 2002), were estimated in order to take school or classroom context effects into account in which students were nested within classrooms. In most of the country samples, the classroom level was equivalent to the school level because typically only one classroom was selected within each school. Therefore, as with other IEA studies, it was not possible to separately estimate, in the analyses presented in the ICCS reports the variance due to the classroom and school levels.

A hierarchical regression model with  $i$  students nested in  $j$  clusters (schools, classrooms) can be estimated as

$$Y_{ij} = \alpha_j + X_{ij}^n \beta_{ij} + X_j^m \beta_j + U_{0j} + \varepsilon_{ij},$$

where  $Y_{ij}$  are the criterion variables,  $X_{ij}^n$  is a vector of student-level variables, with its corresponding vector of regression coefficients  $\beta_{ij}$ , and  $X_j^m$  is a school- (or classroom-) level variable with its corresponding vector of regression coefficients  $\beta_j$ .  $U_{0j}$  is the residual term at the level of the cluster (school or classroom), and  $\varepsilon_{ij}$  is the student-level residual. Both residual terms are assumed to have a mean of 0 and variance that is normally distributed at each level.

The explained variance in hierarchical linear models has to be estimated for each level separately, with the estimate based on a comparison of each prediction model with the baseline (“null”) model (or ANOVA model) without any predictor variables. Thus:

$$Y_{ij} = \alpha_j + U_{0j}^{null} + U_{ij}^{null}.$$

The residual term  $U_{0j}^{null}$  provides an estimate of the variance in  $Y_{ij}$  between  $j$  clusters, and  $\varepsilon_{ij}^{null}$  is an estimate of the variance between  $i$  students within clusters. The intra-class correlation  $IC$ , which reflects the proportion of variance between clusters (in this case, schools), can be computed from these estimates as

$$IC = \frac{U_{0j}^{null}}{U_{0j}^{null} + \varepsilon_{ij}^{null}}.$$

In order to provide a comparable baseline model for the ICCS multilevel analysis, the “null” model estimated. This model is the one from which students with missing data were excluded after “missing treatment” procedures had been completed (see section on missing treatment below). The explained variance at the school level  $EV_j$  was computed as

$$EV_j = \left(1 - \frac{U_{0j}^{null}}{U_{0j}^{null} + \varepsilon_{ij}^{null}}\right) \times 100,$$

and the explained variance at the student level  $EV_{ij}$  was computed as

$$EV_{ij} = \left(1 - \frac{\varepsilon_{ij}^{null}}{U_{0j}^{null} + \varepsilon_{ij}^{null}}\right) \times 100.$$

Because multilevel modeling takes the hierarchical structure of the cluster sample into account, HLM standard errors that took both sampling and imputation errors into account were reported. Data were weighted (with normalized weights) at both levels.



The software package HLM 6.08 (Raudenbush, Bryk, Cheong, & Congdon, 2004) was used to estimate all hierarchical models. Countries that did not meet IEA sample participation requirements were excluded from the analyses, as were countries where there were fewer than 50 schools. The countries to which these provisos applied were Hong Kong SAR, Liechtenstein, Luxembourg, and the Netherlands.

In most countries, one intact classroom per school was sampled, which made it impossible to disentangle classroom- and school-level variance. In two small countries, Cyprus and Malta, two classrooms were sampled in each school; a few other countries had only small numbers of schools with more than one classroom. These differences in sampling design need to be taken into account when interpreting the results of the multilevel analyses of ICCS data.

### Missing data treatment in ICCS multivariate analyses

Multivariate analysis is more prone to missing data problems than are other forms of analysis. A larger number of cases tend to be excluded if the analysis uses only those records that have complete information for all variables. Generally, there are two possible sources of missing data: (1) no questionnaire data for either the student or their school, and (2) missing data for individual variables.

To address the missing data issue, the small proportion of students without any student questionnaire data were excluded from the analyses and a “dummy variable adjustment” was applied for the remaining students (see Cohen & Cohen, 1975). Mean or median values were assigned to students or schools with missing data, and dummy indicator variables (with 1 indicating a missing value and 0 non-missing values) were added to the analysis.

This treatment was applied to both the student and the school levels during the hierarchical linear modeling. At the student level, the variables were as follows:

- Years of expected further education (EXPEDYR);
- Frequency of students’ use of media information on political and social issues (MEDINF);
- Perception of openness in classroom discussions of political and social issues (OPDISC);
- Reported parental interest in political and social issues (RPARINT);
- Frequency of discussion of political and social issues with parents (PARDISC); and
- Recent voting for class representative or school parliament (SCVOTE).

Treatment of the missing school questionnaire data involved including just one indicator variable for completely missing school data. The variable indicated cases for which any of the above variables had missing values. Only two countries had complete—or almost complete—school data.

Table 13.4 shows the unstandardized regression coefficients for missing indicators for the complete hierarchical linear model (Model 4) reported in the international report (Schulz, Ainley, Fraillon, Kerr, & Losito, 2010, pp. 225ff.). Missing indicators in almost all countries were negatively associated with civic knowledge; however, given the substantial standard errors, the associations were often not significant. No consistent association was found between the missing-school-data indicators and civic knowledge.



Table 13.4: Coefficients of missing indicators in hierarchical linear analysis of ICCS data

Country	Coefficients of Missing Indicators for:						
	Expected years of education	Frequency of media information	Perceptions of openness in classroom discussions	Parental interest in political and social issues	Discussion of political and social issues with parents	Recent voting for class representative / school parliament	Missing school questionnaire
Austria	-58.7 (15.0)	-92.7 (22.4)	-18.5 (14.3)	13.7 (17.1)	-20.4 (28.7)	-36.8 (23.3)	0.3 (9.2)
Belgium (Flemish)	-6.4 (14.1)	0.0 (0.0)	-54.4 (25.9)	-30.0 (18.5)	-2.6 (32.9)	-57.1 (22.3)	-31.6 (18.0)
Bulgaria	-71.7 (19.4)	4.9 (13.5)	-33.5 (13.2)	-20.3 (13.2)	-35.2 (11.4)	-29.7 (9.5)	0.0 (0.0)
Chile	-36.6 (17.8)	-41.5 (19.9)	-38.9 (13.4)	-14.6 (13.0)	4.5 (20.2)	-30.6 (12.3)	3.6 (11.3)
Chinese Taipei	20.2 (24.4)	-42.0 (35.2)	7.1 (38.4)	30.6 (29.1)	3.9 (24.8)	-83.4 (28.3)	2.6 (5.4)
Colombia	-62.1 (20.6)	-16.2 (9.8)	-37.0 (10.3)	-25.0 (5.2)	-20.5 (7.8)	-20.4 (7.3)	-20.2 (24.5)
Cyprus^	-15.1 (16.0)	-37.6 (17.3)	-34.7 (10.7)	-38.3 (9.7)	-22.5 (15.6)	-40.8 (10.7)	15.7 (5.6)
Czech Republic †	-18.0 (10.7)	-41.4 (17.1)	-52.6 (18.1)	-29.9 (10.1)	-30.2 (19.2)	-7.6 (12.0)	12.9 (7.2)
Denmark	-29.1 (9.4)	42.4 (40.3)	-43.1 (18.4)	-10.0 (14.8)	-57.8 (43.4)	-40.1 (14.1)	-15.0 (8.3)
Dominican Republic ~	9.4 (9.8)	-8.5 (5.4)	-11.1 (6.1)	-25.0 (6.1)	-2.6 (4.1)	-16.4 (4.2)	-9.5 (8.7)
England ‡	-24.5 (10.6)	-32.7 (24.0)	-59.2 (25.8)	-42.5 (8.3)	1.9 (18.7)	-55.6 (19.4)	3.9 (9.3)
Estonia	-33.1 (15.2)	-101.6 (61.1)	-8.4 (28.0)	-30.0 (23.2)	7.5 (28.5)	-60.1 (24.6)	0.6 (9.1)
Finland	-39.3 (16.4)	-29.5 (46.1)	-25.5 (14.5)	-14.4 (13.7)	34.5 (34.4)	-5.8 (16.0)	1.1 (23.4)
Greece	-33.9 (17.5)	-14.2 (28.1)	-39.9 (24.8)	-51.7 (32.4)	-22.5 (26.6)	-29.4 (14.1)	-23.4 (11.4)
Guatemala¹	3.2 (21.8)	-8.0 (10.2)	-24.0 (10.9)	-21.6 (10.4)	-18.0 (7.0)	-3.2 (9.5)	31.1 (14.6)
Indonesia	-27.7 (11.7)	-0.1 (13.7)	-6.8 (9.8)	-33.5 (8.2)	-9.5 (9.5)	-9.5 (9.8)	26.8 (23.1)
Ireland	-28.6 (8.5)	-54.2 (29.5)	-20.8 (16.3)	-19.4 (9.5)	-44.4 (20.2)	-31.3 (13.7)	-11.5 (15.4)
Italy	-26.4 (24.8)	33.8 (34.2)	-53.7 (28.0)	1.5 (27.6)	-65.2 (23.6)	-3.4 (16.5)	-11.0 (22.3)
Korea, Republic of¹	-37.3 (20.9)	-132.2 (58.4)	-71.6 (26.8)	3.1 (35.7)	21.8 (44.7)	-28.6 (16.5)	-2.3 (6.7)
Latvia	-53.8 (13.9)	11.9 (27.3)	-10.4 (21.7)	-5.5 (19.6)	-17.4 (21.2)	-39.4 (21.0)	-32.7 (14.5)
Lithuania	-51.6 (35.2)	4.4 (32.9)	41.1 (30.9)	-45.4 (13.4)	-5.5 (30.7)	-92.8 (21.7)	-4.1 (12.5)
Malta^	-14.4 (15.4)	-42.2 (34.0)	-6.7 (25.1)	-28.1 (18.2)	-45.0 (20.1)	-32.7 (18.7)	-79.4 (16.8)
Mexico	-49.2 (12.0)	-2.5 (22.2)	-20.2 (16.6)	-39.9 (10.2)	-12.6 (9.2)	-39.7 (9.0)	13.2 (12.6)
New Zealand	-76.2 (17.3)	-1.1 (27.0)	-36.5 (19.6)	-30.2 (10.9)	-6.1 (22.8)	-33.0 (13.9)	2.3 (10.4)
Norway †	-22.2 (14.5)	-22.4 (24.1)	-9.6 (21.2)	-24.8 (8.3)	-8.6 (21.0)	-27.7 (15.5)	12.8 (9.9)
Paraguay¹ ~	-71.1 (27.4)	-6.6 (9.3)	-15.7 (8.6)	-11.3 (11.2)	-13.4 (7.5)	-23.9 (7.2)	-9.2 (14.9)
Poland	5.1 (40.5)	-125.2 (52.4)	43.9 (47.3)	-19.8 (21.4)	22.1 (32.5)	-38.3 (48.2)	0.0 (0.0)





Table 13.4: Coefficients of missing indicators in hierarchical linear analysis of ICCS data (contd.)

Country	Coefficients of Missing Indicators for:						
	Expected years of education	Frequency of media information	Perceptions of openness in classroom discussions	Parental interest in political and social issues	Discussion of political and social issues with parents	Recent voting for class representative / school parliament	Missing school questionnaire
Russian Federation	-56.6 (16.8)	6.0 (11.7)	-49.4 (14.8)	-11.2 (23.8)	-29.2 (11.9)	-18.4 (12.4)	-29.5 (12.7)
Slovak Republic <sup>2</sup>	-36.1 (32.4)	-69.8 (35.0)	-24.4 (33.0)	-32.7 (15.8)	26.3 (23.8)	-46.5 (19.7)	19.5 (18.9)
Slovenia	-30.6 (18.8)	-19.9 (21.1)	-13.1 (37.7)	-56.1 (17.4)	-2.3 (20.4)	-37.8 (16.6)	-13.2 (9.0)
Spain	-35.1 (31.7)	-137.5 (34.6)	-32.9 (28.2)	-40.2 (14.9)	28.2 (27.1)	-36.2 (18.7)	5.6 (9.1)
Sweden	-18.9 (8.1)	-28.2 (52.0)	24.7 (25.9)	-34.2 (10.8)	10.2 (24.7)	-17.2 (13.2)	4.8 (8.2)
Switzerland †	-22.2 (9.3)	3.6 (31.4)	12.4 (25.9)	-21.2 (15.4)	-46.0 (27.9)	16.3 (22.5)	-2.8 (9.9)
Thailand	16.4 (21.9)	-30.1 (23.0)	4.8 (18.9)	-22.4 (11.4)	-21.5 (19.6)	-26.3 (11.2)	-0.3 (11.7)
ICCS average	-30.4 (3.4)	-30.3 (5.3)	-21.1 (4.0)	-22.9 (2.9)	-11.7 (4.1)	-31.9 (3.1)	-4.1 (2.3)

**Notes:**

( ) Standard errors appear in parentheses. Coefficients significant at  $p < .05$  in bold.

Table 13.5 shows (for the countries included in the multilevel analysis) the numbers of students assessed in ICCS and the respective percentages of students included in the analysis after completion of the missing treatment. On average, across countries, 98 percent of assessed students were included in the analysis post treatment. Inclusion percentages across the participating countries ranged from 93 percent in Austria to almost 100 percent in the Slovak Republic.

*Table 13.5: ICCS students included in hierarchical linear models*

Country	Total Number of Assessed Students	Total Number of Students in Analysis	Percentage of Students in Analysis
Austria	3,385	3,152	93.1
Belgium (Flemish)	2,968	2,891	97.4
Bulgaria	3,257	3,207	98.5
Chile	5,192	5,131	98.8
Chinese Taipei	5,167	5,104	98.8
Colombia	6,204	6,151	99.1
Cyprus	3,194	3,032	94.9
Czech Republic	4,630	4,582	99.0
Denmark	4,508	4,295	95.3
Dominican Republic	4,589	4,431	96.6
England	2,916	2,788	95.6
Estonia	2,743	2,681	97.7
Finland	3,307	3,263	98.7
Greece	3,153	3,118	98.9
Guatemala	4,002	3,966	99.1
Indonesia	5,068	4,981	98.3
Ireland	3,355	3,288	98.0
Italy	3,366	3,323	98.7
Korea, Republic of	5,254	5,234	99.6
Latvia	2,761	2,710	98.2
Lithuania	3,902	3,850	98.7
Malta	2,143	2,115	98.7
Mexico	6,576	6,521	99.2
New Zealand	3,979	3,833	96.3
Norway	3,013	2,903	96.3
Paraguay	3,399	3,338	98.2
Poland	3,249	3,216	99.0
Russian Federation	4,295	4,265	99.3
Slovak Republic	2,970	2,957	99.6
Slovenia	3,070	2,988	97.3
Spain	3,309	3,264	98.6
Sweden	3,464	3,312	95.6
Switzerland	2,924	2,836	97.0
Thailand	5,263	5,228	99.3
<b>Overall sample</b>	<b>130,575</b>	<b>127,954</b>	<b>98.0</b>





During the multiple regression analysis of expected electoral and active political participation reported in Chapter 8 of the extended (international) ICCS report (Schulz et al., 2010b, pp. 235ff), an approach similar to that used for the multilevel analysis of civic knowledge was used to account for missing data. However, although civic knowledge scores were at hand for all assessed students during the multivariate analysis of the two questionnaire scales (expected electoral and active political participation), scores for many students for these two dependent variables were not available.

On average, across countries, nine percent of students did not have complete data for all variables in the model; in two countries (the Dominican Republic and Paraguay), the percentages were considerably higher—above 20 percent. For 8 of the 11 predictor variables, missing values were substituted with means (for continuous variables) and medians (for categorical variables), and eight dummy indicators were added to the set of predictor variables. The results for these additional variables are not included in the tables.

Missing indicators for a number of student variables were computed during the multiple regression analysis. These variables were as follows:

- Sense of citizenship self-efficacy;
- Students' internal political efficacy;
- Students' interest in political and social issues;
- Students' trust in civic institutions;
- Students' reported parental interest in political and social issues;
- Students' past or current participation in civic activities in the community;
- Students' past or current participation in civic activities at school; and
- Students' support for political parties.

Missing treatment was not done for those variables in the model that showed lower proportions of missing data.

Table 13.6 shows the unstandardized regression coefficients for the missing indicators for the multiple regression analysis of expected electoral participation; Table 13.7 shows those for the analysis of expected political participation. Missing indicators tended to be negatively associated with expected electoral participation. Most of the coefficients were not, however, significant.

Table 13.8 records the numbers of students included in both regression models in comparison with those that were assessed in ICCS. For both regression models, about 93 percent of cases, on average, remained in the analysis (the range across countries was 70% to 99%). In two countries (the Dominican Republic and Paraguay), more than 15 percent of the respective samples were excluded. The results for these countries are annotated in the reporting tables.



Table 13.6: Coefficients of missing indicators multiple regression analysis of expected electoral participation

Country	Coefficients of Missing Indicators for:								Participation at school	Support for political parties
	Citizenship self-efficacy	Internal political efficacy	Interest in political and social issues	Trust in civic institutions	Parental interest	Participation in community	Participation at school	Support for political parties		
Austria	-3.2 (3.2)	-3.2 (1.8)	-4.4 (2.1)	-4.4 (2.1)	-1.6 (2.0)	-0.8 (1.6)	0.4 (1.5)	-4.4 (2.1)		
Belgium (Flemish)	17.9 (0.5)	-6.2 (0.7)	-3.8 (3.0)	-3.8 (3.0)	-0.1 (2.2)	-3.5 (6.6)	-0.4 (4.8)	-3.8 (3.0)		
Bulgaria	0.2 (2.9)	5.8 (4.6)	-1.8 (2.6)	-1.8 (2.6)	-0.7 (2.0)	-1.7 (2.5)	0.3 (2.5)	-1.8 (2.6)		
Chile	-7.3 (5.2)	-0.2 (2.7)	1.4 (2.1)	1.4 (2.1)	-2.7 (1.6)	-0.7 (1.5)	-1.4 (1.6)	1.4 (2.1)		
Chinese Taipei	3.7 (3.2)	-5.8 (0.3)	1.5 (2.2)	1.5 (2.2)	-1.8 (5.0)	1.0 (1.9)	-5.1 (8.5)	1.5 (2.2)		
Colombia	1.3 (1.5)	-2.0 (2.5)	-0.6 (2.3)	-0.6 (2.3)	-0.2 (0.5)	0.9 (1.3)	-3.0 (1.6)	-0.6 (2.3)		
Cyprus	1.3 (2.1)	3.1 (2.4)	0.7 (1.7)	0.7 (1.7)	0.4 (0.9)	-4.2 (1.5)	-0.1 (1.6)	0.7 (1.7)		
Czech Republic	3.4 (1.4)	0.9 (1.7)	-1.9 (2.5)	-1.9 (2.5)	-1.5 (1.2)	0.9 (1.9)	-2.9 (3.6)	-1.9 (2.5)		
Denmark	4.2 (3.0)	-1.8 (1.9)	-1.2 (2.0)	-1.2 (2.0)	0.3 (1.1)	4.3 (2.6)	-2.7 (2.5)	-1.2 (2.0)		
Dominican Republic	-0.3 (2.0)	-1.5 (1.2)	-1.7 (2.7)	-1.7 (2.7)	-0.1 (0.7)	-0.9 (1.0)	-0.7 (0.8)	-1.7 (2.7)		
England	1.1 (3.4)	-3.7 (1.9)	0.7 (1.2)	0.7 (1.2)	-1.3 (0.7)	-2.4 (2.4)	-2.1 (3.5)	0.7 (1.2)		
Estonia	-4.9 (4.8)	-2.6 (4.0)	-5.8 (3.1)	-5.8 (3.1)	-1.9 (2.1)	3.0 (3.8)	0.2 (3.3)	-5.8 (3.1)		
Finland	1.8 (3.2)	-1.2 (2.1)	-0.5 (2.2)	-0.5 (2.2)	-2.5 (1.1)	2.4 (3.2)	-1.7 (3.2)	-0.5 (2.2)		
Greece	3.4 (3.1)	-0.7 (2.9)	2.1 (2.3)	2.1 (2.3)	-2.8 (2.4)	0.4 (2.8)	3.3 (2.7)	2.1 (2.3)		
Guatemala	0.1 (1.1)	-0.4 (1.4)	2.7 (1.6)	2.7 (1.6)	0.2 (0.7)	-1.6 (1.3)	1.0 (1.6)	2.7 (1.6)		
Hong Kong SAR	5.2 (1.8)	-13.2 (1.8)	-0.6 (4.5)	-0.6 (4.5)	-1.7 (1.6)	-14.0 (0.6)	-0.2 (3.2)	-0.6 (4.5)		
Indonesia	-1.3 (2.6)	1.8 (1.8)	-1.0 (1.5)	-1.0 (1.5)	-1.8 (1.5)	-0.7 (1.3)	-0.8 (1.4)	-1.0 (1.5)		
Ireland	-1.2 (3.0)	-3.5 (3.9)	-1.4 (2.3)	-1.4 (2.3)	-1.2 (1.3)	-1.2 (2.2)	3.3 (2.8)	-1.4 (2.3)		
Italy	-2.9 (3.8)	-0.7 (3.1)	-3.0 (1.9)	-3.0 (1.9)	1.3 (3.7)	-9.3 (0.9)	0.8 (2.6)	-3.0 (1.9)		
Korea, Republic of	7.0 (2.2)	-2.0 (4.1)	-1.0 (2.5)	-1.0 (2.5)	0.1 (3.9)	-1.1 (1.6)	-0.7 (4.3)	-1.0 (2.5)		
Latvia	0.2 (2.1)	4.6 (1.9)	3.7 (1.7)	3.7 (1.7)	-2.6 (2.3)	-2.8 (1.8)	-11.7 (3.5)	3.7 (1.7)		
Liechtenstein	2.0 (1.7)	0.0 (0.0)	-2.5 (1.0)	-2.5 (1.0)	-3.6 (3.8)	0.0 (0.0)	-4.9 (4.0)	-2.5 (1.0)		
Lithuania	-7.3 (6.8)	-1.2 (3.3)	-1.4 (6.5)	-1.4 (6.5)	-2.4 (1.7)	-1.2 (4.1)	1.0 (3.3)	-1.4 (6.5)		
Luxembourg	3.9 (1.5)	-1.5 (3.0)	-4.4 (1.5)	-4.4 (1.5)	-2.2 (0.7)	-2.6 (3.4)	1.6 (2.8)	-4.4 (1.5)		
Malta	1.2 (5.8)	-4.4 (4.3)	-4.3 (4.6)	-4.3 (4.6)	1.4 (3.3)	-15.0 (20.8)	-4.4 (2.0)	-4.3 (4.6)		
Mexico	1.2 (1.6)	-0.7 (2.2)	-0.9 (2.1)	-0.9 (2.1)	1.0 (1.0)	-1.6 (1.6)	0.9 (1.8)	-0.9 (2.1)		
Netherlands	-1.0 (8.3)	3.2 (3.0)	2.1 (7.6)	2.1 (7.6)	1.2 (2.1)	-18.0 (1.5)	21.8 (1.7)	2.1 (7.6)		





Table 13.6: Coefficients of missing indicators multiple regression analysis of expected electoral participation (contd.)

Country	Coefficients of Missing Indicators for:							
	Citizenship self-efficacy	Internal political efficacy	Interest in political and social issues	Trust in civic institutions	Parental interest	Participation in community	Participation at school	Support for political parties
New Zealand	<b>-2.6</b> (1.2)	-1.1 (3.2)	1.1 (3.1)	1.1 (3.1)	0.0 (1.2)	0.4 (1.0)	<b>-3.3</b> (1.6)	1.1 (3.1)
Norway	-2.3 (1.6)	-3.2 (3.2)	1.5 (2.0)	1.5 (2.0)	<b>-2.6</b> (1.2)	0.3 (2.2)	<b>4.6</b> (2.1)	1.5 (2.0)
Paraguay	0.5 (2.1)	-1.5 (2.2)	-0.8 (1.9)	-0.8 (1.9)	0.0 (1.0)	-0.6 (1.0)	0.1 (1.4)	-0.8 (1.9)
Poland	2.5 (2.1)	-1.2 (3.3)	-5.4 (3.6)	-5.4 (3.6)	0.5 (1.5)	1.2 (4.0)	3.3 (3.1)	-5.4 (3.6)
Russian Federation	2.1 (1.6)	1.3 (2.1)	-0.6 (1.7)	-0.6 (1.7)	<b>6.4</b> (2.2)	-0.3 (1.4)	-2.4 (2.2)	-0.6 (1.7)
Slovak Republic	3.3 (2.3)	-0.7 (2.0)	-2.1 (2.0)	-2.1 (2.0)	-2.4 (1.7)	-2.7 (4.0)	4.1 (2.2)	-2.1 (2.0)
Slovenia	1.3 (3.6)	-0.4 (2.5)	2.8 (2.3)	2.8 (2.3)	-1.3 (1.4)	-1.0 (2.9)	2.1 (4.0)	2.8 (2.3)
Spain	-2.5 (1.9)	0.0 (3.2)	-0.9 (2.2)	-0.9 (2.2)	-1.1 (1.6)	1.8 (1.7)	-0.6 (2.0)	-0.9 (2.2)
Sweden	1.8 (2.0)	2.1 (2.7)	<b>2.2</b> (1.0)	<b>2.2</b> (1.0)	<b>-2.2</b> (1.1)	-1.1 (1.1)	0.2 (1.7)	<b>2.2</b> (1.0)
Switzerland	-1.5 (3.7)	1.5 (3.2)	-0.4 (1.7)	-0.4 (1.7)	<b>2.5</b> (1.2)	-1.8 (2.8)	1.2 (4.3)	-0.4 (1.7)
Thailand	4.8 (3.0)	-0.7 (2.8)	<b>-6.5</b> (2.9)	<b>-6.5</b> (2.9)	0.1 (1.6)	1.0 (1.9)	-0.7 (1.3)	<b>-6.5</b> (2.9)
ICCS average	1.0 (0.5)	-1.1 (0.4)	-1.0 (0.5)	-1.0 (0.5)	<b>-0.7</b> (0.3)	<b>-1.9</b> (0.7)	0.0 (0.5)	<b>-1.0</b> (0.5)

**Notes:**

( ) Standard errors appear in parentheses.

\* Statistically significant (  $p < 0.05$  ) coefficients in **bold**.

Unstandardized regression coefficients (standard errors in brackets) for missing indicators of ... \*

Table 13.7: Coefficients of missing indicators multiple regression analysis of expected political participation

Country	Coefficients of Missing Indicators for:								Participation at school	Support for political parties
	Citizenship self-efficacy	Internal political efficacy	Interest in political and social issues	Trust in civic institutions	Parental interest	Participation in community	Participation at school	Support for political parties		
Austria	-3.0 (0.6)	1.6 (0.4)	1.6 (0.9)	1.6 (0.9)	-1.9 (0.8)	1.5 (0.4)	-0.4 (1.0)	1.6 (0.9)		
Belgium (Flemish)	-3.4 (2.9)	5.3 (0.3)	-0.3 (0.8)	-0.3 (0.8)	2.3 (2.0)	13.1 (5.0)	-2.6 (1.4)	-0.3 (0.8)		
Bulgaria	1.6 (1.1)	-2.1 (1.9)	2.4 (0.7)	2.4 (0.7)	-0.8 (0.7)	-1.4 (1.8)	-1.6 (1.1)	2.4 (0.7)		
Chile	-0.2 (4.5)	5.7 (1.0)	1.3 (1.7)	1.3 (1.7)	-1.9 (0.7)	3.8 (1.0)	-2.9 (1.0)	1.3 (1.7)		
Chinese Taipei	-0.6 (3.5)	-10.2 (0.2)	2.6 (1.2)	2.6 (1.2)	-4.0 (2.8)	-2.7 (1.0)	10.4 (8.5)	2.6 (1.2)		
Colombia	4.2 (0.8)	-0.8 (0.8)	-1.0 (0.7)	-1.0 (0.7)	0.2 (0.3)	-0.2 (0.7)	-0.5 (1.2)	-1.0 (0.7)		
Cyprus	-3.2 (1.0)	-2.5 (0.9)	1.4 (2.2)	1.4 (2.2)	0.2 (0.8)	-3.1 (1.3)	-0.3 (1.0)	1.4 (2.2)		
Czech Republic	3.4 (0.4)	-2.7 (0.4)	2.0 (0.3)	2.0 (0.3)	-1.7 (0.7)	0.6 (0.7)	2.8 (0.6)	2.0 (0.3)		
Denmark	4.3 (2.7)	0.4 (1.7)	-4.6 (2.3)	-4.6 (2.3)	-0.7 (0.7)	2.2 (3.1)	-4.6 (2.8)	-4.6 (2.3)		
Dominican Republic	2.5 (1.0)	0.8 (0.4)	-1.1 (1.0)	-1.1 (1.0)	0.4 (0.6)	-0.3 (0.9)	0.5 (1.1)	-1.1 (1.0)		
England	-1.1 (3.8)	-0.9 (1.2)	-1.1 (1.0)	-1.1 (1.0)	0.4 (0.6)	-3.9 (1.1)	-3.5 (2.0)	-1.1 (1.0)		
Estonia	-8.5 (2.8)	-7.0 (2.3)	-3.6 (2.5)	-3.6 (2.5)	0.9 (1.1)	4.1 (2.6)	-0.2 (2.1)	-3.6 (2.5)		
Finland	-4.9 (2.6)	0.1 (1.4)	2.8 (1.8)	2.8 (1.8)	-0.4 (0.8)	-0.4 (1.3)	-0.3 (1.4)	2.8 (1.8)		
Greece	-0.7 (1.7)	0.7 (2.1)	0.0 (0.9)	0.0 (0.9)	-1.9 (0.7)	1.3 (1.9)	5.0 (0.9)	0.0 (0.9)		
Guatemala	0.5 (0.6)	2.5 (1.7)	-0.2 (3.0)	-0.2 (3.0)	0.8 (0.7)	-0.8 (0.8)	-1.1 (0.7)	-0.2 (3.0)		
Hong Kong SAR	-1.4 (10.9)	-11.1 (3.2)	-3.2 (0.5)	-3.2 (0.5)	2.8 (1.4)	-1.7 (0.4)	3.5 (2.7)	-3.2 (0.5)		
Indonesia	1.3 (1.0)	0.5 (0.8)	-2.2 (1.6)	-2.2 (1.6)	1.3 (0.3)	0.8 (0.7)	2.1 (1.0)	-2.2 (1.6)		
Ireland	-0.5 (1.7)	-4.2 (1.3)	2.0 (1.6)	2.0 (1.6)	1.5 (1.0)	-4.1 (1.8)	3.8 (1.0)	2.0 (1.6)		
Italy	-1.6 (0.7)	5.1 (1.0)	3.4 (1.2)	3.4 (1.2)	1.9 (0.4)	0.9 (2.4)	-3.5 (1.2)	3.4 (1.2)		
Korea, Republic of	1.6 (0.9)	0.7 (2.5)	2.9 (1.2)	2.9 (1.2)	4.2 (1.5)	1.6 (1.0)	0.4 (0.6)	2.9 (1.2)		
Latvia	-1.8 (1.1)	-0.9 (0.3)	2.4 (0.3)	2.4 (0.3)	2.8 (0.7)	-0.4 (1.7)	3.2 (5.1)	2.4 (0.3)		
Liechtenstein	-13.6 (1.1)	0.0 (0.0)	7.0 (0.5)	7.0 (0.5)	3.7 (1.5)	0.0 (0.0)	-7.0 (1.7)	7.0 (0.5)		
Lithuania	-11.4 (0.7)	-4.0 (1.7)	2.8 (1.1)	2.8 (1.1)	0.0 (0.3)	4.1 (1.6)	-3.7 (1.4)	2.8 (1.1)		
Luxembourg	2.2 (3.6)	-3.4 (3.2)	-1.7 (0.8)	-1.7 (0.8)	-0.7 (0.6)	0.0 (2.9)	-3.9 (2.7)	-1.7 (0.8)		
Malta	1.5 (1.0)	-0.3 (3.0)	4.7 (1.6)	4.7 (1.6)	4.8 (3.6)	-23.1 (21.8)	5.3 (1.6)	4.7 (1.6)		
Mexico	-0.3 (0.6)	-0.5 (1.7)	1.2 (0.8)	1.2 (0.8)	0.1 (0.8)	1.6 (1.0)	-1.1 (1.2)	1.2 (0.8)		
Netherlands	3.4 (2.2)	1.9 (5.6)	-0.9 (2.2)	-0.9 (2.2)	2.6 (2.4)	-17.2 (2.3)	16.7 (6.4)	-0.9 (2.2)		





Table 13.7: Coefficients of missing indicators multiple regression analysis of expected political participation (contd.)

Country	Coefficients of Missing Indicators for:							
	Citizenship self-efficacy	Internal political efficacy	Interest in political and social issues	Trust in civic institutions	Parental interest	Participation in community	Participation at school	Support for political parties
New Zealand	0.5 (0.9)	0.7 (3.0)	2.6 (2.0)	2.6 (2.0)	0.9 (0.9)	1.1 (1.0)	1.4 (0.6)	2.6 (2.0)
Norway	-0.9 (1.0)	-5.1 (2.0)	1.3 (1.0)	1.3 (1.0)	-0.5 (0.5)	-3.0 (1.1)	-0.2 (2.5)	1.3 (1.0)
Paraguay	-1.0 (0.5)	-1.8 (1.5)	2.2 (0.8)	2.2 (0.8)	0.9 (0.6)	1.6 (0.4)	-1.4 (0.7)	2.2 (0.8)
Poland	3.1 (1.3)	0.8 (1.2)	-1.3 (1.3)	-1.3 (1.3)	0.9 (0.7)	-0.7 (3.1)	3.4 (1.6)	-1.3 (1.3)
Russian Federation	3.6 (2.1)	-4.9 (2.6)	2.0 (0.9)	2.0 (0.9)	1.6 (1.6)	1.6 (1.4)	1.5 (1.7)	2.0 (0.9)
Slovak Republic	2.7 (1.3)	-0.3 (2.6)	-0.9 (0.8)	-0.9 (0.8)	2.4 (0.7)	1.1 (1.0)	1.1 (0.7)	-0.9 (0.8)
Slovenia	6.4 (1.8)	-0.2 (2.5)	6.5 (2.8)	6.5 (2.8)	0.2 (0.8)	-2.9 (0.7)	-3.0 (2.2)	6.5 (2.8)
Spain	-0.4 (1.1)	0.2 (1.9)	3.0 (0.3)	3.0 (0.3)	-0.5 (1.2)	-4.1 (2.1)	5.8 (4.3)	3.0 (0.3)
Sweden	3.5 (1.0)	3.2 (1.6)	3.1 (0.5)	3.1 (0.5)	-0.5 (0.2)	0.5 (1.1)	3.4 (0.8)	3.1 (0.5)
Switzerland	0.7 (0.5)	2.9 (3.3)	-2.9 (1.0)	-2.9 (1.0)	0.6 (0.5)	1.6 (1.1)	-0.2 (1.0)	-2.9 (1.0)
Thailand	-2.8 (1.6)	-0.9 (0.4)	1.9 (0.7)	1.9 (0.7)	2.5 (0.5)	-1.1 (1.0)	2.0 (0.6)	1.9 (0.7)
ICCS average	-0.4 (0.4)	-0.8 (0.3)	1.0 (0.2)	1.0 (0.2)	0.7 (0.2)	-0.7 (0.6)	0.8 (0.4)	1.0 (0.2)

**Notes:**

( ) Standard errors appear in parentheses.

\* Statistically significant (  $p < 0.05$  ) coefficients in **bold**.

Unstandardized regression coefficients (standard errors in brackets) for missing indicators of ... \*

Table 13.8: ICCS students included in multiple regression models

Country	Total Number of Assessed Students	Analysis of Expected Electoral Participation		Analysis of Expected Political Participation	
		Total number of students in analysis	Percentage of students in analysis	Total number of students in analysis	Percentage of students in analysis
Austria	3,385	3,127	92.4	3,130	92.5
Belgium (Flemish)	2,968	2,877	96.9	2,878	97.0
Bulgaria	3,257	2,976	91.4	2,975	91.3
Chile	5,192	4,998	96.3	4,993	96.2
Chinese Taipei	5,167	5,103	98.8	5,103	98.8
Colombia	6,204	5,455	87.9	5,426	87.5
Cyprus	3,194	2,735	85.6	2,722	85.2
Czech Republic	4,630	4,548	98.2	4,544	98.1
Denmark	4,508	4,089	90.7	4,083	90.6
Dominican Republic	4,589	3,287	71.6	3,231	70.4
England	2,916	2,720	93.3	2,721	93.3
Estonia	2,743	2,648	96.5	2,647	96.5
Finland	3,307	3,228	97.6	3,226	97.6
Greece	3,153	2,958	93.8	2,959	93.8
Guatemala	4,002	3,615	90.3	3,604	90.1
Hong Kong SAR	2,902	2,660	91.7	2,660	91.7
Indonesia	5,068	4,717	93.1	4,715	93.0
Ireland	3,355	3,120	93.0	3,120	93.0
Italy	3,366	3,281	97.5	3,276	97.3
Korea, Republic of	5,254	5,207	99.1	5,206	99.1
Latvia	2,761	2,683	97.2	2,686	97.3
Liechtenstein	357	332	93.0	331	92.7
Lithuania	3,902	3,819	97.9	3,816	97.8
Luxembourg	4,852	4,578	94.4	4,574	94.3
Malta	2,143	2,031	94.8	2,031	94.8
Mexico	6,576	5,937	90.3	5,908	89.8
Netherlands	1,964	1,792	91.2	1,793	91.3
New Zealand	3,979	3,631	91.3	3,627	91.2
Norway	3,013	2,674	88.7	2,666	88.5
Paraguay	3,399	2,670	78.6	2,652	78.0
Poland	3,249	3,180	97.9	3,179	97.8
Russian Federation	4,295	4,220	98.3	4,209	98.0
Slovak Republic	2,970	2,940	99.0	2,939	99.0
Slovenia	3,070	2,975	96.9	2,974	96.9
Spain	3,309	3,158	95.4	3,159	95.5
Sweden	3,464	3,282	94.7	3,278	94.6
Switzerland	2,924	2,786	95.3	2,784	95.2
Thailand	5,263	5,141	97.7	5,141	97.7
Overall ICCS sample	140,650	131,178	93.3	130,966	93.1

## Summary

The jackknife repeated replication technique (JRR) was applied in order to allow reporting of sampling errors in ICCS reports. Plausible value methodology was used with respect to reporting civic knowledge scores. This process permitted estimation not only of variance due to sampling but also of imputation variance.

Different types of significance test were used to compare means or percentages between participating countries, with the ICCS average, or between subgroups within the sample. The equating (or link) error was taken into account when averages of civic content knowledge in 2009 were compared with averages of civic content knowledge from the CIVED survey of 1999.

ICCS 2009 data were used to estimate the multiple regression models as well as the hierarchical linear models, and explained variance decomposition was used to estimate the unique contribution of different sets of predictor variables in the multiple regression models. The explained variance at student and school levels was compared separately whenever two-level hierarchical linear models were used.

Missing data problems became more prevalent during multivariate analyses of ICCS data that involved larger numbers of predictor variables. For the reported analyses, data were treated by adding missing indicators and substituting missing values with modes or means. Anyone conducting multivariate analysis of ICCS data needs to take missing data problems into account and should also explore possibilities for applying more advanced methods, including imputation procedures.

## References

- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gonzalez, E. J., & Foy, P. (2000). Estimation of sampling variance. In M. O. Martin, K. D. Gregory, & S. E. Semler (Eds.), *TIMSS 1999: Technical report*. Chestnut Hill, MA: Boston College.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publishers.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Westat (2007). *WesVar ®4.3: User's guide* [computer software]. Rockville MD: Author.
- Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer.



# Appendices

## Appendix A: Organizations and individuals involved in ICCS

### *The international study center and its partner institutions*

The international study center is located at the Australian Council for Educational Research (ACER) and serves as the international study center for ICCS. Center staff at ACER were responsible for designing and implementing the study in close co-operation with the center's partner institutions NFER (National Foundation for Educational Research, Slough, United Kingdom) and LPS (Laboratorio di Pedagogia Sperimentale at the Roma Tre University, Rome, Italy) as well as the IEA Data Processing and Research Center (DPC) and the IEA Secretariat.

#### **Staff at ACER**

John Ainley, *project coordinator*  
Wolfram Schulz, *research director*  
Julian Fraillon, *coordinator of test development*  
Tim Friedman, *project researcher*  
Naoko Tabata, *project researcher*  
Maurice Walker, *project researcher*  
Eva Van De Gaer, *project researcher*  
Anna-Kristin Albers, *project researcher*  
Corrie Kirchhoff, *project researcher*  
Paul Fabian, *project researcher*  
Renee Chow, *data analyst*  
Louise Wenn, *data analyst*

#### **Staff at NFER**

David Kerr, *associate research director*  
Joana Lopes, *project researcher*  
Linda Sturman, *project researcher*  
Jo Morrison, *data analyst*

#### **Staff at LPS**

Bruno Losito, *associate research director*  
Gabriella Agrusti, *project researcher*  
Elisa Caponera, *project researcher*  
Paola Mirti, *project researcher*

### *International Association for the Evaluation of Educational Achievement (IEA)*

IEA provides overall support in coordinating ICCS. The IEA Secretariat in Amsterdam, the Netherlands, is responsible for membership, translation verification, and quality control monitoring. The IEA Data Processing and Research Center (DPC) in Hamburg, Germany, is mainly responsible for sampling procedures and the processing of ICCS data.

#### **Staff at the IEA Secretariat**

Hans Wagemaker, *executive director*  
Barbara Malak, *manager membership relations*  
Paulína Koršňáková, *senior administrative officer*  
Jur Hartenberg, *financial manager*





### **Staff at the IEA Data Processing and Research Center (DPC)**

Heiko Sibberns, *co-director*

Dirk Hastedt, *co-director*

Falk Brese, *ICCS coordinator*

Michael Jung, *researcher*

Olaf Zuehlke, *researcher (sampling)*

Caroline Vandenplas, *researcher (sampling)*

Sabine Meinck, *researcher (sampling)*

Eugenio Gonzalez, *consultant to the Latin American regional module*

### **ICCS project advisory committee (PAC)**

PAC has, from the beginning of the project, advised the international study center and its partner institutions during regular meetings.

### **PAC members**

John Ainley (*chair*), ACER, Australia

Barbara Malak, IEA Secretariat

Heiko Sibberns, IEA Technical Executive Group

John Annette, University of London, United Kingdom

Leonor Cariola, Ministry of Education, Chile

Henk Dekker, University of Leiden, the Netherlands

Bryony Hoskins, Center for Research on Lifelong Learning, European Commission

Rosario Jaramillo, Ministry of Education, Colombia (2006–2008)

Margarita Peña, Colombian Institute for the Promotion of Higher Education (2008–2010)

Judith Torney-Purta, University of Maryland, United States

Lee Wing-On, Hong Kong Institute of Education, Hong Kong SAR

Christian Monseur, University of Liège, Belgium

### **Other project consultants**

Aletta Grisay, University of Liège, Belgium

Isabel Menezes, Porto University, Portugal

Barbara Fratzczak-Rudnicka, Warsaw University, Poland

### **ICCS sampling referee**

Jean Dumais from Statistics Canada in Ottawa was the sampling referee for ICCS. He provided invaluable advice on all sampling-related aspects of the study.

### **National research coordinators (NRCs)**

The national research coordinators (NRCs) played a crucial role in the development of the project. They provided policy- and content-oriented advice on the development of the instruments and were responsible for implementing ICCS in the participating countries.

### **Austria**

*Günther Ogris*

SORA Institute for Social Research and Analysis, Ogris & Hofinger GmbH

### **Belgium (Flemish)**

*Saskia de Groof*

Center of Sociology, Research Group TOR, Free University of Brussels

### **Bulgaria**

*Svetla Petrova*

Center for Control and Assessment of Quality in Education, Ministry of Education and Science



**Chile**

*Marcela Ortiz Guerrero*

Unit of Curriculum and Evaluation, Ministry of Education

**Chinese Taipei**

*Meihui Liu*

Department of Education, Taiwan Normal University

**Colombia**

*Margarita Peña*

Colombian Institute for the Promotion of Higher Education (ICFES)

**Cyprus**

*Mary Koutselini*

Department of Education, University of Cyprus

**Czech Republic**

*Petr Soukup*

Institute for Information on Education

**Denmark**

*Jens Bruun*

Department of Educational Anthropology, The Danish University of Education

**Dominican Republic**

*Ancell Scheker*

Ministry of Education

**England**

*Julie Nelson*

National Foundation for Educational Research

**Estonia**

*Anu Toots*

Tallinn University

**Finland**

*Pekka Kupari*

Finnish Institute for Educational Research, University of Jyväskylä

**Greece**

*Georgia Polydorides*

Department of Early Childhood Education, University of Athens

**Guatemala**

*Luisa Muller Durán*

SINEIE, Ministry of Education

**Hong Kong SAR**

*Lee Wing-On*

Hong Kong Institute of Education

**Indonesia**

*Diah Haryanti*

Ministry of National Education

**Ireland**

*Jude Cosgrove*

Educational Research Centre, St Patrick's College

**Italy**

*Genny Terrinoni*

INVALSI



**Republic of Korea***Tae-Jun Kim*

Korean Educational Development Institute (KEDI)

**Latvia***Andris Kangro*

Faculty of Education and Psychology, University of Latvia

**Liechtenstein***Horst Biedermann*

Institute of Pedagogy, University of Freiburg

**Lithuania***Zivile Urbienie*

National Examination Center

**Luxembourg***Joseph Britz*

Ministry of National Education

*Romain Martin*

University of Luxembourg

**Malta***Raymond Camilleri*

Department of Planning and Development, Ministry of Education

**Mexico***María Concepción Medina*

Ministry of Education

**Netherlands***M. P. C. van der Werf*

GION, University of Groningen

**New Zealand***Kate Lang**Sharon Cox*

Comparative Education Research Unit, Ministry of Education

**Norway***Rolf Mikkelsen*

University of Oslo

**Paraguay***Mirna Vera*

Ministry of Education

**Poland***Krzysztof Kosela*

Institute of Sociology, University of Warsaw

**Russia***Peter Pologevets*

Institution for Education Reforms of the State University Higher School of Economics

**Slovak Republic***Ervin Stava*

Department for International Measurements, National Institute for Certified Educational Measurements (NUCEM)



**Slovenia**

*Marjan Simenc*

University of Ljubljana

**Spain**

*Rosario Sánchez*

Institute of Evaluation, Ministry of Education

**Sweden**

*Marika Sanne*

*Fredrik Lind*

The Swedish National Agency for Education (Skolverket)

**Switzerland**

*Fritz Oser*

Institute of Pedagogy, University of Freiburg

**Thailand**

*Siriporn Boonyananta*

The Office of the Education Council, Ministry of Education

*Somwung Pitiyanuwat*

The Office for National Education Standards and Quality Assessment



## Appendix B: Characteristics of national samples

This appendix describes, for each education system participating in ICCS 2009, the population coverage, exclusion categories, stratification variables, and any deviations from the general ICCS sampling design.

The same sample of schools was selected for the student survey and the teacher survey. However, the school participation status of a school in the student and teacher survey could differ. It was particularly common for school to count as participating in the student survey, but not in the teacher survey; however, the reverse scenario was also possible. If the school participation status in both parts of ICCS 2009 differed, the figures are displayed in two separate tables. If the status counts were identical in both parts, the results are displayed in one combined table.

### B.1. Austria

- Special schools (*Sonderschulen*), elementary schools (*Volksschulen*), technical colleges (*Polytechnische Schulen*), technical and business colleges (*gewerbliche, technische und kunstgewerbliche höhere Schulen*), and commercial colleges and secondary schools (*Oberstufenrealgymnasien*) were all excluded at the school level.
- Explicit stratification was performed by school type (*AHS-Unterstufe*, *Hauptschule* and Other School Type), resulting in three explicit strata.
- Implicit stratification was applied by state (nine levels) in the *AHS-Unterstufe* and *Hauptschule* strata, giving a total of 19 implicit strata.

Table B.1.1: Allocation of student sample in Austria

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
AHS-Unterstufe	43	0	34	3	2	4
Hauptschule	105	0	88	4	2	11
Other School Type	2	0	1	1	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>123</b>	<b>8</b>	<b>4</b>	<b>15</b>

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.1.2: Allocation of teacher sample in Austria

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
AHS-Unterstufe	43	0	12	2	2	27
Hauptschule	105	0	55	2	2	46
Other School Type	2	0	0	0	0	2
<b>Total</b>	<b>150</b>	<b>0</b>	<b>67</b>	<b>4</b>	<b>4</b>	<b>75</b>

**Note:** 37 schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.2. Belgium (Flemish)

- School-level exclusions consisted of schools in the system of education for children with special needs.
- Explicit stratification was performed by school size (large, medium/small), resulting in two explicit strata.
- Implicit stratification was applied by province (five levels), educational network (two levels), and educational stream (two levels), giving a total of 32 implicit strata.
- Two classrooms were sampled in the 10 largest schools (MOS  $\geq 275$ ).
- Schools in the “large” stratum were sampled with equal probabilities.

Table B.2.1: Allocation of student sample in Belgium (Flemish)

School Participation Status— Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Large	40	0	30	7	0	3
Medium/small	120	1	86	21	7	5
Total	160	1	116	28	7	8

Table B.2.2: Allocation of teacher sample in Belgium (Flemish)

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Large	40	0	27	7	0	6
Medium/small	120	1	76	19	6	18
Total	160	1	103	26	6	24

**Note:** 15 schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.3. Bulgaria

- Schools for children with special education needs and schools with fewer than five students in the target grade were excluded at the school level.
- Explicit stratification was performed by region, resulting in 11 explicit strata.
- Implicit stratification was applied by school type (general, profiled, vocational) and size of settlement for school type “general” (small town, medium-size town, large town), giving a total of 54 implicit strata.



Table B.3: Allocation of student sample and teacher sample in Bulgaria

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Region 01	12	1	11	0	0	0
Region 02	24	2	21	1	0	0
Region 03	21	3	18	0	0	0
Region 04	23	0	23	0	0	0
Region 05	17	4	13	0	0	0
Region 06	13	0	13	0	0	0
Region 07	10	2	8	0	0	0
Region 08	10	1	9	0	0	0
Region 09	21	2	19	0	0	0
Region 10	11	1	10	0	0	0
Region 11	13	1	12	0	0	0
<b>Total</b>	<b>175</b>	<b>17</b>	<b>157</b>	<b>1</b>	<b>0</b>	<b>0</b>

#### B.4. Chile

- Special educational needs schools, schools on Easter Island (Rapa Nui) and Juan-Fernández-Islands, and schools that were difficult to access were excluded at the school level.
- Explicit stratification was performed by school administration (municipal, private, subsidized), resulting in three explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), geographical area (north, central, south), and type of education (only primary and lower secondary, primary and secondary), giving a total of 34 implicit strata.
- Over-sampling of private schools. The disproportional allocation was taken into account when sampling weights were computed.

Table B.4: Allocation of student sample and teacher sample in Chile

School Participation Status— Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Municipal	85	1	82	1	0	1
Private	30	0	28	1	1	0
Subsidized	65	1	64	0	0	0
<b>Total</b>	<b>180</b>	<b>2</b>	<b>174</b>	<b>2</b>	<b>1</b>	<b>1</b>



### B.5. Chinese Taipei

- Special education schools and very small schools (fewer than two classes in target grade) were excluded at the school level.
- Explicit stratification was performed by region (north, middle, south, east) and school type (public, private), resulting in eight explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), giving a total of 16 implicit strata.

Table B.5.1: Allocation of student sample in Chinese Taipei

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
North—private	6	0	6	0	0	0
North—public	57	0	55	2	0	0
Middle—private	5	0	5	0	0	0
Middle—public	39	0	39	0	0	0
South—private	3	0	3	0	0	0
South—public	32	0	32	0	0	0
East—private	2	0	2	0	0	0
East—public	6	0	6	0	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>148</b>	<b>2</b>	<b>0</b>	<b>0</b>

Table B.5.2: Allocation of teacher sample in Chinese Taipei

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
North—private	6	0	6	0	0	0
North—public	57	0	53	2	0	2
Middle—private	5	0	5	0	0	0
Middle—public	39	0	36	0	0	3
South—private	3	0	3	0	0	0
South—public	32	0	30	0	0	2
East—private	2	0	2	0	0	0
East—public	6	0	6	0	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>141</b>	<b>2</b>	<b>0</b>	<b>7</b>





### B.6. Colombia

- Night schools, weekend schools, and very small schools (fewer than nine students in the target grade) were excluded at the school level.
- Explicit stratification by hemisphere (Hemisphere A, Hemisphere B) and school schedule (morning, afternoon, full time), resulting in six explicit strata.
- Implicit stratification was applied by urbanization (urban, rural, semi-urban) and school type (public, private), giving a total of 30 implicit strata.

Table B.6.1: Allocation of student sample in Colombia

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Hemisphere A—morning	88	3	81	4	0	0
Hemisphere A—afternoon	50	0	49	0	1	0
Hemisphere A—full time	36	0	33	3	0	0
Hemisphere B—morning	16	0	14	2	0	0
Hemisphere B—afternoon	6	0	5	0	0	1
Hemisphere B—full time	4	0	3	1	0	0
<b>Total</b>	<b>200</b>	<b>3</b>	<b>185</b>	<b>10</b>	<b>1</b>	<b>1</b>

Table B.6.2: Allocation of teacher sample in Colombia

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Hemisphere A—morning	88	3	77	4	0	4
Hemisphere A—afternoon	50	0	47	0	1	2
Hemisphere A—full time	36	0	32	3	0	1
Hemisphere B—morning	16	0	13	2	0	1
Hemisphere B—afternoon	6	0	5	0	0	1
Hemisphere B—full time	4	0	3	1	0	0
<b>Total</b>	<b>200</b>	<b>3</b>	<b>177</b>	<b>10</b>	<b>1</b>	<b>9</b>

**Note:** Four schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.7. Cyprus

- All eligible schools in Cyprus were selected for ICCS 2009.
- There were no school-level exclusions.
- No explicit stratification was performed.
- Implicit stratification was applied by district (Nicosia, Limassol, Larnaca-Famagusta, Paphos) and urbanization (urban, rural), giving a total of eight implicit strata.
- Three classes were sampled in schools with more than six classes; otherwise two classes were sampled where possible.

Table B.7.1: Allocation of student sample in Cyprus

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	68	0	68	0	0	0
Total	68	0	68	0	0	0

Table B.7.2: Allocation of teacher sample in Cyprus

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	68	0	66	0	0	2
Total	68	0	66	0	0	2

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.

### B.8. Czech Republic

- Very small schools (fewer than six students in the target grade), special educational programs, and schools for disabled students were excluded at the school level.
- Explicit stratification was performed by type of education (*Gymnázium*, *Základní škola*), resulting in two explicit strata.
- Implicit stratification was applied by region (14 regions), giving a total of 28 implicit strata.
- All classrooms were sampled in the selected schools.

Table B.8.1: Allocation of student sample in the Czech Republic

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Gymnázium	18	0	16	1	0	1
Základní škola	132	0	105	17	5	5
Total	150	0	121	18	5	6

Table B.8.2: Allocation of teacher sample in the Czech Republic

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Gymnázium	18	0	17	1	0	0
Základní škola	132	0	106	18	5	3
Total	150	0	123	19	5	3

#### B.9. Denmark

- Special schools and schools for severely physically or mentally disabled students were excluded at the school level.
- No explicit stratification was performed.
- Implicit stratification was applied by region (a municipality code number), giving a total of 18 implicit strata.

Table B.9.1: Allocation of student sample in Denmark

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	240	12	121	43	29	35
Total	240	12	121	43	29	35

Table B.9.2: Allocation of teacher sample in Denmark

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	240	12	60	32	21	115
Total	240	12	60	32	21	115

**Note:** 19 schools were regarded as non-participating because the within-school participation rate was below 50%.



### B.10. Dominican Republic

- There were no school-level exclusions.
- Explicit stratification was performed by school type (public, private), resulting in two explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), giving a total of four implicit strata.

Table B.10: Allocation of student sample and teacher sample in the Dominican Republic

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Public	118	2	116	0	0	0
Private	32	2	29	0	0	1
Total	150	4	145	0	0	1

### B.11. England

- Special education schools, very small schools (fewer than 10 students in the target grade) and student-referral units were excluded at the school level.
- No explicit stratification was performed.
- Implicit stratification was applied by GCSE performance (six levels) and school type (comprehensive to 16, comprehensive to 18, independent, other type), giving a total of 24 implicit strata.

Table B.11.1: Allocation of student sample in England

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	160	2	81	24	19	34
Total	160	2	81	24	19	34

Table B.11.2: Allocation of teacher sample in England

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	160	2	77	23	18	40
Total	160	2	77	23	18	40

**Note:** Six schools were regarded as non-participating because the within-school participation rate was below 50%.



### B.12. Estonia

- Schools for adults, special education schools, very small schools (fewer than five students in the target grade), international schools where English is the language of instruction, and Waldorf schools were excluded at the school level.
- Explicit stratification was performed by language (Estonian, Russian, Estonian, and Russian), resulting in three explicit strata.
- Implicit stratification was applied by urbanization (urban, rural), giving a total of six implicit strata.

Table B.12.1: Allocation of student sample in Estonia

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Estonian	116	6	107	2	0	1
Russian	30	3	26	0	1	0
Estonian and Russian*	4	0	3	1	0	0
<b>Total</b>	<b>150</b>	<b>9</b>	<b>136</b>	<b>3</b>	<b>1</b>	<b>1</b>

**Notes:**

One school was regarded as non-participating because the within-school participation rate was below 50%.

\*All students in the four selected schools from this stratum were tested in the Estonian language.

Table B.12.2: Allocation of teacher sample in Estonia

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Estonian	116	6	103	2	0	5
Russian	30	3	23	0	1	3
Estonian and Russian	4	0	3	1	0	0
<b>Total</b>	<b>150</b>	<b>9</b>	<b>129</b>	<b>3</b>	<b>1</b>	<b>8</b>

**Note:** Two schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.13. Finland

- Schools on Åland and special education schools were excluded at the school level.
- Explicit stratification was performed by region (five levels) and urbanization (rural, semi-urban, or rural), resulting in 10 explicit strata.
- No implicit stratification was applied.



Table B.13.1: Allocation of student sample in Finland

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Southern Finland—rural	9	0	8	1	0	0
Southern Finland—semi-urban or urban	76	0	59	11	0	6
Western Finland—rural	9	0	8	0	0	1
Western Finland—semi-urban or urban	34	0	30	4	0	0
Eastern Finland—rural	10	0	9	1	0	0
Eastern Finland—semi-urban or urban	13	1	11	1	0	0
Northern Finland—rural	8	0	7	1	0	0
Northern Finland—semi-urban or urban	17	0	16	0	0	1
Swedish speaking—rural	2	0	2	0	0	0
Swedish speaking—semi-urban or urban	8	0	6	1	0	1
<b>Total</b>	<b>186</b>	<b>1</b>	<b>156</b>	<b>20</b>	<b>0</b>	<b>9</b>

Table B.13.2: Allocation of teacher sample in Finland

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Southern Finland—rural	9	0	8	1	0	0
Southern Finland—semi-urban or urban	76	0	59	10	0	7
Western Finland—rural	9	0	8	0	0	1
Western Finland—semi-urban or urban	34	0	30	4	0	0
Eastern Finland—rural	10	0	9	1	0	0
Eastern Finland—semi-urban or urban	13	1	11	1	0	0
Northern Finland—rural	8	0	7	1	0	0
Northern Finland—semi-urban or urban	17	0	15	0	0	2
Swedish speaking—rural	2	0	2	0	0	0
Swedish speaking—semi-urban or urban	8	0	6	1	0	1
<b>Total</b>	<b>186</b>	<b>1</b>	<b>155</b>	<b>19</b>	<b>0</b>	<b>11</b>

**Note:** Two schools were regarded as non-participating because the within-school participation rate was below 50%.

#### B.14. Greece

- Night schools and schools for children with special needs were excluded at the school level.
- Explicit stratification was performed by school type (public, private) and school location within the “public” stratum, resulting in nine explicit strata.
- Implicit stratification was carried out by region (12 regions), giving a total of 54 implicit strata.
- Unapproved teacher sampling procedures made it necessary to remove Greece from the teacher survey.

Table B.14: Allocation of student sample in Greece

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Public—Athens Center and Piraeus Center	12	0	11	0	0	1
Public—Athens North East and South East	19	0	16	2	1	0
Public—Athens West	16	0	15	1	0	0
Public—Thessaloniki	12	0	10	2	0	0
Public—very large towns	12	0	12	0	0	0
Public—large towns	22	0	21	1	0	0
Public—small towns	40	0	37	2	0	1
Public—rural communities	14	0	13	1	0	0
Private	8	0	6	1	1	0
<b>Total</b>	<b>155</b>	<b>0</b>	<b>141</b>	<b>10</b>	<b>2</b>	<b>2</b>

#### B.15. Guatemala

- Very small schools (fewer than five students in the target grade) and schools that follow an international calendar were excluded at the school level.
- Explicit stratification was performed using the results of a mathematics and reading evaluation conducted in 2006, resulting in five explicit strata.
- Implicit stratification was applied by sector (cooperative, private, public) and urbanization (rural, urban), giving a total of 30 implicit strata.

Table B.15: Allocation of student sample and teacher sample in Guatemala

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Performance Group 1	30	1	29	0	0	0
Performance Group 2	30	0	29	1	0	0
Performance Group 3	30	0	30	0	0	0
Performance Group 4	30	1	28	1	0	0
Performance unknown	30	3	27	0	0	0
<b>Total</b>	<b>150</b>	<b>5</b>	<b>143</b>	<b>2</b>	<b>0</b>	<b>0</b>

### B.16. Hong Kong, SAR

- International schools were excluded at the school level.
- Explicit stratification was performed by finance type (government, aided/caput, direct subsidy scheme), resulting in three explicit strata.
- Implicit stratification was applied by ability (high ability, middle ability, low ability), and region (Hong Kong Island, Kowloon, New Territories), giving a total of 25 implicit strata.

Table B.16.1: Allocation of student sample in Hong Kong, SAR

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Government	10	0	9	0	0	1
Aided/caput	125	0	46	11	1	67
Direct subsidy scheme	15	0	8	1	0	6
<b>Total</b>	<b>150</b>	<b>0</b>	<b>63</b>	<b>12</b>	<b>1</b>	<b>74</b>

Table B.16.2: Allocation of teacher sample in Hong Kong, SAR

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Government	10	0	9	0	0	1
Aided/caput	125	0	58	17	8	42
Direct subsidy scheme	15	0	8	1	0	6
<b>Total</b>	<b>150</b>	<b>0</b>	<b>75</b>	<b>18</b>	<b>8</b>	<b>49</b>

**Note:** Two schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.17. Indonesia

- Schools in Papua and West Papua and very small schools (fewer than 10 students in the target grade) were excluded at the school level.
- No explicit stratification was performed.
- Implicit stratification was applied by school type (junior high, Islamic junior high), school authority (public, private), and national examination score (high performance, medium performance, low performance), giving a total of 12 implicit strata.

Table B.17.1: Allocation of student sample in Indonesia

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	150	8	141	1	0	0
<b>Total</b>	<b>150</b>	<b>8</b>	<b>141</b>	<b>1</b>	<b>0</b>	<b>0</b>



Table B.17.2: Allocation of teacher sample in Indonesia

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	150	8	140	1	0	1
<b>Total</b>	<b>150</b>	<b>8</b>	<b>140</b>	<b>1</b>	<b>0</b>	<b>1</b>

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.

### B.18. Ireland

- Schools located on islands (geographically) were excluded at the school level.
- Explicit stratification was performed by school size ( $\leq 40$  students, 41–80 students, 81–170 students,  $\geq 171$  students) and school sector (community/comprehensive, secondary, vocational), resulting in 10 explicit strata.
- Implicit stratification was applied by SES (in highest SES quarter, in second-highest SES quarter, in third-highest SES quarter, in lowest SES quarter) and gender within all strata except “ $\geq 171$  students” (0% female, 1–45% female, 46–99% female, 100% female), giving a total of 93 implicit strata.
- Two classrooms were sampled in large schools (MOS  $>170$ ); one classroom was sampled in all other schools.
- Schools in the “ $\geq 171$  students” stratum were sampled with equal probabilities.

Table B.18.1: Allocation of student sample in Ireland

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
$\leq 40$ students—community/comprehensive	2	0	2	0	0	0
$\leq 40$ students—secondary	4	0	3	1	0	0
$\leq 40$ students—vocational	6	0	6	0	0	0
41–80 students—community/comprehensive	4	0	4	0	0	0
41–80 students—secondary	25	0	19	0	3	3
41–80 students—vocational	13	0	12	0	0	1
81–170 students—community/comprehensive	20	0	18	0	0	2
81–170 students—secondary	62	0	49	1	2	10
81–170 students—vocational	19	0	15	1	1	2
$\geq 171$ students	10	1	7	0	0	2
<b>Total</b>	<b>165</b>	<b>1</b>	<b>135</b>	<b>3</b>	<b>6</b>	<b>20</b>

Table B.18.2: Allocation of teacher sample in Ireland

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
≤ 40 students—community/comprehensive	2	0	2	0	0	0
≤ 40 students—secondary	4	0	3	1	0	0
≤ 40 students—vocational	6	0	6	0	0	0
41–80 students—community/comprehensive	4	0	4	0	0	0
41–80 students—secondary	25	0	19	0	3	3
41–80 students—vocational	13	0	11	0	0	2
81–170 students—community/comprehensive	20	0	17	0	0	3
81–170 students—secondary	62	0	47	1	2	12
81–170 students—vocational	19	0	12	1	1	5
≥ 171 students	10	1	7	0	0	2
<b>Total</b>	<b>165</b>	<b>1</b>	<b>128</b>	<b>3</b>	<b>6</b>	<b>27</b>

**Note:** Seven schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.19. Italy

- Very few (fewer than 10 students in the target grade) were excluded at the school level.
- No explicit stratification was performed.
- No implicit stratification was applied.

Table B.19.1: Allocation of student sample in Italy

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	172	0	160	11	1	0
<b>Total</b>	<b>172</b>	<b>0</b>	<b>160</b>	<b>11</b>	<b>1</b>	<b>0</b>

Table B. 19.2: Allocation of teacher sample in Italy

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	172	0	157	10	1	4
<b>Total</b>	<b>172</b>	<b>0</b>	<b>157</b>	<b>10</b>	<b>1</b>	<b>4</b>

**Note:** Three schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.20. Korea

- Special education schools, public middle schools in remote places, and islands and branch schools were excluded at the school level.
- Explicit stratification was performed by region (Capital Region, Chungcheong Gangwon, Honam Jeju, Yeongnam), resulting in four explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), giving a total of eight implicit strata.

Table B.20.1: Allocation of student sample in Korea

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Capital Region	72	0	72	0	0	0
Chungcheong Gangwon	20	0	20	0	0	0
Honam Jeju	18	0	18	0	0	0
Yeongnam	40	0	40	0	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>150</b>	<b>0</b>	<b>0</b>	<b>0</b>

Table B.20.2: Allocation of teacher sample in Korea

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Capital Region	72	0	71	0	0	1
Chungcheong Gangwon	20	0	20	0	0	0
Honam Jeju	18	0	18	0	0	0
Yeongnam	40	0	39	0	0	1
<b>Total</b>	<b>150</b>	<b>0</b>	<b>148</b>	<b>0</b>	<b>0</b>	<b>2</b>



### B.21. Latvia

- Schools for mentally or functionally disabled students, part-time schools (students are usually older) and schools where the language of instruction is neither Latvian nor Russian were excluded at the school level.
- Explicit stratification was performed by urbanization (Riga, city, town, rural), resulting in four explicit strata.
- Implicit stratification was applied by language (Latvian, Russian, mixed) and school type (basic, secondary), giving a total of 23 implicit strata.

Table B.21.1: Allocation of student sample in Latvia

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Riga	42	0	39	1	0	2
City	25	0	23	2	0	0
Town	42	0	38	3	0	1
Rural	51	0	41	3	0	7
<b>Total</b>	<b>160</b>	<b>0</b>	<b>141</b>	<b>9</b>	<b>0</b>	<b>10</b>

Table B.21.2: Allocation of teacher sample in Latvia

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Riga	42	0	38	1	0	3
City	25	0	23	1	0	1
Town	42	0	37	2	0	3
Rural	51	0	41	3	0	7
<b>Total</b>	<b>160</b>	<b>0</b>	<b>139</b>	<b>7</b>	<b>0</b>	<b>14</b>

### B.22. Liechtenstein

- All schools were sampled.
- All classrooms were sampled.
- There were no school-level exclusions.
- No explicit stratification was performed.
- No implicit stratification was applied.

Table B.22: Allocation of student sample and teacher sample in Liechtenstein

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	9	0	9	0	0	0
<b>Total</b>	<b>9</b>	<b>0</b>	<b>9</b>	<b>0</b>	<b>0</b>	<b>0</b>

### B.23. Lithuania

- Special needs schools, very small schools (fewer than five students in the target grade), hospital schools, and students taught in Belarusian were excluded at the school level.
- Explicit stratification was performed by language of instruction (Lithuanian, Polish, Russian) and community size (Vilnius (capital), other major cities, medium-size cities, small cities, other), resulting in seven explicit strata.
- Implicit stratification was applied by school type (gymnasium, secondary, basic, other school type), giving a total of 25 implicit strata.
- Over-sampling of Russian and Polish language schools. The disproportional allocation was taken into account when sampling weights were computed.

Table B.23: Allocation of student sample and teacher sample in Lithuania

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Russian	37	0	37	0	0	0
Polish	30	0	29	0	0	1
Lithuanian—Vilnius (capital)	15	0	15	0	0	0
Lithuanian—other major cities	30	0	30	0	0	0
Lithuanian—medium-size cities	14	0	14	0	0	0
Lithuanian—small cities	29	0	29	0	0	0
Lithuanian—other	45	0	44	1	0	0
<b>Total</b>	<b>200</b>	<b>0</b>	<b>198</b>	<b>1</b>	<b>0</b>	<b>1</b>

### B.24. Luxembourg

- All schools were sampled.
- Special needs schools were excluded at the school level.
- All *enseignement secondaire* and *enseignement secondaire technique* classes were selected in each school. In addition, one *modulaire* class was randomly sampled per school whenever the school had these classes.
- No explicit stratification was performed.
- No implicit stratification was applied.

Table B.24.1: Allocation of student sample in Luxembourg

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	31	0	31	0	0	0
<b>Total</b>	<b>31</b>	<b>0</b>	<b>31</b>	<b>0</b>	<b>0</b>	<b>0</b>

Table B.24.2: Allocation of teacher sample in Luxembourg

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	31	0	24	0	0	7
<b>Total</b>	<b>31</b>	<b>0</b>	<b>24</b>	<b>0</b>	<b>0</b>	<b>7</b>

**Note:** Seven schools were regarded as non-participating because the within-school participation rate was below 50%.

#### B.25. Malta

- All eligible schools were sampled for ICCS 2009.
- Two classrooms were sampled per school.
- Very small schools (fewer than 10 students in the target grade) and special schools (schools for children with acute educational needs) were excluded at the school level.
- No explicit stratification was performed.
- Implicit stratification was applied by sector (state, non-state) and school type (only within the state sector: secondary and junior lyceum), giving a total of three implicit strata.

Table B.25: Allocation of student sample and teacher sample in Malta

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	55	0	55	0	0	0
<b>Total</b>	<b>55</b>	<b>0</b>	<b>55</b>	<b>0</b>	<b>0</b>	<b>0</b>

#### B.26. Mexico

- Workers' secondary schools, very small schools (fewer than five students in the target grade), communitarian secondary schools, special education schools, and private telesecondary schools were excluded at the school level.
- Explicit stratification was performed by school type (general, particular, *técnica*, *telesecundaria*), resulting in four explicit strata.
- No implicit stratification was applied.

Table B.26.1: Allocation of student sample in Mexico

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
General	97	0	95	0	0	2
Particular (private)	16	0	16	0	0	0
Técnica (technical)	62	0	61	0	0	1
Telesecundaria (telesecondary)	45	0	43	0	0	2
<b>Total</b>	<b>220</b>	<b>0</b>	<b>215</b>	<b>0</b>	<b>0</b>	<b>5</b>

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.26.2: Allocation of teacher sample in Mexico

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
General	97	0	90	0	0	7
Particular (private)	16	0	16	0	0	0
Técnica (technical)	62	0	53	0	0	9
Telesecundaria (telesecondary)	45	0	43	0	0	2
<b>Total</b>	<b>220</b>	<b>0</b>	<b>202</b>	<b>0</b>	<b>0</b>	<b>18</b>

**Note:** 12 schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.27. The Netherlands

- Special secondary education schools and practice-oriented schools were excluded at the school level.
- Explicit stratification was performed by groups of tracks (vocational education, general education), resulting in two explicit strata.
- No implicit stratification was applied.
- Two classrooms were sampled from the 30 largest schools (MOS  $\geq 254$ ).

Table B.27.1: Allocation of student sample in the Netherlands

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Vocational Education	66	3	27	7	2	27
General Education	84	5	24	4	3	48
<b>Total</b>	<b>150</b>	<b>8</b>	<b>51</b>	<b>11</b>	<b>5</b>	<b>75</b>

Table B.27.2: Allocation of teacher sample in the Netherlands

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Vocational education	66	3	8	1	2	52
General education	84	5	7	3	1	68
<b>Total</b>	<b>150</b>	<b>8</b>	<b>15</b>	<b>4</b>	<b>3</b>	<b>120</b>

**Note:** 34 schools were regarded as non-participating because the within-school participation rate was below 50%.



### B.28. New Zealand

- Westmount Campus schools (“closed Brethren”), correspondence schools, Māori immersion schools, special education schools, and very small schools (fewer than seven students in the target grade) were excluded at the school level.
- Explicit stratification was performed by school size (very large schools, all other schools), resulting in two explicit strata.
- Implicit stratification was applied by school authority (independent (private) schools, state), by school decile (higher socioeconomic disadvantage, moderate socioeconomic disadvantage, lower socioeconomic disadvantage) for state schools, and by school type (co-ed, boys, girls), giving a total of 10 implicit strata (non-certainty schools).

Table B.28.1: Allocation of student sample in New Zealand

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Very large schools	36	0	26	0	0	10
All other schools	139	3	116	3	1	16
<b>Total</b>	<b>175</b>	<b>3</b>	<b>142</b>	<b>3</b>	<b>1</b>	<b>26</b>

Table B.28.2: Allocation of teacher sample in New Zealand

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Very large schools	36	0	21	0	0	15
All other schools	139	0	90	3	1	45
<b>Total</b>	<b>175</b>	<b>0</b>	<b>111</b>	<b>3</b>	<b>1</b>	<b>60</b>

**Note:** Five schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.29. Norway

- Sami schools, special needs schools, and very small schools (fewer than five students in the target grade) were excluded at the school level.
- Explicit stratification was performed by language (Bokmål, Nynorsk), resulting in two explicit strata.
- No implicit stratification was applied.
- Two classrooms were sampled where possible.





Table B.29.1: Allocation of student sample in Norway

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Bokmål	127	0	79	26	6	16
Nynorsk	23	0	16	1	1	5
<b>Total</b>	<b>150</b>	<b>0</b>	<b>95</b>	<b>27</b>	<b>7</b>	<b>21</b>

Table B.29.2: Allocation of teacher sample in Norway

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Bokmål	127	0	44	15	3	65
Nynorsk	23	0	9	1	1	12
<b>Total</b>	<b>150</b>	<b>0</b>	<b>53</b>	<b>16</b>	<b>4</b>	<b>77</b>

**Note:** 32 schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.30. Paraguay

- Very small schools (fewer than nine students in the target grade) and schools in the departments of Boquerón and Alto Paraguay were excluded at the school level.
- Explicit stratification was performed by urbanization (urban, rural) and sector (public, private, op (subsidized)), resulting in six explicit strata.
- No implicit stratification was applied.

Table B.30.1: Allocation of student sample in Paraguay

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Urban—public	71	0	69	2	0	0
Urban—private	9	0	9	0	0	0
Urban—op (subsidized)	15	0	15	0	0	0
Rural—public	51	0	43	7	0	1
Rural—private	2	0	2	0	0	0
Rural—op (subsidized)	2	0	2	0	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>140</b>	<b>9</b>	<b>0</b>	<b>1</b>



Table B.30.2: Allocation of teacher sample in Paraguay

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Urban—public	71	0	63	2	0	6
Urban—private	9	0	9	0	0	0
Urban—op (subsidized)	15	0	14	0	0	1
Rural—public	51	0	42	6	0	3
Rural—private	2	0	2	0	0	0
Rural—op (subsidized)	2	0	1	0	0	1
<b>Total</b>	<b>150</b>	<b>0</b>	<b>131</b>	<b>8</b>	<b>0</b>	<b>11</b>

**Note:** 10 schools were regarded as non-participating because the within-school participation rate was below 50%.

### B.31. Poland

- Special education schools and very small schools (fewer than 10 students in the target grade) were excluded at the school level.
- Explicit stratification was performed by urbanization (rural, small town, medium-size town, big city), resulting in four explicit strata.
- No implicit stratification was applied.

Table B.31: Allocation of student sample and teacher sample in Poland

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Rural	53	0	53	0	0	0
Small town	31	0	31	0	0	0
Medium-size town	31	0	31	0	0	0
Big city	35	0	34	1	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>149</b>	<b>1</b>	<b>0</b>	<b>0</b>

### B.32. Russian Federation

- Very small schools (fewer than five students in the target grade), special needs schools, and evening schools were excluded at the school level.
- A sample of 45 regions out of 86 was first sampled with PPS. The largest 16 regions were sampled with certainty. A sample of schools was then drawn within each region.
- Explicit stratification was performed by region type (certainty, non-certainty). For a better overview, see Table B.32, which gives the school sample allocation for all 45 regions.
- Implicit stratification was applied by school location (nine levels), giving a total of 229 implicit strata.



Table B.32: Allocation of student sample and teacher sample in the Russian Federation

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Altay kr	4	0	4	0	0	0
Arhangelsk obl	4	0	4	0	0	0
Astrakhan obl	4	0	4	0	0	0
Belgorod obl	4	0	4	0	0	0
Bransk obl	4	0	4	0	0	0
Zabaikasky kr (Chita obl)	4	0	4	0	0	0
Evrey-Auto obl	4	0	4	0	0	0
Hanty-Mansii ok	4	0	4	0	0	0
Karachaevo-Cherkessia	4	0	4	0	0	0
Karelia	4	0	4	0	0	0
Kirov obl	4	0	4	0	0	0
Komi	4	0	4	0	0	0
Kostroma obl	4	0	4	0	0	0
Kurgan obl	4	0	4	0	0	0
Kursk obl	4	0	4	0	0	0
Lipetsk obl	4	0	4	0	0	0
Novosibirsk obl	4	0	4	0	0	0
Omsk obl	4	0	4	0	0	0
Orenburg obl	4	0	4	0	0	0
Pensa obl	4	0	4	0	0	0
Primorsk kr	4	0	4	0	0	0
Sakha	4	0	4	0	0	0
Saratov obl	4	0	4	0	0	0
Smolensk obl	4	0	4	0	0	0
Stavropol kr	4	0	4	0	0	0
Udmurtia	4	0	4	0	0	0
Vladimir obl	4	0	4	0	0	0
Volgograd obl	4	0	4	0	0	0
Voronezh obl	4	0	4	0	0	0
Bashkortostan	8	0	8	0	0	0
Chelyabinsk obl	6	0	6	0	0	0
Dagestan	2	0	2	0	0	0
Irkutsk obl	4	0	4	0	0	0
Kemerovo obl	4	0	4	0	0	0
Krasnodar kr	8	0	8	0	0	0
Krasnoyarsk kr	6	0	6	0	0	0
Moscow	10	0	10	0	0	0
Moscow obl	8	0	8	0	0	0
N Novgorod obl	6	0	6	0	0	0
Perm kr	4	0	4	0	0	0
Rostov obl	6	0	6	0	0	0
Samara obl	4	0	4	0	0	0
Sankt-Petersburg	4	0	4	0	0	0
Sverdlovsk obl	6	0	6	0	0	0
Tatarstan	8	0	8	0	0	0
<b>Total</b>	<b>210</b>	<b>0</b>	<b>210</b>	<b>0</b>	<b>0</b>	<b>0</b>

### B.33. Slovak Republic

- Schools in which the language of instruction was not Slovak were not covered by ICCS 2009.
- Very small schools (fewer than five students in the target grade) were excluded at the school level.
- Explicit stratification was performed by region, resulting in eight explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), school type (second level of basic school, second level of eight-years gymnasium—secondary school), giving a total of 24 implicit strata.

Table B.33.1: Allocation of student sample in the Slovak Republic

School Participation Status— Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Bratislava region	14	0	12	1	0	1
Trnava region	12	0	11	0	0	1
Trenčín region	17	0	15	1	0	1
Nitra region	15	0	13	1	1	0
Žilina region	22	0	19	3	0	0
Banská Bystrica region	16	0	15	1	0	0
Prešov region	26	1	21	3	1	0
Košice region	20	0	18	2	0	0
<b>Total</b>	<b>142</b>	<b>1</b>	<b>124</b>	<b>12</b>	<b>2</b>	<b>3</b>

Table B.33.2: Allocation of teacher sample in the Slovak Republic

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Bratislava region	14	0	13	1	0	0
Trnava region	12	0	11	0	0	1
Trenčín region	17	0	15	1	0	1
Nitra region	15	0	13	1	1	0
Žilina region	22	0	19	3	0	0
Banská Bystrica region	16	0	15	1	0	0
Prešov region	26	1	21	3	1	0
Košice region	20	0	18	2	0	0
<b>Total</b>	<b>142</b>	<b>1</b>	<b>125</b>	<b>12</b>	<b>2</b>	<b>2</b>



### B.34. Slovenia

- Dislocated units of larger schools, Waldorf schools, and Italian-language schools were excluded at the school level.
- No explicit stratification was performed.
- No implicit stratification was applied.

Table B.34.1: Allocation of student sample in Slovenia

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	170	0	156	7	0	7
Total	170	0	156	7	0	7

Table B.34.2: Allocation of teacher sample in Slovenia

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
None	170	0	157	7	0	6
Total	170	0	157	7	0	6

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.



### B.35. Spain

- Special education schools and very small schools (fewer than six students in the target grade) were excluded at the school level.
- Explicit stratification was performed by autonomous communities, resulting in 18 explicit strata.
- Implicit stratification was applied by school type (public, private), giving a total of 36 implicit strata.

Table B.35: Allocation of student sample and teacher sample in Spain

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Andalucía	35	0	34	0	0	1
Aragón	4	0	4	0	0	0
Asturias	2	0	2	0	0	0
Balears (Illes)	3	0	3	0	0	0
Canarias	7	0	7	0	0	0
Cantabria	2	0	2	0	0	0
Castilla y León	7	0	7	0	0	0
Castilla-La Mancha	7	0	7	0	0	0
Catalunya	21	0	21	0	0	0
Ceuta y Melilla	2	0	2	0	0	0
Com. Valenciana	15	0	15	0	0	0
Extremadura	4	0	4	0	0	0
Galicia	8	0	8	0	0	0
Madrid	19	0	16	2	0	1
Murcia	5	0	5	0	0	0
Navarra	2	0	2	0	0	0
País Vasco	5	0	5	0	0	0
Rioja (La)	2	0	2	0	0	0
<b>Total</b>	<b>150</b>	<b>0</b>	<b>146</b>	<b>2</b>	<b>0</b>	<b>2</b>

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.



### B.36. Sweden

- Special needs schools, very small schools (fewer than five students in the target grade), small day care schools, and international schools were excluded at the school level.
- Explicit stratification was performed by school administration (independent, municipality) and proportion of students with immigration background (above 40%, below 40%), resulting in four explicit strata.
- No implicit stratification was applied.
- Over-sampling of independent schools and schools with high proportion of students with immigrant background. The disproportional allocation was taken into account when sampling weights were computed.

Table B.36.1: Allocation of student sample in Sweden

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Independent—≥ 40% foreign background	5	1	3	1	0	0
Independent—< 40% foreign background	25	1	20	1	1	0
Municipality—≥ 40% foreign background	25	2	19	3	0	0
Municipality—< 40% foreign background	120	2	114	3	1	0
<b>Total</b>	<b>175</b>	<b>6</b>	<b>156</b>	<b>8</b>	<b>2</b>	<b>3</b>

**Note:** One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.36.2: Allocation of teacher sample in Sweden

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Independent—≥ 40% foreign background	5	1	3	1	0	0
Independent—< 40% foreign background	25	1	21	1	1	1
Municipality—≥ 40% foreign background	25	2	17	3	0	3
Municipality—< 40% foreign background	120	2	107	2	0	9
<b>Total</b>	<b>175</b>	<b>6</b>	<b>148</b>	<b>7</b>	<b>1</b>	<b>13</b>

**Note:** 11 schools were regarded as non-participating because the within-school participation rate was below 50%.



### B.37. Switzerland

- Very small schools (fewer than six students in the target grade) and special education schools were excluded at the school level.
- Explicit stratification was performed by participation of canton in the field trial (Bern, Fribourg, Waadt, other cantons) and by language (German, French, Italian), resulting in five explicit strata.
- Implicit stratification was applied by school type (basic requirements, advanced requirements, high requirements, other school types), giving a total of 19 implicit strata.

Table B.37.1: Allocation of student sample in Switzerland

School Participation Status—Student Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Bern, Fribourg, Waadt—German	22	0	12	2	2	6
Bern, Fribourg, Waadt—French	24	0	21	2	0	1
Other cantons—German	97	0	46	24	6	21
Other cantons—French	24	0	20	2	0	2
Other cantons—Italian	20	0	14	5	0	1
<b>Total</b>	<b>187</b>	<b>0</b>	<b>113</b>	<b>35</b>	<b>8</b>	<b>31</b>

Table B.37.2: Allocation of teacher sample in Switzerland

School Participation Status—Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Bern, Fribourg, Waadt—German	22	0	12	1	2	7
Bern, Fribourg, Waadt—French	24	0	19	2	0	3
Other cantons—German	97	0	42	21	6	28
Other cantons—French	24	0	18	2	0	4
Other cantons—Italian	20	0	14	5	0	1
<b>Total</b>	<b>187</b>	<b>0</b>	<b>105</b>	<b>31</b>	<b>8</b>	<b>43</b>

**Note:** Nine schools were regarded as non-participating because the within-school participation rate was below 50%.





### B.38. Thailand

- Special education schools, welfare education schools, and specific purpose schools that are not under the supervision of the Ministry of Education were excluded at the school level.
- Explicit stratification was performed by region (middle, north, north east, south) and school type (public, private), resulting in eight explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), giving a total of 16 implicit strata.

Table B.38: Allocation of student sample and teacher sample in Thailand

School Participation Status—Student Survey and Teacher Survey						
Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	First replacement	Second replacement	
Middle—private	7	0	6	1	0	0
Middle—public	43	0	33	10	0	0
North—private	3	0	2	1	0	0
North—public	23	0	21	2	0	0
North east—private	2	0	1	1	0	0
North east—public	50	0	33	17	0	0
South—private	7	0	3	4	0	0
South—public	15	1	11	3	0	0
<b>Total</b>	<b>150</b>	<b>1</b>	<b>110</b>	<b>39</b>	<b>0</b>	<b>0</b>



## Appendix C: Descriptions of cognitive test items and allocations to proficiency levels

Table C1: Descriptions of cognitive test items and allocations to proficiency levels

Item	Item Label	Content Domain	Subdomain	Aspect	Cognitive Domain	Civic Knowledge Metric	Proficiency Level
CI2COM1	Most likely purpose of sticker	2	Equity	N/A	2	434 (3.2)	Level 1
CI2MOM1	Purpose of laws limiting media ownership	1	Civil institutions	The media	2	600 (3.2)	Level 3
CI2MLM1	Best argument for government decision	2	Equity	N/A	2	459 (3.2)	Level 1
CI2MLM2	Best argument against decision	2	Equity	N/A	2	591 (3.2)	Level 3
CI2PDO1	Public debate benefits society	2	Social cohesion	N/A	2	623 (2.1)	Level 3
CI2RDM2	Best reason having system law making	1	State institutions	Legislatures/parliaments	2	575 (3.2)	Level 3
CI2SHM1	Reason refuses to wear new shoes	3	Influencing	Selective purchasing of products according to ethical beliefs about the way they were produced	2	440 (3.2)	Level 1
CI2SHM2	Way to promote social cause	3	Influencing	Selective purchasing of products according to ethical beliefs about the way they were produced	2	315 (4.2)	Below Level 1
CI2TGM1	Importance looking at government records	1	Citizens	Rights/responsibilities	2	572 (3.2)	Level 3
CI2TGM2	Record most likely wants to keep secret	1	State institutions	Governments	2	546 (3.2)	Level 2
CI2BPM1	Important find out policies before vote	3	Community participation	Keeping oneself informed	2	495 (3.2)	Level 2
CI2BPM2	Use only friends' information	3	Community participation	Keeping oneself informed	2	483 (3.2)	Level 2
CI2GFM1	Reason best supports opinion playground	1	State institutions	Governments	2	440 (3.2)	Level 1
CI2BIO1	Understanding histories and cultures	2	Social cohesion	N/A	2	722 (2.1)	Level 3
CI2GLM1	Factory benefits local people	1	Civil institutions	Companies/corporations	2	438 (3.2)	Level 1
CI2GLM2	Factory harms local people	1	Civil institutions	Companies/corporations	2	494 (3.2)	Level 2
CI2FDM1	Adult citizens expected to decide	1	Citizens	Citizens' and groups' assigned and desired rights within their civic society	1	481 (3.2)	Level 2
CI2FSM1	Most likely behavior against law	2	Social cohesion	N/A	2	498 (3.2)	Level 2
CI2SCM1	Reason best support district decision	2	Social cohesion	N/A	2	528 (3.2)	Level 2
CI2SCM2	How better players become because rule	2	Social cohesion	N/A	2	519 (3.2)	Level 2
CI2ASM1	Interview helps farmers get more money	3	Influencing	Engaging in demonstration of public support or protest	2	492 (3.2)	Level 2
CI2ASM2	Farmers seeking	2	Equity	N/A	2	443 (3.2)	Level 1
CI2CNM1	Best reason live next door	2	Equity	N/A	2	496 (3.2)	Level 2
CI2CNM2	Need to understand in order to accept	4	Civic self-image	N/A	2	400 (3.2)	Level 1

Table C1: Descriptions of cognitive test items and allocations to proficiency levels (contd.)

Item	Item Label	Content Domain	Subdomain	Aspect	Cognitive Domain	Civic Knowledge Metric	Proficiency Level
CI2ETO1	Help protect farmer's business	3	Influencing	Selective purchasing of products according to ethical beliefs about the way they were produced	2	702 (2.1)	Level 3
CI2ETM2	Argument against protecting business	2	Freedom	Free trade	2	514 (3.2)	Level 2
CI2BCM1	Best reason elect leader	3	Decision-making	Voting	1	534 (3.2)	Level 2
CI2PRM1	Best reason against violent protest	3	Influencing	Engaging in demonstrations of public support or protest	2	575 (3.2)	Level 3
CI2CCM1	Affection by damage to the environment	4	Civic connectedness	N/A	2	462 (3.2)	Level 1
CI2CCM2	What advisor wants people to understand	4	Civic connectedness	N/A	2	516 (3.2)	Level 2
CI2SRM1	Rules help the class decide	2	Social cohesion	N/A	2	491 (3.2)	Level 2
CI2SRM2	Democratic decision	3	Decision-making	Voting	2	546 (3.2)	Level 2
CI2SRM3	Best explanation unfairness of rule	2	Social cohesion	N/A	2	518 (3.2)	Level 2
CI2OMM1	Important ombudsman office independent	1	State institutions	Governments	2	544 (3.2)	Level 2
CI2OMM2	Benefit service free of charge to public	2	Equity	N/A	2	480 (3.2)	Level 2
CI2OMM3	Ombudsman office benefits government	1	State institutions	Governments	2	515 (3.2)	Level 2
CI2RRO1	Benefits of discussing issues	3	Community participation	Participating in religious, cultural, and sporting organisations	2	626 (2.1)	Level 3
CI2DCM1	Most need to understand benefit group	4	Civic self-image	N/A	2	453 (3.2)	Level 1
CI2PFM1	Leader's decision to allow celebration	2	Equity	N/A	2	399 (3.2)	Level 1
CI2PFM2	Most likely action to change opinion	3	Community participation	Participating in religious, cultural, and sporting organisations	2	503 (3.2)	Level 2
CI2PCM1	Best reason for choosing Card B	3	Influencing	Developing proposals for action or advocacy	2	511 (3.2)	Level 2
CI2PCM2	Availability of card on the internet	3	Decision-making	Engaging in organizational governance	2	411 (3.2)	Level 1
CI2VOM1	Best reason for voluntary voting	3	Decision-making	Voting	2	453 (3.2)	Level 1
CI2VOM2	Best support responsibility to vote	3	Decision-making	Voting	2	539 (3.2)	Level 2
CI2VOM3	Reason for using secret ballots	2	Freedom	N/A	2	376 (3.2)	Below Level 1
CI2DLM1	Why laws about information of donations	1	Civil institutions	Political parties	2	607 (3.2)	Level 3
CI2HRM1	Entitled to protection of human rights	2	Equity	N/A	1	459 (3.2)	Level 1
CI2JOM1	Journalists research report news freely	3	Influencing	Engaging in public debate	2	437 (3.2)	Level 1

Table C1: Descriptions of cognitive test items and allocations to proficiency levels (contd.)

Item	Item Label	Content Domain	Subdomain	Aspect	Cognitive Domain	Civic Knowledge Metric	Proficiency Level
CI2WFO2	Argument in favor of requiring people	1	Citizens	Citizens' and groups' assigned and civic society	2	602(2.1)	Level 3
CI2PGM1	Way pressure groups contribute democracy	1	Civil institutions	Pressure groups	2	503(3.2)	Level 2
CI2PGM2	Advantage of independent pressure groups	1	Civil institutions	Pressure groups	2	588(3.2)	Level 3
CI2ECM1	Most likely reason to ask for advice	3	Decision-making	Engaging in organizational governance	2	430(3.2)	Level 1
CI2ECM2	Risk by asking for advice	3	Decision-making	Engaging in organizational governance	2	506(3.2)	Level 2
CI2CEM1	Censorship government responsibility	2	Social cohesion	N/A	2	426(3.2)	Level 1
CI2CEM2	Best reason for opposing censorship	2	Freedom	N/A	2	557(3.2)	Level 2
CI2WFO1	Reasons for giving money to unemployed people	2	Social cohesion	N/A	2	596(2.1)	Level 3
CI2ORM1	Not teach official religion	2	Freedom	N/A	2	502(3.2)	Level 2
CI2RCM1	Need to introduce new emblems	2	Social cohesion	N/A	2	506(3.2)	Level 2
CI2PIM1	Best reason people in jail cannot vote	1	Citizens	Citizens' and groups' assigned and desired rights within their civic society	2	554(3.2)	Level 2
CI2PIM2	Best reason people in jail allowed vote	1	Citizens	Citizens' and groups' assigned and desired rights within their civic society	2	411(3.2)	Level 1
CI2REM2	Help for refugees	2	Freedom	N/A	1	504(3.2)	Level 2
CI2REM3	Responsibility of refugees	1	Citizens	Citizens' and groups' assigned and desired responsibilities within their civic society	1	444(3.2)	Level 1
CI101M1	Role of citizens in democratic countries	1	Citizens	Citizens' and groups' assigned and desired responsibilities within their civic society	1	451(3.2)	Level 1
CI109M1	Most likely serious threat to democracy	1	State institutions	Governments	1	492(3.2)	Level 2
CI108M1	Main purpose of trade unions	1	Civil institutions	Trade unions	1	517(3.2)	Level 2
CI128M1	Country constitution contains	1	State institutions	Legislatures/parliaments	1	494(3.2)	Level 2
CI137M1	Fact about men and women	1	State institutions	Governments	1	448(3.2)	Level 1
CI110M1	Against the law in democratic country	1	Civil institutions	Pressure groups	1	540(3.2)	Level 2
CI113M1	Main task of national legislature	1	State institutions	Legislatures/parliaments	1	488(3.2)	Level 2
CI104M1	Violated right of reporter	1	Citizens	N/A	1	513(3.2)	Level 2
CI115M1	Clearest violation of civil liberties	1	Citizens	N/A	2	588(3.2)	Level 3



*Table C1: Descriptions of cognitive test items and allocations to proficiency levels (contd.)*

Item	Item Label	Content Domain	Subdomain	Aspect	Cognitive Domain	Civic Knowledge Metric	Proficiency Level
CI119M1	Necessity in democratic countries	1	Citizens	N/A	1	525(3.2)	Level 2
CI120M1	Rights of the child	2	Freedom	N/A	1	420(3.2)	Level 1
CI121M1	Multinational businesses	1	Civil institutions	Companies/corporations	1	598(3.2)	Level 3
CI127M1	Central feature of a free-market economy	1	Civil institutions	Companies/corporations	1	611 (3.2)	Level 3
CI132M1	Opinion about flag and anthem	4	Civic connectedness	N/A	2	481 (3.2)	Level 2
CI129M1	Convincing evidence promoting democracy	1	State institutions	Governments	2	560 (3.2)	Level 2
CI130M1	Clearest example of corruption	1	State institutions	Legislatures/parliaments	1	493 (3.2)	Level 2
CI106M1	Main purpose of human rights	2	Equity	N/A	1	445 (3.2)	Level 1

## Appendix D: Tables with coding information

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
International Student Questionnaire					
AGE	Age of student		Imputed from year and month	Copy value (Missing=mean)	Valid=0 Missing=1
IS2G02	Gender	1. girl 2. boy	1 0	Copy value (Missing=0)	Valid=0 Missing=1
IS2G02BN	Student's ethnic background	1. <A> 2. <B> 3. <C> 4. <D>		Dummy coding for national options; first option is reference category	Valid=0 Missing=1
SISCED	Which of the following <levels of education> do you expect to complete?	1. <ISCED level 5A or 6> 2. <ISCED level 4 or 5B> 3. <ISCED level 3> 4. <ISCED level 2> 5. I do not expect to complete <ISCED level 2>	4 3 2 1 0	Length of dummy code varies across countries; first option is reference category (Missing=1)	Valid=0 Missing=1
IS2G04AN	Student's country of birth	1. <Country of test>	1	Length of dummy code varies across countries; first option is reference category (Missing=1)	Valid=0 Missing=1
IS2G04BN IS2G04CN	Mother's country of birth Father's country of birth	2. <Other country/Group A> 3. <Other country/Group B> 4. <Another country>	0 0 0		
IS2G05N	Language used at home	1. <Language of test> 2. <Other language 1> 3. <Other language 2> 4. <Another language>	1 0 0 0	Length of dummy code varies across countries; first option is reference category (Missing=1)	Valid=0 Missing=1
MSEI	Mother's occupational status	ISEI score		Copy value (Missing=mean)	Valid=0 Missing=1
MISCED		1. <ISCED level 5A or 6> 2. <ISCED level 4 or 5B> 3. <ISCED level 3> 4. <ISCED level 2> 5. <ISCED level 1> 6. She did not complete <ISCED level 1>		Length of dummy code varies across countries; first option is reference category (Missing=1)	Valid=0 Missing=1

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
FSEI	Father's occupational status	ISEI score		Copy value (Missing=mean)	Valid=0 Missing=1
FISCED		1. <ISCED Level 5A or 6> 2. <ISCED Level 4or 5B> 3. <ISCED Level 3> 4. <ISCED Level 2> 5. <ISCED Level 1> 6. She did not complete <ISCED Level 1>		Length of dummy code varies across countries; first option is reference category (Missing=1)	Valid=0 Missing=1
IS2G10A IS2G10B	Interest – mother Interest – father	1. Very interested 2. Quite interested 3. Not very interested 4. Not interested at all	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
IS2G11	Books at home	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	0 1 2 3 4 5	Recode as per scoring (Missing=median)	Valid=0 Missing=1
FAMSTRUC	Family structure	Derived index		Dummy coding for FAMSTRUC variable; first option is reference category	Valid=0 Missing=1
IS2G12A IS2G12B IS2G12C IS2G12D IS2G12E IS2G12F	Activities—television videos or DVDs Activities—homework study for school Activities—computer or internet Activities—reading for enjoyment Activities—talk friends phone internet Activities—spend time with friends	1. No time 2. Fewer than 30 minutes 3. About 30–60 minutes 4. About 1–2 hours 5. More than 2 hours	0 1 2 3 4	Recode as per scoring (Missing=median)	Valid=0 Missing=1
POLDISC	Discussion of pol. and soc. issues – wle	Derived scale		Copy scale value (Missing=median)	Valid=0 Missing=1
IS2G13B IS2G13C	Activities—watching the news on TV Activities—reading newspaper	1. Never or hardly ever 2. Monthly (at least once a month)	0 1	Recode as per scoring (Missing=median)	Valid=0 Missing=1
IS2G13E	Activities—using the internet	3. Weekly (at least once a week)	2		
IS2G13H	Activities—special interest group	4. Daily or almost daily	3		

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
PARTCOM	Civic partic. in the community—wle	Derived scale		Copy scale value (Missing=median)	Valid=0 Missing=1
IS2G14G	Participation—religious group	1. Yes, I have done this within the last 12 months 2. Yes, I have done this but more than a year ago 3. No, I have never done this	2  1  0	Recode as per scoring  (Missing=median)	Valid=0 Missing=1
PARTSCHL	Students' civic participation at school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
OPDISC	Students' perceptions of openness in classroom discussions	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2G16A	Regular lesson—disagree openly	1. Never 2. Rarely 3. Sometimes 4. Often	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
STUDINF	Students' perceptions of influence on decisions about school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2G17G	Opinion—extra curricular activities>	1. To a large extent 2. To a moderate extent 3. To a small extent 4. Not at all	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
STUTREL	Students' perceptions of student-teacher relations at school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2G18D IS2G18G	Agree—outsider at school Agree—bullied by other students	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
VALPARTS	Students' perceptions of the value of participation at school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
DEMVAL	Students' support for democratic values	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1



Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
IS2P20B IS2P20C IS2P20D IS2P20G IS2P20J IS2P20K IS2P20L	Democracy—government jobs Democracy—own all newspapers Democracy—jail without trial (r) Democracy—check letter phone email (r) Democracy—protest never violent Democracy—differences in income Democracy—control media (r)	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 (r=0) 2 (r=1) 1 (r=2) 0 (r=3)	Recode as per scoring (Missing=median)	Valid=0 Missing=1
CITCON	Students' perceptions of the importance of conventional citizenship	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
CITSOC	Students' perceptions of the importance of social movement related citizenship	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P21K IS2P21L	Good citizen—working hard Good citizen—obeying laws	1. Very important 2. Quite important 3. Not very important 4. Not important at all	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
INTPOLS	Students' interest in politics and social issues	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P22F IS2P22G	Interest—environmental issues Interest—European politics	1. Very interested 2. Quite interested 3. Not very interested 4. Not interested at all	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
INPOLEF	Students' sense of internal political efficacy	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
GENEQL	Students' attitudes towards gender equality	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P24G	Gender rights—women raising children	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
ETHRGHT	Students' attitudes toward equal rights for all ethnic/racial groups	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IMMGHT	Students' attitudes towards equal rights for immigrants	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P26F	Immigrants—not many jobs restrict immigration	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
INTRUST	Students' trust in civic institutions	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P27G	Trust institutions—the media	1. Completely 2. Quite a lot 3. A little 4. Not at all	3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
IS2P27H	Trust institutions—the armed forces		2		
IS2P27I	Trust institutions—schools		1		
IS2P27J	Trust institutions—the United Nations		0		
IS2P27K	Trust institutions—people				
IS2P27L	Trust institutions—European Commission	Trust institutions—European Parliament Trust institutions—<state/province> govt			
IS2P27M	Trust institutions—European Parliament				
IS2P27N	Trust institutions—<state/province> govt				
ATTCNT	Students' attitudes toward their country	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P28E	Agree—prefer live in other country	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
			1		
			2		
			3		
SUPPORT	Students' support for political parties	Derived index		Copy index value (Missing=median)	Valid=0 Missing=1
CITEFF	Students' citizenship self-efficacy	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
LEGPROT	Students' expected participation in future legal protest	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ILLPROT	Students' expected participation in future illegal protest	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ELECPART	Students' expected adult electoral participation	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
POLPART	Students' expected adult participation in political activities	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P32H	Adult—vote in <state elections>	1. I will certainly do this 2. I will probably do this 3. I will probably not do this 4. I will certainly not do this	3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
IS2P32I	Adult—vote in European elections		2		
			1		
			0		
INFPART	Students' expected future informal political participation	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
IS2P33A	Expect—volunteer time	1. I will certainly do this 2. I will probably do this 3. I will probably not do this 4. I will certainly not do this	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
IS2P34N	Students' religion	1. No religion 2. [A] 3. [B] 4. [C] 5. [D] 6. Other religion	0 1 1 1 1 1	Length of dummy code varies across countries; first option is reference category (Missing=median)	Valid=0 Missing=1
IS2P35	Religious practices	1. Never 2. Less than once a year 3. At least once a year 4. At least once a month 5. At least once a week	1 2 3 4 5	Recode as per scoring (Missing=1)	Valid=0 Missing=1
RELINF	Students' attitudes toward the influence of religion on society	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
IS2P36F	Agree—religion no longer relevant	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
European Regional Questionnaire					
EUIDENT	Students' sense of European identity	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ES2P01B	Identity—first Europe then <country>	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
ES2P01G	Identity—feel part of the European Union		2		
ES2P01H	Identity—proud member European Union		1		
ES2P01I	Identity—more part of Europe as <region>		0		
EUPART	Students' participation in activities or groups at the European level	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ES2P02E	Participation—info about Euro country	1. Yes, I have done this within the last 12 months 2. Yes, I have done this but more than a year ago 3. No, I have never done this	2	Recode as per scoring (Missing=median)	Valid=0 Missing=1
ES2P02J	Participation—by <organisation>		1 0		

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
EUROPP	Students' reports on opportunities for learning about Europe at school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EUROCOM	Students' participation in communication about Europe	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EUPRLANG	Students' proficiency in another European language	Derived index		Copy index value (Missing=median)	Valid=0 Missing=1
EUATLANG	Students' attitudes toward European language learning	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EUMOVE	Students' attitudes toward freedom of migration within Europe	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EURESTR	Students' attitudes toward restricting migration within Europe	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ES2P08G	Travel-restrictions prevent know Europe	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
EUCITOPP	Students' attitudes toward equal opportunities for other European citizens	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EUCOMPOL	Students' attitudes towards common policies in Europe	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EURUNION	Students' attitudes toward European unification	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ES2P10A ES2P10G	Organization—lose individual identities Organization—free to decide own affairs	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
EUCURR	Students' attitudes toward common European currency	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EUSIZE	Students' attitudes toward further expansion of the European Union	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EUKNOW	Self-reported student knowledge about the European Union	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
Latin American Regional Questionnaire					
AIDENT	Students' sense of Latin American identity	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
AUTGOV	Students' attitudes toward authoritarianism in government	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
S2P03D S2P03E	Dictatorships justified order security Dictatorships justified economic benefit	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
ATTCORR	Students' attitudes toward corrupt practices in government	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
S2P04G S2P04H	Government guarantee right resource use Government to be accountable to society	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing=median)	Valid=0 Missing=1
DISLAW	Students' attitudes toward disobeying the law	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
S2P05G	When only way publicly fight unfair law	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing=median)	Valid=0 Missing=1
ATTDIFF	Students' attitudes toward neighborhood diversity	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
S2P07A S2P07B S2P07C S2P07D	Peace by dialogue and negotiation (r) For peace the end justifies the means When authorities do not fulfill duties Felony against my family	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 (=0) 2 (=1) 1 (=2) 0 (=3)	Recode as per scoring (Missing=median)	Valid=0 Missing=1
ATTVIOL	Students' attitudes toward the use of violence	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EMPATH	Students' feelings of empathy toward classmates	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
EXPAGG	Students' personal experience of physical and verbal aggression at school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1

Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
LS2P11A	People of same sex right to marry	1. Strongly agree	3 (r=0)	Recode as per scoring	Valid=0
LS2P11B	Morality suffers due to homosexuality (r)	2. Agree	2 (r=1)	(Missing=median)	Missing=1
LS2P11C	People of same sex deserve same rights	3. Disagree	1 (r=2)		
LS2P11D	People of same sex not accep in school (r)	4. Strongly disagree	0 (r=3)		
LS2P11E	Homosexuality as a mental illness (r)				
SCHDISC	Students' reports on frequencies of discussion about civic issues at school	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
<b>Asian Regional Questionnaire</b>					
UNDEMG0V	Students' acceptance of authoritarian government practice	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
AS2P01A	Government—take care the way parents do	1. Strongly agree	3	Recode as per scoring	Valid=0
AS2P01C	Government—promote religious life	2. Agree	2	(Missing=median)	Missing=1
		3. Disagree	1		
		4. Strongly disagree	0		
OBAUTH	Students' attitudes towards obedience to authority	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
AS2P02B	Best behave—tolerate wrong behaviors	1. Strongly agree	3	Recode as per scoring	Valid=0
AS2P02E	Best behave—not argue with each other	2. Agree	2	(Missing=median)	Missing=1
AS2P02G	Best behavepoint out what was wrong	3. Disagree	1		
		4. Strongly disagree	0		
TRADCL	Students' attitudes toward the preservation of traditional culture	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
LEGSYS	Students' perceptions of the integrity of the legal system	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
CORRPUB	Students' attitudes toward corruption in public service	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
MORALPOL	Students' attitudes toward personal morality of politicians	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
ASIAID	Students' sense of Asian identity	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1





Table D.1: List of international and regional student-level conditioning variables for ICCS 2009 (contd.)

Variable	Variable Label	Response Options	Scoring	Conditioning Variables	
				Scaling Variable	Missing Variable
AS2P06B	Asian identity—myself as world citizen	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 (r=0) 2 (r=1) 1 (r=2) 0 (r=3)	Recode as per scoring (Missing=median)	Valid=0 Missing=1
AS2P06C	Asian identity—develop an Asian Union				
AS2P06I	Asian identity—Western cntry more proud (r)				
AS2P06J	Asian identity—determine development				
AS2P06K	Asian identity—Western have to learn				
ASIACIT	Students' perceptions of good citizenship	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1
GUANXI	Students' attitude toward the use of connections to hold public office	Derived scale		Copy scale value (Missing=mean)	Valid=0 Missing=1

**Note:** (r) indicates that variable is reverse coded.

Table D.2: Years of parental education corresponding to ISCED levels per country

Country	ISCED 1	ISCED 2	ISCED 3	ISCED 4/5B	ISCED 5A/6
Austria	4	8	12	15	17
Belgium (Flemish)	6	8	12	14	17
Bulgaria	4	8	12	15	17
Chile	6	8	12	16	17
Chinese Taipei	6	9	12	14	16
Colombia	5	9	11	14	15
Cyprus	6	9	12	15	16
Czech Republic	5	9	13	16	16
Denmark	6	9	12	15	17
Dominican Republic	6	8	12	14	16
England	6	9	13	15	16
Estonia	6	9	12	15	16
Finland	6	9	12	14	16
Greece	6	9	12	15	17
Guatemala	6	9	11	14	16
Hong Kong SAR	6	9	13	14	16
Indonesia	6	9	12	15	16
Ireland	6	9	12	14	16
Italy	5	8	13	16	17
Korea, Republic of	6	9	12	14	16
Latvia	6	9	12	15	16
Liechtenstein	5	9	12	14	17
Lithuania	4	10	12	15	16
Luxembourg	6	9	13	15	17
Malta	6	11	13	16	17
Mexico	6	9	12	14	16
Netherlands	6	9	12	15	16
New Zealand	7	10	12	14	15
Norway	6	9	12	14	16
Paraguay	6	9	12	16	16
Poland	4	8	12	15	16
Russian Federation	4	9	11	14	15
Slovak Republic	4	8	12	13	17
Slovenia	4	8	11	15	16
Spain	5	8	11	13	16
Sweden	6	9	12	14	15
Switzerland	6	9	12	14	17
Thailand	6	9	12	14	16





*Table D.3: Years of further schooling beyond target grade corresponding to ISCED levels per country*

Country	ISCED 2	ISCED 3	ISCED 4/5B	ISCED 5A/6
Austria	0	4	7	9
Belgium (Flemish)	0	4	6	9
Bulgaria	0	4	7	9
Chile	0	4	8	9
Chinese Taipei	1	4	6	8
Colombia	1	3	6	7
Cyprus	1	4	7	8
Czech Republic	1	5	8	8
Denmark	1	4	7	9
Dominican Republic	0	4	6	8
England	0	4	6	7
Estonia	1	4	7	8
Finland	1	4	6	8
Greece	1	4	7	9
Guatemala	1	3	6	8
Hong Kong SAR	1	5	6	8
Indonesia	1	4	7	8
Ireland	1	4	6	8
Italy	0	5	8	9
Korea, Republic of	1	4	6	8
Latvia	1	4	7	8
Liechtenstein	1	4	6	9
Lithuania	2	4	7	8
Luxembourg	1	5	7	9
Malta	2	4	7	8
Mexico	1	4	6	8
Netherlands	1	4	7	8
New Zealand	1	3	5	6
Norway	1	4	6	8
Paraguay	1	4	8	8
Poland	1	4	7	8
Russian Federation	1	3	6	7
Slovak Republic	1	5	7	9
Slovenia	1	5	7	8
Spain	2	4	6	8
Sweden	1	4	6	7
Switzerland	1	4	6	9
Thailand	1	4	6	8



Table D.4: Transformation parameters for international and regional questionnaire scales (means and standard deviations of original IRT scores)

International questionnaire scales

Student Questionnaire		
Scale	Mean	SD
ATTCNT	0.011	1.526
CITCON	0.003	1.166
CITEFF	-0.006	1.434
CITSOC	-0.003	1.571
DEMVAL	-0.022	1.356
ELECPART	-0.168	1.968
ETHRGHT	-0.151	2.052
GENEQL	-0.070	1.516
ILLPROT	0.257	2.081
IMMRGHT	-0.035	1.695
INFPART	0.010	2.043
INPOLEF	-0.002	1.732
INTPOLS	0.009	2.122
INTRUST	-0.015	1.732
LEGPROT	0.000	1.379
OPDISC	-0.005	1.164
PARTCOM	0.260	1.060
PARTSCHL	0.038	0.974
POLDISC	0.056	1.287
POLPART	0.020	1.882
RELINF	0.168	2.324
STUDINF	-0.005	1.648
STUTREL	-0.020	1.674
VALPARTS	-0.022	1.548
Teacher Questionnaire		
Scale	Mean	SD
CONFCH	0.014	1.159
TCASSESS	-0.031	1.438
TCHACT	0.008	1.221
TCHPART	0.182	1.299
TCIVACT	-0.004	1.464
TCIVCONF	0.118	1.536
TCLCLIM	0.193	3.030
TSCPROB	0.020	1.757
TSTCLACT	-0.011	1.439
TSTSBEH	-0.109	2.570
TSTUDACT	0.021	1.663
TSTUDINF	-0.017	1.737
School Questionnaire		
Scale	Mean	SD
COMSOCT	-0.013	1.448
CSCPROB	0.008	1.716
CSTUDBEH	1.032	2.793
CSTUDINF	-0.001	1.606
NSCSBEL	-0.928	2.502
RESCOM	-0.097	2.100
SCAUTON	0.011	1.194
SCPARACT	-0.005	1.521
SCSTUDOP	0.000	0.956
SCTCPART	0.009	1.948
SSCSBEL	-0.216	2.591
TSCSBEL	-0.689	2.304

Regional questionnaire scales

European Regional Questionnaire		
Scale	Mean	SD
EUIDENT	0.020	1.540
EUPART	0.020	1.080
EUROPP	0.010	1.380
EUROCOM	0.010	1.160
EUATLANG	-0.030	1.740
EURESTR	0.000	1.230
EUMOVE	-0.020	1.420
EUCITOPP	-0.010	2.140
EUCOMPOL	-0.010	1.310
EURUNION	0.010	1.760
EUCURR	-0.070	1.940
EUSIZE	0.010	1.500
EUKNOW	0.000	1.860
Asian Regional Questionnaire		
Scale	Mean	SD
UNDEMGOV	0.004	1.651
OBAUTH	0.091	2.160
TRADCL	0.084	1.805
LEGSYS	-0.038	1.197
CORRPUB	0.174	1.612
MORALPOL	0.056	1.348
ASIAID	0.017	1.919
ASIACIT	0.097	1.363
GUANXI	0.187	1.953
Latin American Regional Questionnaire		
Scale	Mean	SD
LAIDENT	0.029	1.175
AUTGOV	-0.025	1.196
ATTCORR	-0.026	1.445
DISLAW	-0.002	1.075
ATTDIFF	0.011	1.528
ATTVIOL	-0.007	1.644
EMPATH	-0.108	1.810
ABUSE	-0.061	1.099
SCHDISC	0.005	1.162





This volume contains the technical report for the International Civic and Citizenship Education Study (ICCS) sponsored by the International Association for the Evaluation of Educational Achievement (IEA). Over the past 50 years, IEA has conducted 30 comparative research studies focusing on educational policies, practices, and outcomes in various school subjects taught in more than 80 countries around the world.

ICCS studied the ways in which young people in lower-secondary schools are prepared to undertake their roles as citizens in a range of countries. It investigated student knowledge and understanding of civics and citizenship as well as student perceptions, attitudes, and activities related to civics and citizenship. It also examined differences among countries in these outcomes and the relationship of these outcomes to students' individual characteristics and family background, to teaching practices, and to school and broader community contexts. Three regional modules for Asia, Europe, and Latin America addressed issues of civic and citizenship education of special interest in those parts of the world.

Thirty-eight countries from around the world participated in ICCS. Data gathered from more than 140,000 Grade 8 students and 62,000 teachers in over 5,300 schools have provided evidence that may be used to improve policy and practice in civic and citizenship education. The study's data have also provided a new baseline for future research on civic and citizenship education.

This technical report follows publication of several international and regional reports that presented the results of ICCS. It includes detailed information on the development of the instruments used, including their translation to national languages and translation verification, as well as on sampling design and implementation, sampling weights and participation rates, survey operation procedures, quality control of data collection, data management, and creation of the international database. Separate chapters present scaling of the ICCS test and questionnaires and describe the procedures that were used to report results in the ICCS publications.

This technical report will enable fellow researchers in the field to evaluate published reports, monographs, and articles based on the ICCS data. It will also help them conduct their own analyses of the data included in the ICCS international database, available from IEA, as is the database user guide.