ICCS 2016 Technical Report

EDITORS:
Wolfram Schulz
Ralph Carstens
Bruno Losito
Julian Fraillon

IEA INTERNATIONAL CIVIC AND CITIZENSHIP EDUCATION STUDY 2016

International Civic and Citizenship Education Study 2016 Technical Report

ICCS 2016 Technical Report

IEA International Civic and Citizenship Education Study 2016

Edited by

Wolfram Schulz
Bruno Losito
Ralph Carstens
Julian Fraillon

Contributors

Gabriela Agrusti
John Ainley
Christine Busch
Tim Friedman
Eveline Gebhardt
Juliane Kobelt
Hannah Köhler
Gabriela Nausica Noveanu
Duygu Savaşcı
Sabine Tieck
Sabine Weber

Wolfram Schulz
The Australia Council for Educational Research
Camberwell, Victoria
Australia

Ralph Carstens
International Association for the Evaluation
of Educational Achievement
Hamburg, Germany

Bruno Losito
Laboratorio di Pedagogia Sperimentale
Università degli studi Roma Tre
Rome, Italy

Julian Fraillon
The Australia Council for Educational Research
Camberwell, Victoria
Australia

IEA Secretariat
Keizersgracht 311
1016 EE Amsterdam, the Netherlands
Telephone: +31 20 625 3625
Fax: + 31 20 420 7136
email: secretariat@iea.nl
Website: www.iea.nl

ISBN: 9789079549337

The International Association for the Evaluation of Educational Achievement (IEA), with headquarters in Amsterdam, is an independent, international cooperative of national research institutions and governmental research agencies. It conducts large-scale comparative studies of educational achievement and other aspects of education, with the aim of gaining in-depth understanding of the effects of policies and practices within and across systems of education.

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

Contents

List of tables and figures ix

Chapter 1: Overview of the IEA International Civic and Citizenship Education Study 2016 1
- Overview 1
- Participating countries, populations, and sample design 2
- Study framework 3
- Instruments and data collection 5
- Links and key changes from ICCS 2009 6
- Outline of the technical report 7
- References 8

Chapter 2: ICCS test development 9
- Overview 9
- Test scope and format 10
- Test development process 10
- Field trial test design and content 15
- Main survey 18
- Released test items 20
- References 20

Chapter 3: ICCS questionnaire development 21
- Introduction 21
- Conceptual framework for questionnaire development 21
- Student questionnaire development 24
- Teacher questionnaire development 25
- School questionnaire development 27
- European questionnaire development 28
- Latin American questionnaire development 29
- Development and implementation of the national contexts survey 30
- References 32

Chapter 4: Translation and national adaptations of ICCS instruments 33
- Introduction 33
- Translation and adaption of ICCS 2016 instruments 33
- International verification processes 38

Chapter 5: Sampling design and implementation 43
- Overview 43
- Target population definitions 43
- Coverage and exclusions 44
- Sample size requirements 46
- School sampling design 48
- Within-school sampling design 50
- References 51

Chapter 6: Field operation procedures and data preparation 53
- Overview 53
- Field operation personnel 53
- Field operation resources 54
- Field operation processes 56
- Scoring the ICCS assessment and checking scorer reliability 63
- Field trial procedures 64

Chapter 7: Quality assurance procedures for the ICCS data collection 67
The international quality assurance program 67
Findings from the ICCS 2016 administration 68
Findings from the survey activities questionnaire 78

Chapter 8: Data management, cleaning and creation of the ICCS international database 87
Introduction 87
Data sources 87
Cleaning the international and national databases 89
Interim data products 96
Final product: the ICCS 2016 international database 96
Reference 97

Chapter 9: Weighting procedures 99
Overview 99
Within-school participation requirements 99
Calculating student weights 100
Calculating teacher weights 102
Calculating school weights 104
Calculating participation rates 105
ICCS standards for sampling participation 110
Reference 115

Chapter 10: Scaling procedures for ICCS test items 117
Overview 117
The scaling model 117
Test coverage and item dimensionality 118
Assessment of item fit 119
Assessment of scorer reliabilities 121
Differential item functioning by gender 122
Cross-national measurement equivalence 124
Missing data issues 125
Item adjudication 125
International item calibration and test reliability 131
International ability estimates 133
Equating ICCS 2016 to ICCS 2009 133
The development of proficiency levels for civic knowledge 135
References 137

Chapter 11: Scaling procedures for ICCS questionnaire items 139
Introduction 139
Simple indices 139
Scaling procedures 142
Scaled indices 151
References 242

Chapter 12: The reporting of ICCS 2016 results 245
Overview 245
Estimation of sampling variance 245
Estimation of imputation variance for civic knowledge scores 247
Reporting of differences 249
Hierarchical linear modeling 251
Multiple regression modeling 253
References 256

Appendices 257
Appendix A: Organizations and individuals involved in ICCS 2016 257
Appendix B: Characteristics of national samples 261
Appendix C: Descriptions of civic knowledge test items and allocations to proficiency levels 284
Appendix D: Mapping of questionnaire items common to ICCS 2009 and 2016 288
Appendix E: Coding information for the items in the questionnaires and the test 292

List of tables and figures

Table 2.1:	Test development process timeline	9
Table 2.2:	ICCS 2016 item development plan by item format	11
Table 2.3:	ICCS 2016 item development plan by content and cognitive domain	11
Table 2.4:	Composition of the field trial test instrument by item format and origin	15
Table 2.5:	Field trial cluster composition	16
Table 2.6:	Field trial test booklet design	16
Table 2.7:	ICCS 2016 test development plan	17
Table 2.8:	Field trial item mapping to assessment framework	17
Table 2.9:	Composition of the main study test instrument by item format and origin	18
Table 2.10:	Main study cluster composition	18
Table 2.11:	Main study test booklet design	19
Table 2.12:	Main study item mapping to assessment framework	19
Table 2.13:	Characteristics and mapping of ICCS released test items	20
Table 3.1:	Mapping of variables to contextual framework (examples)	23
Table 4.1:	Languages used for the ICCS 2016 survey instruments	35
Table 5.1:	Population coverage and exclusion rates	45
Table 5.2:	School, student, and teacher sample sizes	47
Table 6.1:	Hierarchical identification (ID) system	57
Table 6.2	Weighted percentages of online mode administration for school and teacher questionnaires	61
Table 6.3	Time allowed for administering the ICCS student instruments	63
Table 7.1:	IQOs' observations on ICCS 2016 administration preparation	70
Table 7.2:	IQOs' observations on test administration sessions	71
Table 7.3:	IQOs' evaluation of the test administrators' adherence to the test administration script	71
Table 7.4:	IQOs' evaluation of student cooperation during test administration	72
Table 7.5:	IQOs' evaluation of the international student questionnaire administration	73
Table 7.6:	IQOs' evaluation of the regional student questionnaire administration	73
Table 7.7:	IQOs' general observations on the ICCS 2016 administration	75
Table 7.8:	IQO reported interview responses of the school coordinators to questions about their perceptions of the ICCS 2016 administration	76
Table 7.9:	IQO reported interview responses of the school coordinators to questions evaluating the ICCS 2016 administration	77
Table 7.10:	NRC responses to questions related to sampling	79
Table 7.11	NRC responses to questions about contacting schools and recruiting school coordinators	80
Table 7.12:	NRC responses to questions about adapting and translating the ICCS assessment materials	81
Table 7.13:	NRC responses to questions about assembling and printing the ICCS assessment materials	82
Table 7.14:	NRC responses to questions about booklet and questionnaire administration	83
Table 7.15:	NRC responses to questions about preparing and administering the online questionnaire	84

Table 7.16	NRC responses to questions about scoring open-ended response items and coding occupational data	85
Table 7.17	NRC responses to questions about entering and submitting data	85
Table 7.18	NRC responses to questions about their impressions of the manuals provided	86
Table 9.1:	Unweighted participation rates in the student survey	107
Table 9.2:	Weighted participation rates in the student survey	108
Table 9.3:	Unweighted participation rates in the teacher survey	109
Table 9.4:	Weighted participation rates in the teacher survey	111
Table 9.5:	Participation by country in the student and teacher surveys	114
Table 10.1:	Item total-score correlations and weighted item fit for international sample	120
Table 10.2:	Percentages of scorer agreement for open-ended ICCS test items	121
Table 10.3:	Gender DIF estimates for test items	123
Table 10.4:	Percentages of omitted and invalid responses for test items	126
Table 10.5:	Example of item statistics provided to national centers	128
Table 10.6:	National items excluded from scaling	130
Table 10.7:	Final set of ICCS 2016 item parameters used to scale international test items	132
Table 11.1:	Transformation parameters for new ICCS 2016 questionnaire scales (means and standard deviations of original IRT logit scores)	146
Table 11.2:	Transformation parameters and linking error for equated questionnaire scales	148
Table 11.3:	Factor loadings and reliabilities for the national index of students' socioeconomic background	152
Table 11.4:	Reliabilities for scales measuring students' participation in out-of-school activities	154
Table 11.5:	Item parameters for scales measuring students' participation in out-of-school activities	155
Table 11.6:	Reliabilities for scales measuring students' civic participation in the community and at school	157
Table 11.7:	Item parameters for scale measuring students' civic participation in the community and at school	158
Table 11.8:	Reliabilities for scales measuring students' perceptions of civic learning and participation at school	160
Table 11.9:	Item parameters for scale measuring students' perceptions of civic learning and participation at school	161
Table 11.10:	Reliabilities for scales measuring students' perceptions of school climate and interactions	163
Table 11.11:	Item parameters for scales measuring students' perceptions of school climate and interactions	164
Table 11.12:	Reliabilities for scales measuring students' perceptions of the importance of citizenship behaviors	166
Table 11.13:	Item parameters for scales measuring students' perceptions of the importance of citizenship behaviors	167
Table 11.14:	Reliabilities for scales measuring students' endorsement of equal rights and opportunities	169

Table 11.15:	Item parameters for scales measuring students' endorsement of equal rights and opportunities	170
Table 11.16:	Reliabilities for scales measuring students' attitudes toward civic institutions and their country of residence	172
Table 11.17:	Item parameters for scales measuring students' attitudes toward civic institutions and their country of residence	173
Table 11.18:	Reliabilities for scales measuring students' dispositions toward civic engagement	175
Table 11.19:	Item parameters for scales measuring students' dispositions toward civic engagement	176
Table 11.20:	Reliabilities for scales measuring students' expectations to engage in activities to express their opinion	178
Table 11.21:	Item parameters for scale measuring students' expectations to engage in activities to express their opinion	179
Table 11.22:	Reliabilities for scales measuring students' expected political participation	181
Table 11.23:	Item parameters for scales measuring students' expected political participation	182
Table 11.24:	Reliabilities for scale measuring students' endorsement of religious influence in society	183
Table 11.25:	Item parameters for scale measuring students' endorsement of religious influence in society	184
Table 11.26:	Reliabilities for scales measuring students' perceptions of European identity and learning about Europe at school	186
Table 11.27:	Item parameters for scales measuring students' perceptions of European identity and learning about Europe at school	187
Table 11.28:	Reliabilities for scales measuring students' attitudes toward migration in Europe	189
Table 11.29:	Item parameters for scales measuring students' attitudes toward migration in Europe	190
Table 11.30:	Reliabilities for scales measuring students' attitudes toward European integration	192
Table 11.31:	Item parameters for scales measuring students' attitudes toward European integration	193
Table 11.32:	Reliabilities for scales measuring students' expectations for Europe's and their own individual future	195
Table 11.33:	Item parameters for scales measuring students' expectations for Europe's and their own individual future	196
Table 11.34:	Reliabilities for scales measuring students' attitudes toward authoritarian and corrupt government practices	198
Table 11.35:	Item parameters for scales measuring students' attitudes toward authoritarian and corrupt government practices	198
Table 11.36:	Reliabilities for scales measuring students' attitudes toward violence and disobedience to the law	200
Table 11.37:	Item parameters for scales measuring students' attitudes toward violence and disobedience to the law	201
Table 11.38:	Reliabilities for scale measuring students' acceptance of neighborhood diversity	203

Table 11.39: Item parameters for scale measuring students' acceptance of neighborhood diversity 203
Table 11.40: Reliabilities for scale measuring students' feelings of empathy toward classmates 205
Table 11.41: Item parameters for scale measuring students' feelings of empathy toward classmates 205
Table 11.42: Reliabilities for scale measuring students' acceptance of homosexuality 206
Table 11.43: Item parameters for scale measuring students' acceptance of homosexuality 207
Table 11.44: Reliabilities for scale measuring teachers' perceptions of teacher participation at school 208
Table 11.45: Item parameters for scale measuring teachers' perceptions of teacher participation at school 208
Table 11.46: Reliabilities for scale measuring teachers' perceptions of social problems at school 210
Table 11.47: Item parameters for scale measuring teachers' perceptions of social problems at school 210
Table 11.48: Reliabilities for scale measuring teachers' perceptions of student activities in the community 212
Table 11.49: Item parameters for scale measuring teachers' perceptions of student activities in the community 213
Table 11.50: Reliabilities for scales measuring teachers' perceptions of student behavior, classroom climate and bullying at school 215
Table 11.51: Item parameters for scales measuring teachers' perceptions of student behavior, classroom climate and bullying at school 216
Table 11.52: Reliabilities for scales measuring teachers' reports of civic-related activities in class and PD activities for teaching methods 218
Table 11.53: Item parameters for scales measuring teachers' reports of civic-related activities in class and PD activities for teaching methods 219
Table 11.54: Reliabilities for scales measuring teachers' reports of their confidence to teach CCE topics and participation in CCE PD activities 221
Table 11.55: Item parameters for scales measuring teachers' reports of their confidence to teach CCE topics and participation in CCE PD activities 222
Table 11.56: Reliabilities for scale measuring principals' perceptions of teacher participation in school governance 224
Table 11.57: Item parameters for scale measuring principals' perceptions of teacher participation in school governance 224
Table 11.58: Reliabilities for scales measuring principals' perceptions of students' and teachers' sense of belonging at school 226
Table 11.59: Item parameters for scales measuring principals' perceptions of students' and teachers' sense of belonging at school 227
Table 11.60: Reliabilities for scale measuring principals' perceptions of student opportunities to participate in community activities 228
Table 11.61: Item parameters for scale measuring principals' perceptions of student opportunities to participate in community activities 229
Table 11.62: Reliabilities for scales measuring principals' perceptions of bullying and reports on activities against bullying at school 231
Table 11.63: Item parameters for scales measuring principals' perceptions of bullying and report on activities against bullying at school 232

Table 11.64:	Reliabilities for scale measuring principals' perceptions of engagement of the school community	234
Table 11.65:	Item parameters for scale measuring principals' perceptions of engagement of the school community	235
Table 11.66:	Reliabilities for scale measuring principals' reports on environment-friendly practices at school	236
Table 11.67:	Item parameters for scale measuring principals' reports on environment-friendly practices at school	237
Table 11.68:	Reliabilities for scale measuring principals' perceptions of the availability of resources in local community	238
Table 11.69:	Item parameters for scale measuring principals' perceptions of the availability of resources in local community	239
Table 11.70:	Reliabilities for scales measuring principals' perceptions of social tension, poverty and crime in the community	241
Table 11.71:	Item parameters for scales measuring principals' perceptions of social tension, poverty and crime in the community	242
Table 12.1:	Example of the computation of replicate weights	246
Table 12.2:	National averages for civic knowledge with standard deviations, sampling, and overall errors	248
Table 12.3:	Estimates of equating error for percentages at or above Level B	251
Table 12.4:	ICCS 2016 students included in multilevel analyses of variation in civic knowledge	253
Table 12.5:	ICCS 2016 students included in multiple regression analyses of expected electoral and active political participation	255
Table B.1.1:	Allocation of student sample in Belgium (Flemish)	261
Table B.1.2:	Allocation of teacher sample in Belgium (Flemish)	261
Table B.2.1:	Allocation of student sample in Bulgaria	262
Table B.2.2:	Allocation of teacher sample in Bulgaria	262
Table B.3.1:	Allocation of student sample in Chile	263
Table B.3.2:	Allocation of teacher sample in Chile	263
Table B.4.1:	Allocation of student sample in Chinese Taipei	263
Table B.4.2:	Allocation of teacher sample in Chinese Taipei	264
Table B.5.1:	Allocation of student sample in Colombia	264
Table B.5.2:	Allocation of teacher sample in Colombia	264
Table B.6.1:	Allocation of student sample in Croatia	265
Table B.6.2:	Allocation of teacher sample in Croatia	265
Table B.7.1:	Allocation of student sample in Denmark	266
Table B.7.2:	Allocation of teacher sample in Denmark	266
Table B.8.1:	Allocation of student sample in the Dominican Republic	266
Table B.8.2:	Allocation of teacher sample in the Dominican Republic	267
Table B.9.1:	Allocation of student sample in Estonia	267
Table B.9.2:	Allocation of teacher sample in Estonia	267
Table B.10.1:	Allocation of student sample in Finland	268
Table B.10.2:	Allocation of teacher sample in Finland	269
Table B.11:	Allocation of student sample in Hong Kong SAR	270
Table B.12:	Allocation of student and teacher samples in Italy	270

Table B.13.1:	Allocation of student sample in the Republic of Korea	271
Table B.13.2:	Allocation of teacher sample in the Republic of Korea	271
Table B.14.1:	Allocation of student sample in Latvia	272
Table B.14.2:	Allocation of teacher sample in Latvia	273
Table B.15.1:	Allocation of student sample in Lithuania	274
Table B.15.2:	Allocation of teacher sample in Lithuania	274
Table B.16:	Allocation of student and teacher samples in Malta	275
Table B.17.1:	Allocation of student sample in Mexico	276
Table B.17.2:	Allocation of teacher sample in Mexico	276
Table B.18.1:	Allocation of student sample in the Netherlands	277
Table B.18.2:	Allocation of teacher sample in the Netherlands	277
Table B.19.1:	Allocation of student sample in North Rhine-Westphalia (Germany)	278
Table B.19.2:	Allocation of teacher sample in North Rhine-Westphalia (Germany)	278
Table B.20.1:	Allocation of student sample in Norway	279
Table B.20.2:	Allocation of teacher sample in Norway	279
Table B.21:	Allocation of student and teacher samples in Peru	280
Table B.22:	Allocation of student sample in the Russian Federation	281
Table B.23.1:	Allocation of student sample in Slovenia	282
Table B.23.2:	Allocation of teacher sample in Slovenia	282
Table B.24.1:	Allocation of student sample in Sweden	283
Table B.24.2:	Allocation of teacher sample in Sweden	283
Table C1:	Descriptions of civic knowledge test items and allocations to proficiency levels	284
Table D.1:	Mapping of student questionnaire items between 2009 and 2016	288
Table D.2:	Mapping of European questionnaire items between 2009 and 2016	289
Table D.3:	Mapping of Latin American questionnaire items between 2009 and 2016	289
Table D.4:	Mapping of teacher questionnaire items between 2009 and 2016	290
Table D.5:	Mapping of school questionnaire items between 2009 and 2016	291
Table E.1:	List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge	292

Figures

Figure 3.1:	Contexts for the development of learning outcomes related to civic and citizenship education	23
Figure 5.1:	Systematic PPS sampling of schools	49
Figure 6.1:	Procedures for working with schools to prepare for test administration	58
Figure 8.1:	Overview of the data cleaning process	91
Figure 9.1:	Categories into which countries were placed with respect to student sampling participation rates	112
Figure 9.2:	Categories into which countries were placed with respect to teacher sampling participation rates	113
Figure 10.1:	Mapping of student abilities and item thresholds	118
Figure 10.2:	Item characteristic curve by category for item CI2WFO2	119
Figure 10.3:	Gender DIF plot for item CI3RPM1	124

Figure 10.4: Example of an item-by-country interaction graph for item CI3MPO2 125
Figure 10.5: Scatterplot for link item parameter estimates from ICCS 2009 and ICCS 2016 134
Figure 10.6: Proficiency level cut-points and percentage of students at each level 136

Figure 11.1: Summed category probabilities for a fictitious item 149
Figure 11.2: Example of questionnaire item-by-score map 150
Figure 11.3: Confirmatory factor analysis of items measuring students' participation in out-of-school activities 153
Figure 11.4: Confirmatory factor analysis of items measuring students' civic participation in the community and at school 156
Figure 11.5: Confirmatory factor analysis of items measuring students' perceptions of civic learning and participation at school 159
Figure 11.6: Confirmatory factor analysis of items measuring students' perceptions of school climate and interactions 162
Figure 11.7: Confirmatory factor analysis of items measuring students' perceptions of the importance of citizenship behaviors 165
Figure 11.8: Confirmatory factor analysis of items measuring students' endorsement of equal rights and opportunities 168
Figure 11.9: Confirmatory factor analysis of items measuring students' attitudes toward civic institutions and their country of residence 171
Figure 11.10: Confirmatory factor analysis of items measuring students' dispositions toward civic engagement 174
Figure 11.11: Confirmatory factor analysis of items measuring students' expectations to engage in activities to express their opinion 177
Figure 11.12: Confirmatory factor analysis of items measuring students' expected political participation 180
Figure 11.13: Confirmatory factor analysis of items measuring students' endorsement of religious influence in society 182
Figure 11.14: Confirmatory factor analysis of items measuring students' perceptions of European identity and learning about Europe at school 185
Figure 11.15: Confirmatory factor analysis of items measuring students' attitudes toward migration in Europe 188
Figure 11.16: Confirmatory factor analysis of items measuring students' attitudes toward European integration 191
Figure 11.17: Confirmatory factor analysis of items measuring students' expectations for Europe's and their own individual future 194
Figure 11.18: Confirmatory factor analysis of items measuring students' attitudes toward authoritarian and corrupt government practices 197
Figure 11.19: Confirmatory factor analysis of items measuring students' attitudes toward violence and disobedience to the law 199
Figure 11.20: Confirmatory factor analysis of items measuring students' acceptance of neighborhood diversity 202
Figure 11.21: Confirmatory factor analysis of items measuring students' feelings of empathy toward classmates 204
Figure 11.22: Confirmatory factor analysis of items measuring students' acceptance of homosexuality 206

Figure 11.23: Confirmatory factor analysis of items measuring teachers' perceptions of teacher participation at school 207
Figure 11.24: Confirmatory factor analysis of items measuring teachers' perceptions of social problems at school 209
Figure 11.25: Confirmatory factor analysis of items measuring teachers' perceptions of student activities in the community 211
Figure 11.26: Confirmatory factor analysis of items measuring teachers' perceptions of student behavior, classroom climate and bullying at school 214
Figure 11.27: Confirmatory factor analysis of items measuring teachers' reports on civic-related activities in class and PD activities for teaching methods 217
Figure 11.28: Confirmatory factor analysis of items measuring teachers' reports of their confidence to teach CCE topics and participation in CCE PD activities 220
Figure 11.29: Confirmatory factor analysis of items measuring principals' perceptions of teacher participation in school governance 223
Figure 11.30: Confirmatory factor analysis of items measuring principals' perceptions of students' and teachers' sense of belonging at school 225
Figure 11.31: Confirmatory factor analysis of items measuring principals' perceptions of student opportunities to participate in community activities 227
Figure 11.32: Confirmatory factor analysis of items measuring principals' perceptions of bullying and reports on activities against bullying at school 230
Figure 11.33: Confirmatory factor analysis of items measuring principals' perceptions of engagement of the school community 233
Figure 11.34: Confirmatory factor analysis of items measuring principals' reports on environment-friendly practices at school 235
Figure 11.35: Confirmatory factor analysis of items measuring principals' perceptions of the availability of resources in local community 237
Figure 11.36: Confirmatory factor analysis of items measuring principals' perceptions of social tension, poverty and crime in the community 240

CHAPTER 1:

Overview of the IEA International Civic and Citizenship Education Study 2016

Ralph Carstens and Wolfram Schulz

Overview

The International Association for the Evaluation of Educational Achievement (IEA) International Civic and Citizenship Education Study (ICCS) 2016 investigated the ways in which young people are prepared to undertake their roles as citizens in a range of countries in the second decade of the 21st century. It studied students' knowledge and understanding of civics and citizenship, as well as their attitudes, perceptions, and activities related to civics and citizenship. Based on nationally representative samples of students, the study further examined differences among countries in relation to these outcomes of civic and citizenship education, and explored how cross-national differences relate to student characteristics, school and community contexts, and national characteristics.

Building on IEA's previous studies of civic education, the IEA established ICCS in order to meet the need for continuing research on civic and citizenship education and as a response to widespread interest in conducting regular international assessment of this field of education. As the second cycle of this study, ICCS 2016 is a continuation and an extension of ICCS 2009, intended as an exploration of the enduring and emerging challenges of educating young people in a world where contexts of democracy and civic participation continue to change at the national, regional, and global levels. ICCS 2016 extends the scope previous assessments, but is designed to maintain continuity with ICCS 2009, enabling enduring aspects of civic and citizenship educational contexts, processes, and outcomes to be measured, while supporting comparison of outcomes and contexts between 2009 and 2016. Some key materials and variables are statistically linked to enable changes to be investigated.

ICCS 2016 addressed the following research questions (see Schulz, Ainley, Fraillon, Losito, & Agrusti, 2016):

(1) The way civic and citizenship education is implemented in participating countries, including the aim and principles for this learning area, the curricular approaches chosen to provide it, and changes and/or developments since 2009.

(2) The extent of students' knowledge and understanding of civics and citizenship, and the factors associated with its variation across and within countries.

(3) Students' current and expected future involvement in civic-related activities, their perceptions of their capacity to engage in these activities, and their perceptions of the value of civic engagement.

(4) Students' beliefs about contemporary civil and civic issues in society, including those concerned with civic institutions, rules, and social principles (democracy, citizenship, and diversity), as well as their perceptions of their communities and threats to the world's future.

(5) The ways in which schools organize civic and citizenship education, with a particular focus on general approaches, the processes used to facilitate civic engagement, interaction with their communities, and schools' and teachers' perceptions of the role of this learning area.

ICCS 2016 gathered data from more than 94,000 students in their eighth year of schooling in about 3800 schools from 24 countries. Most of these countries had participated in ICCS 2009. The student data were augmented by data from more than 37,000 teachers in those schools and by contextual data collected from school principals and national research centers. An additional European student questionnaire in ICCS 2016 gathered data from almost 53,000 students in 14 European countries and one benchmarking participant (North Rhine-Westphalia, Germany). The Latin American student questionnaire in ICCS 2016 gathered data from more than 25,000 students in five Latin American countries.

All materials for the study were developed with reference to the ICCS 2016 assessment framework (Schulz et al., 2016). Initial results were reported in the ICCS 2016 international report (Schulz, Ainley, Fraillon, Losito, Agrusti, & Friedman, 2018b), the European report (Losito, Agrusti, Damiani, & Schulz, 2018), and the Latin American report (Schulz, Ainley, Cox, & Friedman, 2018a). A user guide describes the organization, content, and use of the international database from a practical perspective (Köhler, Weber, Brese, Schulz, & Carstens, 2018).

ICCS 2016 was organized and carried out by a consortium of three partner institutions, which closely cooperated with the national research coordinators (NRCs) from the participating countries: (i) the Australian Council for Educational Research (ACER) in Melbourne acted as the international study center, with responsibilities for the general development of study design, assessment framework and international student instruments; (ii) the IEA in Hamburg acted as the project coordination center, with responsibilities for data processing, sampling, scaling, data analysis and general coordination activities; and (iii) the Laboratorio di Pedagogia Sperimentale (LPS) at the Roma Tre University in Rome was the associated study center, with responsibilities for the development of the teacher and school surveys and a European regional module.

Dedicated groups were established to help manage the study, including: (i) a joint management committee (JMC) composed of key staff from all centers involved in the consortium, and (ii) a project advisory committee (PAC); individual experts were also consulted. NRCs played a crucial role in this study, as they coordinated the work of the national research centers, oversaw the local implementation of survey procedures, and contributed to the development of the assessment framework, instruments and reporting through a series of face-to-face meetings, and through a regular communication network established by the international project team.

The next sections provide a compressed overview of key parameters for ICCS 2016, followed by a description of the contents of this technical report. More detailed information on the study's design, implementation, results, and any limitations are provided in the further chapters of this report and in the assessment framework (Schulz et al., 2016), the international and regional reports (Losito et al., 2018; Schulz et al., 2018a,b), and the user guide (Köhler et al., 2018).

This ICCS 2016 technical report provides a comprehensive account of the conceptual, methodological, and analytical implementation of the study, which should complement the other, substantive publications. Using all these publications in combination will allow researchers and analysts to understand the procedures used and correctly undertake new analyses.

Participating countries, populations, and sample design

Twenty-four countries[1] participated in ICCS 2016. Sixteen of those countries were from Europe, five from Latin America, and three from Asia. In two of the participating countries, only sub-national entities participated. In Belgium, ICCS 2016 was implemented only in the Flemish education system. In Germany, one state (Land), North Rhine-Westphalia, took part in as a benchmarking participant. As is the case with other IEA studies, participation in ICCS is open to all IEA member countries and affiliated countries and education systems. Each country decides whether or not it will participate in an IEA study.

1 In this report, the term "country" refers to both the countries and the sub-national entities within countries that participated in the study.

The student and teacher population definitions and sampling methods were the same as those used in ICCS 2009. The ICCS student population is defined as all students in Grade 8 (students approximately 14 years of age), provided that the average age of students in this grade was 13.5 years or above at the time of the assessment. If the average age of students in Grade 8 was below 13.5 years, Grade 9 became the target population. The population for the ICCS teacher survey was defined as all teachers teaching regular school subjects to students enrolled in the country's target grade at each sampled school.

The school samples were designed as stratified two-stage cluster samples; schools were randomly selected at the first stage with probability proportional to size, and intact classrooms were sampled at the second stage. Typically, each country aimed for a sample size of 150 schools, the total number of sampled students ranging between 3000 and 4500. Furthermore, ICCS aimed to sample 15 teachers from all teachers teaching the target grade at each sampled school.

As in other IEA studies, sample participation requirements for both student and teacher surveys were 85 percent participation of the selected schools and 85 percent of the selected students within the participating schools, or a weighted overall participation rate of 75 percent. Countries that did not meet the required response rates, even after replacement, are reported separately below the main section of each table.

Study framework

The ICCS 2016 assessment framework (Schulz et al., 2016) provided the conceptual underpinnings for the study. The 2016 framework was developed as an extension and refinement of the ICCS 2009 framework (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008). This approach not only supported the measurement and ongoing reporting of core elements of ICCS (as measured and reported in ICCS 2009) but also allowed consideration of the newer, often global, developments likely to have influenced civic and citizenship education since ICCS 2009.

The structure of the ICCS 2016 framework and the suggested analytical implications of this structure are consistent with the corresponding features of the ICCS 2009 framework. The 2016 framework differs from the 2009 framework only in terms of the addition of new content areas and some revisions to content within the framework.

The 2016 framework consists of two parts:

- The *civics and citizenship* framework outlines the outcome measures addressed by the cognitive test and the international and regional student questionnaires;
- The *contextual* framework maps the contextual factors expected to influence outcomes and explain their variation.

The civics and citizenship framework is organized around three dimensions.

- A *content dimension* specifying the subject matter to be assessed within civics and citizenship (with regard to both affective-behavioral and cognitive aspects);
- A *cognitive dimension* describing the thinking processes to be assessed in the student test;
- An *affective-behavioral dimension* describing the types of student perceptions and activities measured by the student questionnaire.

The four content domains in the ICCS assessment framework are civic society and systems, civic principles, civic participation, and civic identities. Each of these contains a set of sub-domains that incorporate elements referred to as "aspects" and "key concepts."

- *Civic society and systems* (three sub-domains): (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* (four sub-domains): (i) equity (all people having the right to fair and just treatment), (ii) freedom (of belief, of speech, from fear, and from want), (iii) sense of community (sense of belonging, connectedness, and common vision among individuals and communities within a society), and (iv) rule of law (equal and fair application of the law to all; separation of powers and legal transparency).
- *Civic participation* (three sub-domains): (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* (two sub-domains): (i) civic self-image (individuals' experience of their place in each of their civic communities), and (ii) civic connectedness (sense of connection to different civic communities and the civic roles individuals play within each community). ICCS also includes global citizenship as a key concept relating to students' civic identities.

The two cognitive processes in the ICCS framework are:

- *Knowing:* This refers to the learned civic and citizenship information students use when engaging in the more complex cognitive tasks that help them make sense of their civic worlds.
- *Reasoning and applying:* This refers to the ways in which students use civic and citizenship information to reach conclusions that are broader than the contents of any single concept. This process also refers to how students use these conclusions in real-world contexts.

The assessment framework identified the different types of student perceptions and behaviors relevant to civics and citizenship along two affective-behavioral domains: (i) attitudes, and (ii) engagement.

- *Attitudes:* These refer to judgments or evaluations regarding ideas, persons, objects, events, situations, and/or relationships. They include students' beliefs about democracy and citizenship, students' attitudes toward the rights and responsibilities of groups in society, and students' attitudes toward institutions.
- *Engagement:* This refers to students' civic engagement, students' expectations of future civic-related action, and students' dispositions to actively engage in society (interest, sense of efficacy). The notion of engagement includes concepts such as preparedness to participate in forms of civic protest, anticipated future political participation as adults, and anticipated future participation in citizenship activities.

The ICCS 2016 research team, in close collaboration with participating countries, identified three areas of civic and citizenship education that warranted a stronger profile in ICCS 2016 than they had been afforded in ICCS 2009. The likely relevance of this content in future cycles of ICCS also influenced its inclusion.

- *Environmental sustainability in civic and citizenship education:* Over recent decades, countries have increasingly concluded that responsible citizenship includes regard for the environment and its long-term protection, requisite for future sustainable development. Today, many education systems emphasize protection of the environment or education for environmental sustainability in their citizenship curricula.

- *Social interaction at school:* Reviews of civic and citizenship education curricula across countries suggest that at the outset of the 21st century a large number of countries were emphasizing the non-formal aspects of civic learning through participation and engagement or social interaction within their schools. Scholars studying this learning area have also started to give greater recognition to the role of social learning within schools.
- *The use of social media for civic engagement:* Research continues to emphasize the growing importance of social media on civic life and to provide evidence of how these media influence young people's engagement in society.

Instruments and data collection

ICCS 2016 comprised the following instruments administered to students, teachers, school principals, and national centers:

- Eighty-eight items measuring civic and citizenship knowledge, analysis, and reasoning contained in the *student cognitive test* were assigned to eight booklets (each of which contained three of a total eight 11-item clusters) according to a balanced rotated design. Each student completed one of the 45-minute booklets. The test items were generally presented with contextual material that served as a brief introduction to each item or set of items.
- The *student questionnaire* took between 30 and 40 minutes to complete and was used to obtain students' perceptions about civics and citizenship, and information about each student's background.
- The *teacher questionnaire* (designed to take about 30 minutes) asked respondents about their perceptions of civic and citizenship education in their schools. It also asked them to provide information about their schools' organization and culture, as well as their own teaching assignments and backgrounds.
- The *school questionnaire*, which also took 30 minutes to complete, asked school principals to provide information about school characteristics, school culture and climate, and the provision of civic and citizenship education in the school.
- National research coordinators (NRCs) compiled and synthesized the information procured from national experts in response to an online *national contexts survey*. This information concerned the structure of the education system, civic and citizenship education in the national curricula, and recent developments in civic and citizenship education.

In addition to the international and regional instruments, ICCS 2016 offered several international options within the questionnaires and invited the national centers to consider using them. These options contained items concerning students' ethnicity, household composition, and religion, and a number of specific questions for teachers of civic and citizenship education.

The regional instruments, an innovative feature first introduced in ICCS 2009, were again made available to countries in regions with five or more participating countries, in ICCS 2016 this was the case for Europe and Latin America.[2] The purpose of regional instruments in ICCS is to allow collection of region-specific aspects of civic and citizenship education. The questions in the ICCS 2016 regional instruments, which took roughly 15 minutes to answer, focused on particular issues associated with civics and citizenship in the respective geographical region.

The ICCS 2016 design included three major components – a pilot study, a field trial and the main survey. As preparation for the more qualitative pilot study, the international research team asked participating countries to establish focus groups that included students, teachers, and principals,

2 An Asian regional component was not included in ICCS 2016 at the international level, however, the three countries in the Asian region re-administered the ICCS 2009 regional questionnaire as a national option under their own responsibility.

with the aim to discuss any proposed new materials in terms of, for example, relevance, clarity and answerability. The field trial followed a quantitative approach that required all participating countries to run a trial data collection according to the standardized procedures outlined in manuals and other related guidance materials.

The main survey data collection eventually took place in the 24 participating countries between October 2015 and June 2016. The survey was carried out in countries with a Southern Hemisphere school calendar between October and December 2015, and in those with a Northern Hemisphere school calendar between February and June 2016. These collection phases were followed by data processing and cleaning starting in mid-2016, then weighting, adjudication item analysis, scaling, and eventually reporting in November of 2017. The further chapters in this report provide a detailed account of these activities.

Links and key changes from ICCS 2009

ICCS builds on previous IEA studies of civic education and is a response to the challenge of educating young people in changing contexts of democracy and civic engagement. 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, Schwille, & Amadeo, 1999); an additional survey, of upper-secondary students, took place in 2000 (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002). A technical report detailed the design and implementation (Schulz & Sibberns, 2004).

ICCS 2009 was designed in a way that provided explicit links to CIVED, yet more as a baseline study for continuous and sustainable study cycles. The main results from this study were reported in Schulz, Ainley, Fraillon, Kerr, & Losito (2010). Like its predecessor, ICCS 2009 included a student test of civic knowledge and understanding, as well as questionnaires for students, teachers, and school principals. ICCS 2009 adopted the term *civic and citizenship education* to emphasize a broadening of the concepts, processes, and practices that had occurred in this area of educational provision since the turn of the century. As is common in IEA studies, a technical report documented the design and implementation of ICCS 2009 (Schulz, Ainley, & Fraillon, 2011).

The test design established for ICCS 2009 included provision for a set of secure common items that makes it possible to compare the test performance of students in countries participating in across ICCS cycles. Twenty-one of the countries that participated in ICCS 2009 also participated in ICCS 2016. About half of the student test items used in ICCS 2016 were secure items from ICCS 2009. The ICCS 2016 questionnaires, in particular the international and regional student instruments, also include a large number of items (with identical format and wording) which allow a review of changes over time between the first two ICCS cycles.

Although 21 countries participated in both ICCS 2009 and ICCS 2016, the international and regional reports present only the changes for those countries where data collection met the technical standards associated with sampling, instrument preparation, field operations, scoring, and data management during both cycles. This means that the reporting of changes over time does not cover all 21 countries or all questions and instruments. The number of countries included in comparisons of data collected by the various questions and instruments consequently vary.

Outline of the technical report

This report is structured so as to provide comprehensive technical information about the conceptual, methodological, and analytical aspects of ICCS 2016. Following this overview there is a series of chapters that provide detailed information about different aspects of ICCS 2016 and its implementation in a chronological order from the conceptual and instrument development at the beginning of the study to data processing, analysis and reporting towards the end.

- Chapters 2, 3, and 4 are concerned with the development, adaptation and translation of instruments. Chapter 2 provides information about the development and properties of the ICCS 2016 cognitive test, while Chapter 3 describes the development and properties of the international questionnaires, including the student questionnaire, the teacher questionnaire, the school questionnaire, the European student questionnaire, the Latin American student questionnaire, and the national contexts survey. Chapter 4 describes the procedures for the translation and adaptation of any instruments. A description of the actual adaptations made to the international materials is included in the study's user guide (Köhler et al., 2018).
- Chapters 5 and 9 are concerned with aspects of sampling before and after collection. Chapter 5 describes the sampling design, national variations, and implementation, while Chapter 9 documents the weighting and adjudication procedures that were used to ensure the results from ICCS 2016 represented the defined populations in each country.
- Chapters 6 and 7 are concerned with the survey implementation in the field. Chapter 6 describes the field operation procedures and the process of preparing data files at the country level. Chapter 7 documents the quality control protocols and observation procedures that were used during the ICCS 2016 data collection.
- Chapters 8, 10, 11 and 12 document the post-collection work, in particular the psychometric and statistical analyses used in this study. Chapter 8 provides an account of data management and the iterative building of the international databases. Chapter 10 reports on the scaling procedures for the cognitive test, so how responses to the items were used to construct scores on the civic knowledge scale. Chapter 11 describes the methods and procedures used to derive scales from sets of questionnaire items, and Chapter 12 provides an account of how the study results were reported, including the types of statistical analyses that were conducted to produce the reporting tables.

Finally, the technical report also contains a set of appendices including:

- Lists of the organizations and individuals involved in ICCS (Appendix A);
- Characteristics of national samples (Appendix B);
- Descriptions of civic knowledge test items and their allocations to proficiency levels, noting that the majority of items will remain secure for use in future ICCS cycles (Appendix C);
- A mapping of questionnaire items common to ICCS 2009 and 2016 (Appendix D); and
- Tables with coding information for the items in the questionnaires and the test (Appendix E).

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).

Köhler, H., Weber, S., Brese, F., Schulz, W., & Carstens, R. (Eds.). (2018). *ICCS 2016 user guide for the international database*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Losito, B., Agrusti, G., Damiani, V., & Schulz, W. (2018). *Young people's perceptions of Europe in a time of change. IEA International Civic and Citizenship Education Study 2016 European report*. Cham, Switzerland: Springer. Retrieved from https://www.springer.com/gp/book/9783319739595.

Schulz, W., Ainley, J., Cox, C., & Friedman, T. (2018a). *Young people's views of government, peaceful coexistence and diversity in five Latin American countries. The International Civic and Citizenship Education Study 2016 Latin American report*. Cham, Switzerland: Springer. Retrieved from https://www.springer.com/gp/book/9783319953922.

Schulz, W., Ainley, J., & Fraillon. J. (Eds.) (2011). *ICCS 2009 technical report*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

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 thirty-eight countries*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA International Civic and Citizenship Education Study 2016: Assessment framework*. Cham, Switzerland: Springer. Retrieved from: http://www.springer.com/gp/book/9783319393568.

Schulz, W., Ainley, J., Fraillon, Losito, B., Agrusti, G., & Friedman, T. (2018b). *Becoming citizens in a changing world. IEA International Civic and Citizenship Education Study 2016 International Report*. Cham, Switzerland: Springer. Retrieved from https://www.springer.com/gp/book/9783319739625.

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. (Eds.) (2004). *IEA civic education study: Technical report*. 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 case studies from the IEA Civic Education Project*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

CHAPTER 2:

ICCS test development

Julian Fraillon

Overview

The ICCS civic knowledge assessment was developed over an 18-month period from February 2013 to August 2014. The ICCS 2016 test 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 2016 field trial and main survey (Table 2.1).

Table 2.1: Test development process timeline

Year	Month	Group	Activity
2013	February	ICCS International Study Center	Establishment of test specifications
2013	June	First meeting of National Research Coordinators (Hamburg)	Review of proposed test development process and test specifications Item development workshop
2013	June	ICCS International Study Center	Drafting, review and refinement of test items Call for item submissions by NRCs and PAC members
2014	January	National Research Coordinators and Project Advisory Committee	Web-based item review
2014	February	Selected countries	Pilot testing
2014	March	ICCS International Study Center	Item revision based on web-based review and pilot testing
2014	April	First meeting of Project Advisory Committee (Philadelphia)	Review of items proposed for inclusion in field trial test and confirmation of test design
2014	April	ICCS International Study Center	Item revision based Project Advisory Committee review
2014	May	Second meeting of National Research Coordinators (Brussels)	Review of items proposed for inclusion in field trial test and confirmation of test design
2014	July	ICCS Field Trial Scoring Trainers (Hamburg)	Review of field trial scoring guides for constructed-response items (as part of scoring training)
2014	July	ICCS International Study Center	Revision of field trial scoring guides for constructed-response items
2015	February	ICCS International Study Center	Analysis of field trial item data and recommendations for items to be included in main survey test (field trial analysis report)
2015	April	Second Meeting of Project Advisory Committee (Chicago)	Review of field trial analysis report and recommendations for test design and items proposed for inclusion in main survey
2015	May	Third meeting of National Research Coordinators (Tallin)	Review of field trial analysis report and recommendations for test design and items proposed for inclusion in main survey
2015	July	ICCS Main Survey Scoring Trainers (Hamburg)	Review of main survey scoring guides for constructed-response items (as part of scoring training)

Test scope and format

ICCS Assessment framework

The cognitive test items for this study were developed with reference to the ICCS 2016 assessment framework (Schulz, Ainley, Fraillon, Losito, & Agrusti, 2016) and designed to measure a single trait labeled *civic knowledge* in the international reports on ICCS 2016 (Schulz, Ainley, Fraillon, Losito, Agrusti, & Friedman, 2018). 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 2016 was mapped to both the civics and citizenship content and cognitive process required by students to respond correctly to the item.

The ICCS 2016 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 applying.

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 1.2 items per unit in the main survey test instrument.

Two item formats were used: 79 of 88 test items in the final item pool had a multiple-choice format with four response options; the remaining nine items were constructed-response items for which students were required to write between one and three sentences.

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.

Test specifications planning

The item development process began formally with the development of draft test specifications. These outlined the proposed approximate number of items planned for use in the main survey, incorporating the available secure trend items from ICCS 2009, as well as specifying the target number of items by format (multiple-choice or constructed-response) and by the content and cognitive domains specified in the assessment framework (Tables 2.2 and 2.3).

The test specifications were developed with the intention of having the ICCS 2016 main survey item pool match the proportions of items by domain and item format match those established for ICCS 2009 (see final two columns of Table 2.3). The proposed test specifications and item development plan were presented to and approved by NRCs at the first meeting of NRCs in June 2013.

Table 2.2: ICCS 2016 item development plan by item format

Items	Pilot study	Field trial	Main study
Secure ICCS 2009 trend items (total –90% multiple choice, 10% constructed response)		~22	~30—35+
New multiple choice items	~100	~80	~50
New constructed response items	~10	~8	~4—5
Total	~110	~110	~84—90

Table 2.3: ICCS 2016 item development plan by content and cognitive domain

Domain		ICCS 2016 new items (*n*)			ICCS 2009 secure trend items (*n*)	ICCS 2016 main survey items		ICCS 2009 items (%)
		Pilot	Field trial	Main survey		Total (*n*)	Total (%)	
Content domain	1: Civic society and systems	52	42	26	8	34	40	39
	2: Civic principles	32	26	16	12	28	33	33
	3: Civic participation	20	16	9	10	19	22	23
	4: Civic identities	6	4	3	2	5	6	6
	Total	110	88	54	32	86	100*	100*
Cognitive domain	1: Knowing	37	29	18	3	21	24	24
	2: Reasoning and applying	73	59	36	29	65	76	76
	Total	110	88	54	32	86	100*	100*

Note:
*Some percentages may not add to 100%, due to rounding.

Incorporating ICCS 2009 secure trend items in ICCS 2016

ICCS 2009 included a set of 17 secure items from CIVED (the previous IEA Civic and Citizenship Education Study conducted in 1999). These 17 items were used as the basis for reporting achievement trends on the CIVED scale in countries that participated in both CIVED and ICCS 2009 (see Schulz, Ainley & Fraillon, 2011). The 17 CIVED items were included as a single cluster in ICCS 2009 and the entire cluster was released following the publication of the ICCS 2009 results.

The ICCS 2009 and ICCS 2016 assessment frameworks specified broader test content than that of CIVED. Sixteen of the CIVED items included in ICCS 2009 were mapped to the *civic society and systems* content domain and 14 of the 17 items mapped to the *knowing* cognitive domain in the ICCS 2009 and the ICCS 2016 assessment frameworks. All 17 items were multiple-choice.

When the ICCS 2009 items were developed, the balance of items by content and cognitive domains, and by item format planned for in the test specifications was met for the total item pool including the 17 CIVED items. The net result of this, together with the inclusion of the CIVED items in ICCS 2009 as an intact cluster (and their subsequent release), meant that the ICCS 2009 secure item pool and test clusters available for use in ICCS 2016 had a relative under-representation of items mapping to the *civic society and systems* content and *knowing* cognitive domains when compared to the test specifications. We planned to correct the imbalance in the item pool by developing slightly higher proportions of items mapping to the *civic society and systems* content and *knowing* cognitive domains than the specified total target proportions in the framework (see Table 2.3).

In order to create test clusters that met the planned test specifications for content, cognitive processes and item format, we decided to disassemble the ICCS 2009 clusters and redistribute the secure trend items across the ICCS 2016 field trial item clusters. This would result in each ICCS 2016 item cluster having a combination of new and secure trend items. We planned to use the field trial data to determine whether or not the reconfiguration had any effect on the measurement properties of the ICCS 2009 secure trend items. If no effect was found then we planned to use a similar reconfiguration of the clusters for the ICCS 2016 main survey. The IEA Technical Executive Group (TEG) and ICCS 2016 NRCs supported this decision.

NRC item development workshop

An item development workshop was conducted as part of the first meeting of NRCs in June 2013 in Hamburg. At this workshop, national representatives were provided with information about the framework and procedures for ICCS 2016 test development, as well as the test specifications. 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 review of the test specifications and 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 formats, 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 develop cognitive test items and submit them to the ISC for further 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 their units 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 of test developers (between three and six) 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. A robust discussion is required 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 2016. 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 2016 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 2016 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 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

Once all ICCS 2016 draft test material (stimulus, items and scoring guides) had been developed, it was made available to PAC members and NRCs for review. Reviewers were invited to make comments on the items and complete a brief rating (1 to 4) of the suitability of each item for inclusion in the test. The rating categories were:

4: Item should be included in the test without being altered.

3: Item should be included in the test with some minor changes.

2: Item should only be included in the test if specific changes are made.

1: Item should not be included in the test.

The web-based external review took place in January and February 2014, after which the test draft material was further revised in accordance with the feedback provided following the review.

Pilot testing

Small-scale pilot testing with convenience samples of students was conducted in six countries (Australia, Belgium [Flemish], Chile, Colombia, Malta and Slovenia). Three non-overlapping test forms were created with a total of 129 items (120 newly developed items and nine secure trend items). The test items were translated into the target language at national centers (at this stage without translation verification or adaptation review) and administered to students under test conditions.

The items were scored and the resulting pilot data were returned to the ISC for review. There were not sufficient numbers of students to support scaling analysis, but response frequencies for multiple-choice options and proportions of correct responses for all items were checked as part of the review. In addition, national centers were invited to provide qualitative feedback from post-test interviews with students.

Information collected from the pilot was used to inform further revision of the items and the 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 112 items, including 22 items that were secure trend items from the ICCS 2009 (Table 2.4).

Table 2.4: Composition of the field trial test instrument by item format and origin

Item format	New items (*n*)	Trend items (*n*)	Total items (*n*)	Total score points (*n*)	Score points (%)
Multiple choice	80	20	100	100	83
Constructed response	10	2	12	20	17
Total	90	22	112	120	100

The items were allocated to 10 clusters (C1 to C10) that were presented in a fully balanced rotated test design across 10 test booklets (B1 to B10). The ICCS 2009 trend items were distributed across all 10 clusters. This allowed for the clusters to be balanced for content and cognitive processes, reading load and item format (Table 2.5). A rotated cluster test design was used in the field trial (see Tables 2.5 and 2.6).

Table 2.5: Field trial cluster composition

Cluster*	Multiple choice items	Constructed response items
C1	10	1
C2	10	1
C3	10	1
C4	10	1
C5	10	1
C6	10	2
C7	10	1
C8	10	2
C9	10	1
C10	10	1

Table 2.6: Field trial test booklet design

Booklet	Position		
	1	2	3
B1	C1	C2	C4
B2	C2	C3	C5
B3	C3	C4	C6
B4	C4	C5	C7
B5	C5	C6	C8
B6	C6	C7	C9
B7	C7	C8	C10
B8	C8	C9	C1
B9	C9	C10	C2
B10	C10	C1	C3

Note:
See Table 2.5 for information about the field trial cluster composition.

Coverage of the assessment framework

The ICCS 2016 assessment framework contained a test development plan for content and cognitive domains (Table 2.7). This includes the target percentages of item content for the sub-domains within each domain. The final set of field-trial test items contained slightly lower proportions of items addressing the content domains civic participation and civic identities in the ICCS 2016 assessment framework (Table 2.8).

Table 2.7: ICCS 2016 test development plan

Domain	Sub-domain	All items (%)
Content domain	Civic society and systems	40
	Civic principles	30
	Civic participation	20
	Civic identities	10
	Total	100
Cognitive domain	Knowing	30
	Reasoning and applying	70
	Total	100

Table 2.8: Field trial item mapping to assessment framework

Domain	Sub-domain	New items (*n*)	Trend items (*n*)	Total items (*n*)	All items (%)
Content domain	Civic society and systems	41	3	44	39
	Civic principles	27	9	36	32
	Civic participation	17	7	24	21
	Civic identities	5	3	8	7
	Total	90	22	112	100
Cognitive domain	Knowing	34	2	36	32
	Reasoning and applying	56	20	76	68
	Total	90	22	112	100

The decision to have relatively lower proportions of items addressing these content domains 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. Nonetheless, field-trial items achieved the planned coverage of the content and cognitive domains in the assessment framework.

Main survey

Selection of items

We evaluated the measurement properties of the field trial test items based on the data collected among students in all participating countries. The analysis procedures used to review measurement properties are described in Chapter 11. We reviewed those items with unsatisfactory measurement properties further to determine whether they could be revised or had to be deleted from the item set. We made minor revisions to a small number of items. Items were only modified when there was clear evidence that the revision would improve their measurement properties. Twenty-four items were removed from the field trial test item pool, leaving 80 newly developed items available for inclusion in the main survey test.

Test design and content

The main survey test consisted of 88 items, including 42 items that were secure items from ICCS 2009 and 46 items newly developed for ICCS 2016 (Table 2.9).

Table 2.9: Composition of the main study test instrument by item format and origin

Item format	New items (*n*)	Trend items (*n*)	Total items (*n*)	Total score points (*n*)	Score points (%)
Multiple choice	41	38	79	79	83
Constructed response	5	4	9	16	17
Total	46	42	88	95	100

For the main survey, test items were allocated to eight clusters (C1 to C8) that were assembled into a fully balanced rotated test design comprising eight test booklets (B1 to B8), each with a testing time of 45 minutes.

Data from the field trial analyses showed that reconfiguring the clusters (see section on test specifications planning) to include both ICCS 2009 secure trend and newly developed items for ICCS 2016 did not affect the measurement properties of the ICCS 2009 trend items. As a consequence, we decided to create the ICCS 2016 main survey clusters with similar combinations of secure trend and newly developed items (Table 2.10). A balanced rotated cluster test design was used in the main survey (Table 2.11).

Table 2.10: Main study cluster composition

Cluster	Multiple choice items	Constructed response items	New items	Trend items
C1	10	1	5	6
C2	10	1	6	5
C3	9	2	6	5
C4	10	1	6	5
C5	10	1	6	5
C6	10	1	5	6
C7	10	1	6	5
C8	10	1	6	5

Table 2.11: Main study test booklet design

Booklet	Position		
	1	2	3
B1	C1	C2	C4
B2	C2	C3	C5
B3	C3	C4	C6
B4	C4	C5	C7
B5	C5	C6	C8
B6	C6	C7	C1
B7	C7	C8	C2
B8	C8	C1	C3

Mapping to framework

The main survey test items were mapped to the ICCS assessment framework (Table 2.12).

Table 2.12: Main study item mapping to assessment framework

Domain	Sub-domain	New items (*n*)	Trend items (*n*)	Total items (*n*)	All items (%)
Content domain	Civic society and systems	24	11	35	40
	Civic principles	12	18	30	34
	Civic participation	8	10	18	20
	Civic identities	2	3	5	6
	Total	46	42	88	100
Cognitive domain	Knowing	18	5	23	26
	Reasoning and applying	28	37	65	74
	Total	46	42	88	100

The final test instrument provided a good coverage of the assessment framework and matched proportions of items by content and cognitive domain from the item development plan (see Table 2.3). The only difference was the decision to use a higher than originally planned proportion of secure trend items in the ICCS 2016 instrument. The original test development plan was established to allow the release of a third cluster of ICCS 2009 secure test items. Once the IEA decided to keep all available ICCS 2009 test material secure, we decided to include 42 secure trend items in the ICCS 2016 instrument to maximize the number of items available for use in equating ICCS 2016 data with those from ICCS 2009.

Released test items

Two clusters of test items have been released following publication of the ICCS 2016 international report. The two released clusters were main survey clusters C6 and C7 (see Table 2.13 for a summary of the released items by item format and the coverage of the ICCS 2016 assessment framework by the released item material).

Table 2.13: Characteristics and mapping of ICCS released test items

Characteristic		Released cluster C6 items (*n*)	Released cluster C7 items (*n*)	Total items (*n*)	Total items (%)
Item format	Multiple choice	10	10	20	91
	Constructed response	1	1	2	9
	Total	11	11	22	100
Content domain	Civic society and systems	3	3	6	27
	Civic principles	5	5	10	45
	Civic participation	3	1	4	8
	Civic identities		2	2	9
	Total	11	11	22	100
Cognitive domain	Knowing	6	1	7	32
	Reasoning and applying	5	10	15	68
	Total	11	11	22	100

References

Schulz, W., Ainley, J., Fraillon, Eds. (2011). *ICCS 2009 Technical Report*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA International Civic and Citizenship Education Study 2016: Assessment framework*. Cham, Switzerland, Springer. Retrieved from: http://www.springer.com/gp/book/9783319393568.

Schulz, W., Ainley, J., Fraillon, Losito, B., Agrusti, G., & Friedman, T. (2018). *Becoming citizens in a changing world. IEA International Civic and Citizenship Education Study 2016 International Report*. Cham, Switzerland: Springer. Retrieved from: https://www.springer.com/gp/book/9783319739625.

CHAPTER 3:

ICCS questionnaire development

Gabriella Agrusti, John Ainley, Bruno Losito and Wolfram Schulz

Introduction

This chapter describes the development of the international questionnaires for students, teachers, schools and national research centers in the second cycle of ICCS. The student questionnaire was designed to measure contextual information, as well as aspects related to students' attitudes and engagement. 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, climate and 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 sub-regional) level about the characteristics of education systems, aims, contexts and implementation of civic and citizenship education, as well as 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 2016 (Schulz, Ainley, Fraillon, Losito, & Agrusti, 2016). The assessment framework consisted of two parts:

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

The ICCS 2016 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 aspects of students' attitudes and engagement that ICCS 2016 set out to measure.

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 composed of a set of sub-domains that incorporated elements referred to as "aspects" and "key concepts".

- *Civic society and systems: three sub-domains.* (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: four sub-domains.* (i) equity (all people having the right to fair and just treatment); (ii) freedom (of belief, of speech, from fear, and from want); (iii) social cohesion (sense of belonging, connectedness, and common vision amongst individuals and communities within a society); and (iv) rule of law.
- *Civic participation: three sub-domains.* (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: two sub-domains*. (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 and the civic roles individuals play within each community).

The assessment framework identified the two different types of student beliefs, perceptions and behaviors relevant to civics and citizenship. We identified two affective-behavioral domains: attitudes, and engagement.

- *Attitudes:* these include beliefs about democracy and citizenship, attitudes toward equal rights for groups in society, perceptions of threats to the world's future, trust in groups and institutions, as well as attitudes toward the country of residence.
- *Engagement:* these refer to students' self-confidence in undertaking civic activities, past and current engagement at school or in the community, as well as their expectations of future civic action, including constructs such as preparedness to participate in legal or illegal activities to express their opinion, anticipated future political participation as adults, and anticipated future participation in citizenship activities at school.

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, supra-national 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 (these 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 variables such as 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 are variables related to civic-related learning and the acquisition of its outcomes, and they 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).

A variety of contextual factors potentially influence the learning outcomes of civic and citizenship education (Figure 3.1). The framework assumes a reciprocal relationship between *processes* and *outcomes* (as indicated by the double-headed arrow in Figure 3.1). 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. Relationships between *antecedents* and *processes* are assumed to be unidirectional (as denoted by a single-headed arrow in Figure 3.1).

The ICCS researchers collected variables (or groups of variables) through their use of the various ICCS 2016 instruments (Table 3.1). 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, and about some process-related variables (e.g., learning activities). The student test and the student perceptions questionnaire 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.

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

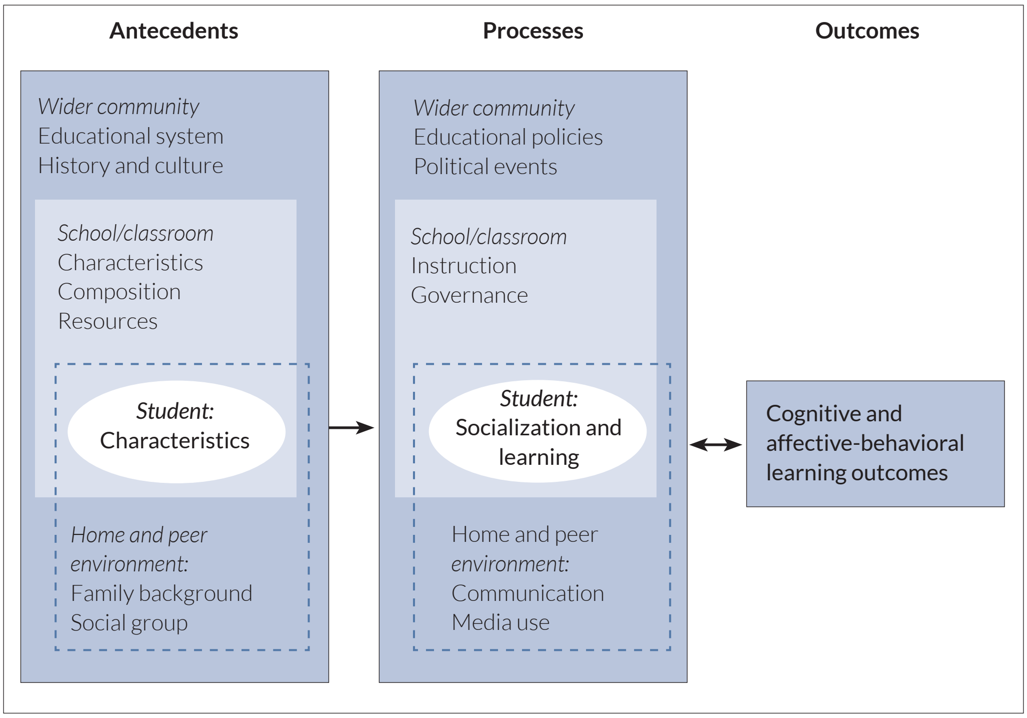


Table 3.1: Mapping of variables to contextual framework (examples)

Level of...	Antecedents	Processes	Outcomes
Wider community	NCS and other sources: Democratic history Structure of education	NCS and other sources: Intended curriculum Political developments	StT and StQ/RQ: Test results Student attitudes and engagement
School/classroom	ScQ and TQ: School characteristics Resources	ScQ and TQ: Implemented curriculum Policies and practices	
Student	StQ: Gender Age	StQ: Civic engagement Practiced engagement	
Home and peer environment	StQ: Parent SES Ethnicity Language Country of birth	StQ: Family communication Communication with peers Media information	

Notes:
NCS = national contexts survey, ScQ = school questionnaire, TQ = teacher questionnaire, RQ = regional questionnaires, StQ = student questionnaire, StT = student test, and SES = socioeconomic status.

The context of the wider community can be viewed as multi-layered: the local community comprising the students' schools and home environments, which, in turn, are embedded within the broader regional, national, and (possibly) supra-national contexts. Within the scope of ICCS 2016, the level of the local community and the level of the national context were the most relevant levels.

Student questionnaire development

The development of the international student questionnaire was conducted in three phases:

- The first phase included discussions about retaining or modifying ICCS 2009 questionnaire items and reviews of first new draft material by national centers and experts, accompanied by piloting of a draft questionnaire in five countries;
- In the second phase, first draft material was finalized following an expert and national center review and tested in an international field trial undertaken in all participating ICCS 2016 countries;
- During the final phase, the results from the field trial were discussed with experts and national centers, and a final selection of main survey items was undertaken.

During each of these phases, the following criteria were applied when selecting proposed item material:

- Relevance with regard to the ICCS 2016 assessment framework;
- Appropriateness for the national contexts in participating countries;
- Psychometric properties of items that were designed to measure latent traits postulated in the initial formulation and found in pilot or field trial data.

Piloting of international student questionnaire material was carried out in five countries (Australia, Belgium [Flemish], Chile, Colombia, and Malta). Two forms each of about 20 minutes duration were administered to convenience samples of target grade students to pilot a total of 252 affective-behavioral questionnaire items. The results from this pilot study were used in conjunction with feedback from national centers and experts to elaborate a draft student questionnaire for the field trial.

The ICCS 2016 field trial student questionnaire included a total of 247 items (including 24 items that were optional for countries) and was administered to samples in all 24 participating countries. Three different questionnaire forms were used to trial a larger pool of questionnaire items than would have been possible in a single form. We allocated items to these forms in a way that allowed analysis of all possible combination of item sets and scales. In each participating country, we collected data from about 600 students.

The analyses of field trial data were designed to provide empirical evidence for the selection of the main survey material with a particular emphasis on the review of the cross-national validity of measures derived from the ICCS 2016 questionnaires (see a discussion of different approaches in Schulz, 2009). Field trial outcomes and a draft student questionnaire for the final data collection proposed by the international study center were discussed with national coordinators and experts before the final selection.

The final international student questionnaire consisted of 179 items. Twenty-two of these items were designed to capture student background information and 157 were designed to measure the affective-behavioral domains specified in the ICCS 2016 assessment framework. Another 26 items in the ICCS 2016 student questionnaire were optional, and 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:* questions about the students' age, gender and expected education;
- *Your home and your family:* questions about characteristics of the household and the students' parents;
- *Your activities outside school:* questions about activities reported by students at their home, peer and community environment;

- *Your school:* questions about students' reports on different aspects of their schools, as well as their participation in civic activities at school;
- *Citizens and society:* questions about democracy and the importance of different behaviors for good citizenship;
- *Rights and responsibilities:* questions designed to measure student attitudes toward equal rights for gender groups, ethnic/racial groups and immigrants, and perceptions of how to avert threats from political violence;
- *Institutions and society:* questions regarding student trust in civic institutions, perceptions of their country of residence, and their level of concern regarding different potential threats to the world's future;
- *Participation in society:* questions about students' self-confidence with regard to active participation, and about their expectations of future participation in different civic activities; and
- *You and religion:* questions pertaining to the international option about religious background and practices, and about their attitudes toward the influence of religion on society.

Most of the items were developed by the international study center, but the national centers also proposed new items or modifications to student questionnaire material that were included in the final survey instrument.

Optional items were designed to capture variables that were perceived as either not relevant or inappropriate in some of the participating countries, but regarded as crucial in a large number of other countries. In these cases, single questions or sets of questions were included as international options where each national center could choose whether it wanted to administer this material. In addition, there was some interest in measuring aspects of the European region (for example, trust in European institutions) within the context of the international student questionnaire, and a number of optional European items were added to some of international item sets.

The following international options were offered to countries participating in ICCS:

- Students' ethnicity;
- Household composition (people living with the student at home); and
- Religion.

Five national centers chose to include the item on ethnicity, 23 national centers opted to include the item on household composition, and 19 chose to include items measuring student perceptions of religion in the student questionnaire.

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 classroom levels. The relationships between schools and local communities included civic-related activities carried out by teachers with their target grade students within the local 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, we included some questions about school climate, social relations at school, environmental sustainability, the priority assigned to civic and citizenship education and on its different objectives in both teacher and school questionnaires.

Under the assumption that teaching staff constitute an important factor in determining school climate and culture, and the students' school experience, the teacher questionnaire was designed to be completed by teachers teaching across all subject areas in the school curriculum for the ICCS target grade. The maximum time envisaged 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. This international option was chosen by all participating countries. It included questions about teaching and assessment approaches to civic and citizenship education, teachers' preparedness and training in civic and citizenship education.

The questionnaire development process took place in four phases:

- The first phase included development and review of draft material by the international project team and experts. The first questionnaire draft was piloted in a small-scale pilot study implemented in four countries (Belgium [Flemish], Colombia, Finland, and Italy). The pilot included discussions of draft material with focus groups consisting of teachers who had completed the teacher questionnaire, following guidelines provided by the international project team. The results from the pilot were used to refine the item material.
- In the second phase, material from the pilot was revised, and we developed a first draft field trial questionnaire, which was subsequently reviewed by national centers and by experts from the Project Advisory Committee.
- During the third phase, the revised item material was administered to samples of teachers as part of the international field trial in all of the participating countries.
- The final phase comprised a review of the results from the international field trial and the final selection of main survey item material following discussions with the Project Advisory Group and national center staff.

For the selection of main survey item material, we applied similar criteria to those used for the student questionnaire (see previous section). When selecting items we considered the empirical evidence from the field trial in conjunction with the national research centers' review of the appropriateness of the questionnaire material for the diversity of national contexts, in particular in light of existing differences between participating education systems, as well as between schools within each system.

The field trial teacher questionnaire consisted of 24 questions, with a total of 141 items and one filter question introducing the international option. It was administered to target grade teachers in the schools selected for the field trial. Sixty of these items were included in the international option for teachers of civic-related subjects. On average, field trial teacher samples consisted of about 300 teachers in each participating country.

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

- *General:* questions about teacher background characteristics;
- *The school:* questions about the school environment and issues related to participation in teaching and learning activities;
- *Civic and citizenship education at school:* questions about the delivery of civic and citizenship education at a school level; and
- *Teaching of civic and citizenship education:* questions for teachers of subjects directly related to civic and citizenship education and offered as an international option to countries.

School questionnaire development

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

The school questionnaire collected data about variables related to the school context. These included school characteristics, such as: school size and resources; the school environment; school autonomy in planning civic and citizenship education; student, teacher, and parent participation in the running of the school; social relations at school, and teachers' and students' sense of belonging to the school; initiatives related to environmental sustainability; and approaches to civic and citizenship education.

Furthermore, it also measured aspects related to the local community context, such as resources available to students in the local area, and issues of social tension within the local community.

Some of the constructs measured through the school questionnaire were also assessed through the teacher questionnaire, with the aim of collecting data about these issues which considered the perspectives of both teachers and school principals.

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

The questionnaire development process took place in four phases:

- The first phase included discussions and reviews of the draft material with the Joint Management Committee and international experts. The first questionnaire draft was piloted in a small-scale pilot study implemented in five countries (Belgium [Flemish], Colombia, Finland, Italy, and Malta), which relied mainly on focus groups with small groups of school principals, following the guidelines provided by the Joint Management Committee.
- In the second phase, material from the pilot was revised and a subsequent first draft of the field trial school questionnaire was reviewed by national centres and by the experts on the Project Advisory Committee.
- During the third phase, draft material was administered to smaller samples of schools in an international field trial undertaken in all participating countries.
- The final phase consisted of a review of field trial results, which formed the basis for the final selection of main survey questions and items. The selection took place following discussions with the Project Advisory Group and national center representatives.

During each of these phases, we adopted the same criteria used for the other questionnaires when selecting material. During the process of instrument development, we gave particular attention 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 field trial version of the school questionnaire included 20 questions with a total of 107 items and was administered to the principals of schools that were selected in all ICCS countries that participated in the field trial. In most countries, principals from 20–30 schools 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, there were some limitations with regard to the scope of analyses that could be carried out with the field trial data gathered with this instrument. The school questionnaire results from the field trial were discussed with national center representatives and experts of the Project Advisory Group prior to the final selection of item material for the main survey. The revisions following the field trial also included a rewording of some of the items.

The final school questionnaire consisted of 21 questions containing a total of 106 items, and was divided into the following five sections:

- *General:* this section included only one question about principals' career;
- *The school environment:* questions about school governance and students', teachers' and parents' participation at school; teachers' and students' sense of belonging to the school; social relations at school; and initiatives related to environmental sustainability;
- *The local community:* 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:* questions about how civic and citizenship education was implemented at the school; and
- *School size and resources:* questions about basic school characteristics like school size, numbers of teachers or school location.

European questionnaire development

The European questionnaire was developed in accordance with the ICCS 2016 assessment framework (Schulz et al., 2016) with a focus on elements that were viewed as pertinent to the region. Ideas for the incorporation of region-specific aspects in the ICCS 2016 framework were developed by the international project team by identifying potential elements for inclusion from a review of existing developments and mapping these against the broader ICCS assessment framework. This identification of relevant aspects was helpful to determine the attitudes to be measured by the regional instrument in addition to those already measured within the international student questionnaire. ICCS researchers then discussed the conceptual ideas, their mapping to the broader framework and draft material in a series of meetings with European national research coordinators (NRCs).

All decisions about the scope and focus of the regional instrument were made in close cooperation between the international project team and national center representatives. It was decided that the regional student instrument would only consist of a student questionnaire in contrast to ICCS 2009, where both cognitive and questionnaire items were included. The European student questionnaire was developed to focus on the following specific regional issues deemed to be of high importance within the regional context: European identity, opportunities for learning about Europe at school, immigration and freedom of migration within Europe, cooperation among European countries, political and ethical consumerism, the European Union, and the future of Europe.

The development of the European instrument was a collaborative effort undertaken in in close cooperation between the international project team, the Project Advisory Group and national centers. The development process consisted of four interrelated phases:

- International project staff made recommendations for the initial instrument development and discussed them with country representatives at the first NRC meeting in Hamburg and during a meeting of representatives from European ICCS 2016 countries in Rome in December 2013. NRCs provided suggestions for a refinement of draft proposals on both occasions. Several countries also submitted materials to be included in the questionnaire.

- A first version of the questionnaire was drafted by international project staff based on the feedback received from country representatives.
- The draft version of the European questionnaire was tested in an international field trial. Except the Russian Federation, all of the European countries participating in ICCS 2016 administered the field trial version of the European student questionnaire.
- The international project team conducted a final revision of the European questionnaire, informed by the field trial analysis and results, in cooperation with the Project Advisory Group and national centers.

In each of these phases, we adopted the same criteria as for the other questionnaires when selecting material. The version of the European questionnaire administered in the field trial contained 13 questions comprised of 83 items. This version of the questionnaire also comprised a cognitive item asking students about their country's membership in the European Union. This cognitive item was not included in the final version of the instrument, since EU membership was widely acknowledged by almost all the students participating in the field trial.

The final European student questionnaire consisted of 11 questions comprising a total of 70 items related to:

- students' sense of European identity;
- students' reports on opportunities for learning about Europe at school;
- students' attitudes toward freedom and restriction of movement within Europe;
- students' attitudes toward equal rights for immigrants;
- students' attitudes toward cooperation among European countries;
- students' perceptions of discrimination in their country;
- students' expectations for European future;
- students' expectations for their own individual future;
- students' attitudes toward political and ethical consumerism;
- students' opinions on the age limits for different rights and obligations to be acquired; and
- students' attitudes toward European Union.

Latin American questionnaire development

The Latin American questionnaire used in ICCS 2016 was adapted from the instrument used in the previous cycle. Its content was linked to the ICCS 2016 assessment framework (Schulz et al., 2016) and it focused on aspects that were deemed particularly relevant to the Latin American context.

The development of the Latin American questionnaire was a collaborative effort undertaken under the supervision of the international study center involving the national research coordinators and advice from the Project Advisory Committee. The development process comprised three phases:

- The first phase involved the review of ICCS 2009 questionnaire items, and the modification and writing of additional item material: this work was guided by the ICCS assessment framework and was undertaken in extensive consultation with the national research coordinators.
- The second phase involved 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 was undertaken by the international study center and consultants in light of the field trial results and further feedback from national centers and experts.

The ICCS 2016 Latin American field trial questionnaire comprised nine questions consisting of 84 questionnaire items, and assessment time took 15 to 20 minutes. Field trial data from the five participating Latin American ICCS 2016 countries provided a basis for a review of the psychometric properties of the item material. The final item selection was undertaken in collaboration with the national research coordinators from the Latin American ICCS 2016 countries and informed by the empirical evidence derived from the field trial, as well as conceptual considerations.

For the final regional questionnaire used in the ICCS 2016 main survey, we retained nine questions with a total of 72 items; two items were optional for countries. Thirty-five items from four questions remained unchanged (both with regard to stem and item wording) and were designed to enable the measurement of changes in perceptions since 2009. The other 37 items from the remaining five questions were either new or modified from the previous ICCS 2009 questionnaire.

The final Latin American questionnaire had a stipulated assessment time of 15 minutes and addressed the following region-specific affective-behavioral aspects:

- Students' perceptions of government and law (attitudes toward authoritarian government, dictatorships, and corrupt practices);
- Students' perceptions regarding peaceful coexistence (attitudes toward violence, acceptance of disobedience to the law, and feelings of empathy); and
- Students' perceptions of discrimination in their country, acceptance of social minorities, and attitudes towards homosexuality.

Development and implementation of the national contexts survey

The ways in which students acquire civic knowledge and understanding, and the ways they develop dispositions toward civic society and civic engagement are likely to be strongly influenced by factors at the country or *national context* level. Factors of interest include, among others, the historical background, the nature of the political system, the structure of the educational system, and the curricular framework for civic and citizenship education. The national contexts survey was designed to systematically collect those relevant data and information at the country- or system-level about both antecedents and processes that were not readily available from published sources.

The development, coordination, analyses, verification and reporting of the national contexts survey was coordinated by ICCS researchers at the international study, and the development process was conducted in liaison with partner institutions and in cooperation with national research coordinators (NRCs) from participating countries. Staff at the IEA in Hamburg were responsible for the implementation of this data collection as an online survey completed by national centers and drawing on available expertise in their countries.

One of the major changes from ICCS 2009 was that, for ICCS 2016, there was only one data collection shortly after the conclusion of the main survey of schools, students and teachers. Another important change was a change in response formats by reducing the use of rating-scale formats, restricting the scope of questions to more factual aspects and by asking for a wider array of reference documents related to particular aspects of the survey; national centers had greater opportunities to provide open-ended responses for clarification.

The development process and implementation was conducted in four phases:

- The initial development phase was undertaken between the first and second meetings of national researcher coordinators, including initial discussion about the scope and a revision of the questions used in ICCS 2009.
- The second phase consisted of review of a draft questionnaire by national coordinators and experts from the Project Advisory Group, with the aim of further revision and refinement.

- The third phase consisted of extensive reviews by three NRCs from different types of education systems (who provided feedback on the appropriateness, completeness and relevance of questions, which was used to finalize the instrument).
- The fourth phase consisted in setting up the instrument as an online survey that was completed by the national centers after the respective main data collection from students, teachers, and schools in their countries.

During the process of the development of the content, we applied the following criteria when considered the contexts and questions to be included in the national contexts survey:

- Its relevance with regard to the ICCS assessment framework;
- Its additional value in relation to information about national contexts already in the public domain;
- Its appropriateness for the national contexts in participating countries; and
- Its validity in terms of comparability, analysis and reporting.

Based on feedback received from national centers and previously received expert advice, the international project staff modified the scope, content and format of the ICCS 2009 material as follows:

- It added content that was not present in the 2009 NCS but was deemed to be relevant;
- It retained, but modified, some questions where concerns about the format had been expressed by the national centers and experts from the Project Advisory Group;
- It removed some aspects included in the NCS in ICCS 2009 because these were no longer deemed as appropriate or relevant for ICCS 2016.

The ICCS 2016 National Contexts Survey included questions related to the following aspects of the national context in each participating country:

- Education system (background and structure of the education system);
- Civic and citizenship education in the curriculum (education policies, civic and citizenship education at school and at the target grade, current reforms and debates);
- Teachers and teacher education (general structure, teacher education for civic and citizenship education, in-service teacher education for civic and citizenship education); and
- Assessments and quality assurance.

The online facility enabled national center staff to complete the survey in several administration sessions (i.e., they could log in and out in order to complete the questionnaire as needed information became available).

The online data received from each of the 24 participating countries were thoroughly checked for consistency and plausibility by project staff at the international study center. National centers were invited to review draft tables, and corrections were applied where appropriate following feedback from participating countries. The international project team used the outcomes of the national contexts survey in conjunction with data from published sources to inform the descriptions of the education systems and contexts for civic and citizenship in ICCS 2016 countries (see Schulz, Ainley, Fraillon, Losito, Agrusti, & Friedman, 2018).

References

Schulz, W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study. *IERI Monograph Series*, *2*, 113–135. Retrieved from: http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_05.pdf.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *International civic and citizenship education study 2016: Assessment framework*. Cham, Switzerland: Springer. Retrieved from: http://www.springer.com/gp/book/9783319393568.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., & Friedman, T. (2018). *Becoming citizens in a changing world. IEA International Civic and Citizenship Education Study 2016 international report*. Cham, Switzerland: Springer. Retrieved from: https://www.springer.com/gp/book/9783319739625.

CHAPTER 4:

Translation and national adaptations of ICCS instruments

Gabriela Nausica Noveanu, Tim Friedman and Hannah Köhler

Introduction

This chapter describes the procedures and activities related to countries' preparation of national versions of the ICCS assessment instruments, focusing on the following major activities:

- Translation and adaptation of the international source version of the ICCS assessment instruments into national languages;
- International verification of the national translations and adaptations;
- International layout verification of the final national instruments; and
- Verification of online questionnaires.

In close collaboration with the participating countries, the international project team developed the international source version of the ICCS 2016 assessment and questionnaires in English. After the release of the international source version of these materials, all the participating countries were required to translate and/or adapt the international version into their language(s) of instruction. Linguistic and assessment experts performed multiple rounds of reviews to ensure that the translated national instruments were equivalent to the international source version.

The translation and verification processes aimed to ensure that national instruments incorporated high quality translations that were internationally comparable and were adapted appropriately for each country's context and education system. As part of the ICCS 2016 international quality assurance program, each country's instruments underwent a formal external review of the translations and adaptations prior to the assessment. The International Study Center (ISC) managed the adaptation consultation and also undertook layout verification of the final instruments. The IEA coordinated translation verification. These processes were all carefully documented during various stages of translation, adaptation and layout verification by the different parties involved in the instrument preparation, and occurred twice, firstly before the field trial and then again prior to the main survey.

Translation and adaption of ICCS 2016 instruments

ICCS 2016 instruments to be translated and adapted

The international project team provided NRCs with all materials to be adapted and translated, together with a spreadsheet to document each step of the adaptation, translation, and verification processes.

The following ICCS 2016 materials were required to undergo adaptation and translation:

- The student cognitive test (including covers and instructions);
- The international questionnaires for school principals, teachers, and students (including covers and introductions); and
- The regional student questionnaires (including covers and introductions), where applicable.

The ICCS 2016 assessment was designed so that each cluster of cognitive items appeared in multiple booklets. The item clusters, covers and instructions were prepared as separate files to facilitate translation. This approach allowed countries to translate each component only once before assembling the booklets. Finally, each booklet needed to be prepared as a separate file, including cover, instructions and three clusters of items.

The ICCS 2016 procedural manuals and scoring guides for the constructed-response items were typically translated, but were not subject to the international verification procedures.

Languages used in ICCS 2016

Identifying the language of the assessment (the target language) is relatively straightforward for most countries because it will be the dominant language used in both public and private sectors of society, including the education system. However, some countries use more than one language of instruction in their educational system. In such cases, the student instruments were commonly translated into the different languages of instruction to ensure that the assessment was appropriate for all students.

For ICCS 2016, a total of 23 countries and one benchmarking entity prepared 32 different sets of materials in 22 different languages (see Table 4.1). Of the participating entities, seven administered the instruments in more than one language. As all countries that participated in the Latin American module had Spanish as the assessment language, the source version of the questionnaire was developed in this language for use and adaptation by these countries.

All countries were required to follow the standardized, internationally agreed-upon procedures from the initial translation through to the final printing of their national instruments. At the national level, countries were responsible for translating and/or adapting the international cognitive test and questionnaires according to the international guidelines for ICCS 2016.

The IEA arranged for each country's translated and adapted instruments to undergo translation verification. When the verified materials were returned, the national research coordinators (NRCs) were tasked with reviewing the feedback of translation verification, revising their materials as needed, and updating their documentation for use during data processing and analyses.

Translation process

The IEA provided the countries with written guidance on the procedures required to translate the cognitive items and questionnaires appropriately. Each country was responsible for ensuring that skilled and experienced translators translated the instruments. To ensure that national versions of the ICCS instruments were consistent with the international source version, the translation guidelines allowed for national adaptations where necessary. Following translation of the instruments, one or more qualified reviewers independently reviewed the completed translations when necessary to ensure the nationally translated instruments were of the highest quality and appropriate for the intended audience. Some countries employed multiple translators and reviewers, either working together to complete the tasks, or working independently to provide two or more perspectives. Countries were responsible for ensuring consistency of the reviews across the translated materials.

If countries prepared translations in more than one language, IEA suggested that professionals familiar with the various languages be involved in order to ensure that the translations were equivalent across the national languages.

Countries were strongly advised to hire highly qualified translators and reviewers who were well suited to the task of working with the ICCS materials. Essential qualifications for translators and reviewers included:

- Excellent knowledge of English;
- Excellent knowledge of the target language;
- Experience in the country's cultural context; and
- Experience translating texts in the subject areas related to civics and citizenship.

Table 4.1: Languages used for the ICCS 2016 survey instruments

Country	Language	Instruments				
		Test	International student questionnaire	Regional student questionnaire	Teacher questionnaire	School questionnaire
Belgium (Flemish)	Dutch	•	•	•	•	•
Bulgaria	Bulgarian	•	•	•	•	•
Chinese Taipei	Chinese	•	•	n/a	•	•
Chile	Spanish	•	•	•	•	•
Colombia	Spanish	•	•	•	•	•
Croatia	Croatian	•	•	•	•	•
Denmark	Danish	•	•	•	•	•
Dominican Republic	Spanish	•	•	•	•	•
Estonia	Estonian	•	•	•	•	•
	Russian	•	•	•	•	•
Finland	Finnish	•	•	•	•	•
	Swedish	•	•	•	•	•
North-Rhine Westphalia (Germany)	German	•	•	•	•	•
Hong Kong SAR	Chinese	•	•	n/a	•	•
	English	•	•		•	•
Italy	Italian	•	•	•	•	•
Korea, Republic of	Korean	•	•	n/a	•	•
Latvia	Latvian	•	•	•	•	•
	Russian	•	•	•	•	
Lithuania	Lithuanian	•	•	•	•	•
	Polish	•	•	•		
	Russian	•	•	•		
Malta	Maltese	•	•	•		
	English	•	•	•	•	•
Mexico	Spanish	•	•	•	•	•
Netherlands	Dutch	•	•	•	•	•
Norway	Bokmål	•	•	•	•	•
	Nynorsk	•	•	•	•	•
Peru	Spanish	•	•	•	•	•
Russian Federation	Russian	•	•	n/a	•	•
Slovenia	Slovenian	•	•	•	•	•
Sweden	Swedish	•	•	•	•	•

Note:
n/a = not applicable.

The IEA also advised the countries that the translators and/or reviewers should have experience of test development and in working with students in the target grade. The reviewers were primarily responsible for assessing the readability and accuracy of the translation for the target population.

Guidelines for translation and adaptation

The general purpose of the translation and adaptation processes was to maintain the same meaning and level of difficulty as the international source version while concurrently following the rules of the target language and the country's cultural context to ensure the comparability of the national versions of the instruments. For countries with English language instruments, this included adapting the international source version to suit the national context. National centers needed to implement cultural adaptations: the content of the materials could be modified by deliberately altering some content to make the resulting materials more suitable for another sociocultural context.

Translators and reviewers were asked to ensure that:

- The translated text did not clarify or remove text from the source text, did not add more information, and was at an appropriate level for the target population;
- The translated text used terminology equivalent to that of the international version;
- The translated text had the same level of difficulty and register (language level and degree of formality) as the international version;
- The translated text used equivalent social, political, and historical terminology appropriate to the target language;
- Idiomatic expressions were translated appropriately, not necessarily word for word; and
- The translated text used correct grammar, punctuation, qualifiers, and modifiers, as appropriate for the target language.

Translation and adaptation of cognitive items

One of the main challenges when translating the ICCS 2016 cognitive test material was to ensure that appropriate terms and expressions in the target language(s) of each country conveyed the same meaning and that the style of the text was consistent with the international source version. When adapting and translating expressions with more contextually appropriate terms, translators needed to ensure that the meaning and difficulty of the item remained the same as the international source version. It was important that adaptation/translation of an item did not simplify or clarify the text in such a way as to provide a hint or definition of the meaning of a question. Translators also needed to ensure the consistency of adaptations and translations from item to item. For multiple-choice items, translators were instructed to pay particular attention to common text that appeared in the question stem and answer options in the source version, to ensure these were maintained in the translated national version.

Although NRCs were strongly encouraged to keep adaptations to a minimum, some adaptations were necessary in order to prevent students from facing unfamiliar contexts or vocabulary that could hinder their ability to read and understand the item. In some cases, changes to the instruments were necessary to follow national conventions, punctuation, and expressions of date and time. Where the names of fictional cities or towns were adapted, translators were advised against using real place names to prevent students' responses from being influenced by their perception and knowledge of the names. These references needed to be adapted to names in the target language of similar length, familiarity, and complexity, the aim being to convey the same meaning and maintain the style of text used in the international source version. In the specific area of civics and citizenship, particular concepts that were not common to all countries, such as specific institutions and organizations, required adaptation. For example, the term "parliament" (intended to refer to a legislative body at the national level) might be used in some unicameral systems (e.g., "Parliament of Denmark") but in bicameral systems contexts may vary, and the appropriate term would have to be selected (e.g., the “Staten-Generaal” or the parliament of the Netherlands consists of the “Tweede Kamer”, or the lower house, and the “Eeerste Kamer”, or the upper house, and each has different tasks).

The goal of such adaptations was to make the questions equally familiar to all students, while maintaining the same meaning and level of difficulty. Some terms in the text were not to be changed or adapted beyond translation. Examples include proper names of actual people and institutions. Countries were not permitted any national additions to the cognitive test.

Cognitive items for comparison over time

According to a carefully specified design, a substantial number of items (about 60 %) were carried over from the previous cycle in order to measure changes in student achievement over time. These items provide the basis for comparisons of changes in achievement from the earlier assessment for those countries that participated in both studies with the same target population. To ensure the quality of comparisons over time, these items need to be administered in exactly the same way in every cycle. For countries that previously participated in ICCS 2009, the translations of the items used in the previous assessment were compared against the ICCS 2016 translations.

If a national center determined that changes to these specific items were necessary (e.g., in order to correct a mistranslation discovered in a previous translation), the changes were carefully documented and referenced during the data analysis. Some items with changes were not included in the comparison analyses over time for that country.

For countries not participating in the comparison over time, the preparation of these specific items follows the same general procedures as the newly developed test items for the current cycle.

Translation and adaptation of the questionnaires

The translation of the questionnaires differed from the translation of the test items in that participating countries were required to adapt some terms, and to ensure that questions were appropriate for the national context and education system. Concepts and expressions that were not common to all cultures and not related to the substance of the questions needed to be changed. The terms requiring adaptation were listed in angle brackets in the international source version. For instance, <language of test> and <target grade> were adapted to the name of the actual language and grade in which the assessment was administered; for example, in Croatia, these terms would be replaced by equivalents “Croatian” and “grade 8”. Some terms related to specific aspects of students’ participation in school leadership were also designated for adaptation. For example, <school parliament> needed to be adapted to the local term that denoted any type of student representation that was democratically elected by students at the school; for example, in Estonia this would be “students’ representative body”. Items assessing levels of education used the current version of the International Standard Classification of Education (ISCED) system. Categories needed to be adapted to the national context and mapped to the ISCED classification of educational levels (http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf) for each participating country.

NRCs were provided with detailed notes on all required adaptations, which contained comprehensive descriptions of the intent of each required adaptation with the aim of clarifying the meaning of the terms used and thus enabling the translators to select the appropriate national term or expression necessary to convey the intended meaning.

Countries were permitted to add a small number of national interest questions or categories within the existing questionnaires. NRCs were instructed to place all substantive national items at the end of the international questionnaires. All national interest questions needed to be documented and approved by the ICCS 2016 international project team before they could be included in the questionnaires.

The national adaptation form

All changes, selections, and adaptations to the survey instruments were done with the goal of creating an international database containing comparable data from all participating countries. Such modifications were documented in a national adaptation form (NAF) for each language and set of instruments. The NAF was an Excel document containing the complete translation, adaptation, and verification history of each set of national instruments. It consisted of several worksheets for the cognitive test, international student questionnaire, teacher questionnaire, school questionnaire, regional student questionnaire, and additional country documentation (language information, version number, and inclusion status of international optional questions). During various stages of the instrument preparation process, sections of the form were entered and reviewed.

While translating and adapting a set of national instruments, the first version of the NAF was filled out in collaboration with the translator(s), reviewer(s), and NRC. The translator and reviewer documented the initial adaptations made to the instruments, which the NRC then reviewed and consolidated. International verifiers and NRC updated and revised the NAF after each round of international verifications.

Documenting an adaptation in the NAF required entering the proposed adaptation in the target language, an English back translation of the adaptation, recoding instructions (if applicable), and a justification for any proposed changes to adaptations. For ease of use and documentation of the different stages of verification, the NAF included designated areas for each item, respondent, and instrument. For those countries measuring comparisons in achievement over time, adaptations for existing ICCS 2009 items had to be documented in the NAF, in the same way as the other national adaptations.

The NAF was an important record of each country's final instruments, as it contained information used throughout the different stages of the translation and verification processes. The international quality observers also used the NAF after data collection to review the implementation of translation and layout verification feedback (see Chapter 7: Quality Assurance for ICCS 2016). The NAF was referenced when adding national data to the international database and during data analysis.

International verification processes

Adaptation reviews

NRCs were required to consult with ISC staff to review all proposed national adaptations. In particular, they were strongly encouraged to discuss any adaptation that might result in a serious deviation from the international instruments. National centers began completing the NAF (Version I) after reviewing the international source version of the survey instruments, and submitted the NAF to the ISC for consultation. Following its review, the ISC provided the national centers with feedback on their adaptations and, where appropriate, made alternative suggestions that better aligned to the source version. Some of the common issues that were identified during adaptation review were:

- Inconsistent adaptations used within or across test booklets or questionnaires;
- Difficulties in establishing country appropriate adaptations for ISCED levels;
- Deviations from adaptations used in previous cycle; and
- Difficulties identifying an appropriate set of subjects that were related to civic and citizenship education in the national context.

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 verification

The national translations of the instruments underwent international translation verification. The IEA managed the international translation verification process in coordination with an external translation verification company, *Today Translations* (based in London, United Kingdom).

The international translation verifiers were responsible for reviewing and documenting the quality and comparability of the national instruments to the international instruments. The required qualifications for verifiers included:

- Fluency in English;
- Mother tongue proficiency in the target language;
- Formal credentials as translators working in English;
- University-level education and (if possible) familiarity with the subject area; and
- Residency in the target country, or close contact with the country and its culture.

Verifiers were trained through web-based seminars, where they received detailed instructions for reviewing the survey instruments and registering deviations from the international source 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.

The IEA supplied verifiers with instructional materials to support their work. Each verifier received the relevant manuals and instruments, the translation guidelines, and a comprehensive set of directions, instructions and relevant examples for reviewing the national instruments and registering deviations from the international source version. The IEA also ensured all international translation verifiers received continuous training, and provided the verifiers with constructive ongoing feedback.

The instructions and training given to the verifiers emphasized the importance of maintaining the same meaning and difficulty level in the translations and adaptations as in the international source version, and ensuring that translations and adaptations were adequate and consistent within and across national instruments and languages of administration. The international translation verification process involved:

- Checking the accuracy, linguistic correctness, and comparability of the translation and adaptations of the cognitive items and questionnaires;
- Documenting any deviations between the national and international source versions, including additions, deletions, and mistranslations; and
- Suggesting an alternative translation/adaptation to improve the accuracy and comparability of the national instruments.

Verifiers provided feedback from translation verification on both the set of instruments and in the associated NAF. Verifiers were asked to correct the text of the cognitive test items and questionnaires and/or to add notes specifying errors using the "Track Changes" function in Microsoft Word. In general, translation verifiers considered the national translations/ adaptations to be well documented and of very high quality. Translation verifiers of the main study instruments also noted the great care taken in implementing their verification feedback after the field trial.

During translation verification, some of the typical errors identified by the verifiers included mistranslations, omissions/additions of text, inconsistent translations, and grammar. In the main survey, for some items the order of choices in multiple-choice items was wrong, as some countries mistakenly kept the order from the field trial. Some of the domain-specific concepts in civics and citizenship (e.g., "parliament") were a particular challenge to adapt for some countries. Any adaptations reported in the NAF were also reviewed by the verifiers, who were asked to comment on their adequacy. An important part of the verifiers' feedback was focused on the changes in the text of the items that were meant to be kept identical for the comparison over time. With the documented comments and suggestions from the verifiers, NRCs were able to revise and improve their national versions.

All comments that the verifiers deemed to be severe deviations from the source (according to the criteria) were entered into the NAF. Each deviation was allocated one of four codes that indicated to the NRCs the severity and type of deviation of the translated text from the international source version:

Code 1: Major Change or Error. Examples include incorrect translation resulting in the answer being suggested by the question; an incorrect translation which changes the meaning or difficulty of the item; incorrect order of choices in a multiple-choice item; omission of a question; incorrect order of items/questions; deviations from the text of the over time comparison items (trend items); and agreed national adaptations not being implemented in the files. Where the verifier was in doubt over the severity of an error or where the verifier was unsure how to correct a possible error, code 1? was assigned. It was also used when the verifier suggested corrections to national adaptations or trend items when the translation was incorrect.

Code 2: Minor Change or Error. Examples include purely linguistic errors and spelling errors that do not affect comprehension.

Code 3: Suggestion for Alternative. The translation may be adequate, but this code indicates the verifier suggested a different wording.

Code 4: Acceptable Adaptation. The adaptation (either required, where <text is in carets>, or non-required yet necessary) is acceptable and appropriate.

For all countries included in comparisons over time, the international verification procedure included a 'trend check' to ensure that the trend items had not been changed. This involved:

- Checking that each of the trend items for the current cycle remained identical to the trend items as they were administered in the previous cycle; and
- Documenting any differences in content.

The verifiers were instructed to record any discrepancies found in the trend items in the NAF. NRCs were instructed to carefully review all discrepancies and were instructed to discuss any proposed changes with the international project team.

Layout verification of paper-based instruments

After 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) reflecting any changes resulting from translation verification.

These files were accessed by staff at the ISC for layout verification. All layout issues identified were documented in a worksheet added to the NAF. The layout issues in each set of instruments were grouped by 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 were identified across countries. These included formatting issues (e.g., spacing, font size, margins, layout consistency across booklets), incorrect order of questions, incorrect labelling of questions, missing text, and the addition of questions not agreed upon in the 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 2016 instruments were finalized, a final version of the NAF (Version IV) was prepared and used by the IEA for data processing.

Verification of online questionnaires

For countries administering the school and teacher questionnaires online, instrument preparation comprised an additional verification step. To set up the survey online, countries were asked to set up their online questionnaires once the paper-based instruments had been verified (as described above). Countries then used the IEA Online Survey System, a process that was primarily a matter of copying and pasting text elements from the already verified paper instruments.

The *Designer* component of the IEA Online Survey System enabled users 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 that enabled users to verify and test-drive the survey exactly as if under live conditions. Once conversion was complete, the Designer also allowed users to export converted questionnaire files to the IEA in Hamburg for final verification.

To ensure that data from both administration modes were comparable, the IEA conducted a systematic check of the paper and online questionnaires. Except for a few inevitable exceptions, which were necessary because of the different administration modes and which were set down for national centers in "online adaptation notes," any deviations with regard to content and layout between paper and online instruments were reported back to the countries. In such cases, the IEA asked national centers to update their online instruments to match the paper instruments. At the final stage of the production of national online instruments, IEA staff checked the layout and structure of all online questionnaires.

IEA staff also conducted visual checks, using the same standards and procedures used for verification of the paper layout. Staff then checked the structure of the national online instruments against the structure of the international online instruments (e.g., number of categories and width of non-categorical questions). The only intended deviations they approved were those documented on the NAF. All inconsistencies that were found were listed in the NAF and reported back to the national centers for their review and revision. IEA staff then verified the revised version of the instruments. This procedure was repeated until no more inconsistences could be found. For the majority of languages, one to two rounds were needed before the final layout and structure of the online instruments were agreed.

As a last check, the IEA set all instruments online and asked the national centers to review the questionnaires one more time in an online environment. In a few cases, this check resulted in additional minor changes (such as correction of spelling errors). It was only after completion of this final check that respondents received notification that the questionnaires were ready and were given the link and login information they needed to access them.

Quality observers' review

International quality observers (IQOs) from each country were hired by the IEA to document the quality of the ICCS 2016 administration, including the survey materials.[3] An important part of the IQOs' responsibilities was to carefully review the national instruments used during the main survey data collection. The IQOs compared the final (printed) country versions of the questionnaires and test booklets against the translation verifiers' comments and suggestions to ensure that the recommendations of the translation verifiers were addressed appropriately by the national centers.

3 For more information about the ICCS quality assurance procedures, please see Chapter 7.

CHAPTER 5:

Sampling design and implementation

Sabine Weber

Overview

In an international comparative survey such as ICCS, the selection of high-quality samples is critically important. Students and teachers must be selected through the use of sound methods, to produce accurate, precise, and internationally comparable estimates. ICCS followed all requirements for sampling quality specified in the *Technical standards for IEA studies* (Martin, Rust, & Adams, 1999).

The international sample design used for ICCS was a stratified two-stage probability design, as established for the 2009 cycle, and the content of this chapter is largely based on Zuehlke (2011). During the first stage, schools were sampled with probability proportional to the size of the schools (defined by the number of students in the schools). During the second stage, one intact class of target-grade students was randomly selected for the student survey. Further, a fixed number of target-grade teachers was randomly selected in each school for the teacher survey. 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 the parts of the population not covered by or excluded from ICCS;
- The international sample design; and
- 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 that they intend to study. Survey results from a representative sample allow inferences to be drawn from 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,[4] 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" (namely 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".

4 ISCED stands for International Standard Classification of Education (UNESCO, 2011).

Teacher target population

ICCS defined the target population of teachers as follows:

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. The teacher target population in ICCS consists of 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, and have been employed at schools since the beginning of the school year.

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 (and thus were out of scope):

- Staff attending to the needs of the target-grade students but not teaching any lessons (e.g., psychological counselors, or chaplains);
- Assistant teachers and parent-helpers; and
- 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" will be used to describe "teachers of students in the target population".

Coverage and exclusions

Population coverage

The ICCS international sampling team encouraged ICCS countries to include all students and teachers covered by the target population definition in the study. 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.

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 based on information provided by the NRCs) and the weighted within-sample exclusion rate (which was estimated based on 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. Five participating countries exceeded this limit.

National centers were able to define those groups of schools that had to be excluded from the ICCS student survey according to their respective national contexts. 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 (Characteristics of national samples) of this report.

Teacher exclusions

Unlike the student survey, there were no reasons that might lead to teacher exclusions from the ICCS survey. If a teacher was part of the teacher target population, he or she was eligible to participate in the study, and consequently 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 also have 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, and only rough estimates, or no estimates at all.

Overview of exclusions

The population coverage and the exclusion rates for the student survey and the teacher survey are reported for all ICCS countries (Table 5.1).

Table 5.1: Population coverage and exclusion rates

Country	Student survey				Teacher survey
	Population coverage (%)	School-level exclusions (%)	Within-sample exclusions (%)	Overall exclusions (%)	Overall exclusions (%)*
Belgium (Flemish)	100	4.8	0.1	4.9	4.8
Bulgaria	100	1.6	0.9	2.5	10.6
Chile	100	1.1	2.4	3.5	4.1
Chinese Taipei	100	1.6	1.7	3.3	4.8
Colombia	100	0.2	0.2	0.4	1.1
Croatia	100	0.5	4.6	5.2	0.5
Denmark	100	1.7	2.7	4.4	4.6
Dominican Republic	100	1.1	0.0	1.1	3.1
Estonia	100	5.1	1.6	6.7	10.4
Finland	100	2.2	1.1	3.3	2.2
Hong Kong SAR	100	4.7	0.0	4.7	n/a
Italy	100	0.8	3.9	4.8	0.8
Korea, Republic of	100	1.7	3.0	4.7	0.6
Latvia	100	4.3	2.2	6.5	4.3
Lithuania	100	3.5	1.8	5.3	10.0
Malta	100	1.6	0.2	1.8	1.6
Mexico	100	0.9	1.1	2.0	3.3
Netherlands	100	3.0	0.9	3.9	8.6
Norway	100	1.3	4.2	5.5	4.4
Peru	100	3.0	0.0	3.1	11.1
Russian Federation	100	2.1	3.0	5.1	2.1
Slovenia	100	1.8	0.8	2.7	3.4
Sweden	100	2.2	4.3	6.4	2.2
Benchmarking participant					
North Rhine-Westphalia (Germany)	100	1.4	5.6	7.0	1.4

Notes:
Results are rounded to the nearest whole number, and thus some totals may appear inconsistent.
n/a = not applicable.
*NRC estimate.

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 participate 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 variable of interest. This means that the complex sample design of ICCS should 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 that did not exceed 5.0 score points for civic knowledge scores and that did not exceed 0.5 score points for questionnaire scales.

In the ICCS student survey, the ICCS sampling team asked 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 3000 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 several 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;
- The average class size in a country was so small that it was not possible to reach, through selection of 150 schools, the student sample size requirement of 3000 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 opted to select two classes to enable future 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 3000 tested students. Non-response, school closures, decreasing student populations or other reasons meant that some countries did not meet this requirement. The ICCS sampling team did not regard this outcome as problematic provided 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 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 few teachers were not included in the survey. ICCS did not specify a minimum achieved teacher sample size.

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

The IEA developed and provided the participating countries with specialized software called Windows® Within-school Sampling Software (WinW3S). 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

ICCS 2016 recorded the intended and achieved school sample sizes, the achieved student sample sizes and the achieved teacher sample sizes for all participating countries (Table 5.2).

Table 5.2: School, student, and teacher sample sizes

Country	Originally sampled schools (*n*)	Student survey		Teacher survey	
		Participating schools (*n*)	Participating students (*n*)	Participating schools (*n*)	Participating teachers (*n*)
Belgium (Flemish)	165	162	2931	157	2021
Bulgaria	150	147	2966	140	1549
Chile	180	178	5081	169	1452
Chinese Taipei	150	141	3953	144	2239
Colombia	150	150	5609	136	1580
Croatia	178	175	3896	176	2723
Denmark	240	184	6254	59	489
Dominican Republic	150	141	3937	128	754
Estonia	175	164	2857	49	403
Finland	185	179	3173	170	2097
Hong Kong SAR	150	91	2653	n/a	n/a
Italy	170	170	3450	170	2331
Korea, Republic of	150	93	2601	106	1368
Latvia	156	147	3224	144	1946
Lithuania	187	182	3631	183	2674
Malta	47	47	3764	47	737
Mexico	223	213	5526	210	1918
Netherlands	150	123	2812	112	1374
Norway	150	148	6271	143	2010
Peru	210	206	5166	206	2384
Russian Federation	352	352	7289	140	1743
Slovenia	150	145	2844	143	2056
Sweden	158	155	3264	135	1542
Benchmarking participant					
North Rhine-Westphalia (Germany)	174	59	1451	n/a	n/a

Note:
n/a = not applicable.

School sampling design

The IEA undertook the school sample selection for all of the ICCS countries. The IEA 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) within each stratum. The following subsections outline the school sample design for ICCS.

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 (such as geographic region, urbanization level, or 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

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 usually was the number of students in the target grade or an adjacent grade;
- Values for each of the intended stratification variables.

School sample selection

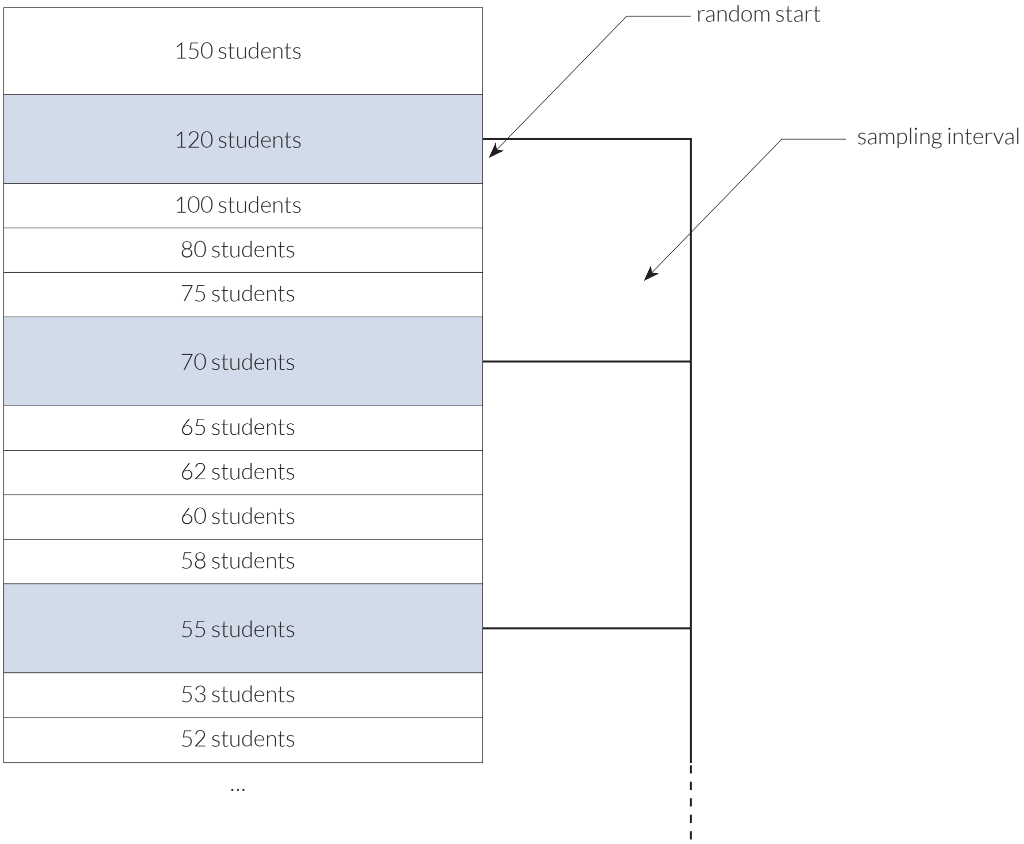
In order to select the school samples for the ICCS main survey, the sampling team used stratified PPS (probabilities proportional to size) systematic sampling. As noted earlier, this method is customary in most large-scale social surveys, and notably in most IEA surveys.

The process of selecting the school samples for each country started with the sorting of the school sampling frame. Within each explicit stratum, schools are sorted 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 the following tasks:

- Calculating a sampling interval in each explicit stratum, as process that involved dividing the total MOS in this stratum by the number of units to sample 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 equalled or exceeded the corresponding value, the team selected the corresponding unit.

Figure 5.1: Systematic PPS sampling of schools



Notes:
A box represents a school in the sampling frame. Schools in the sampling frame are sorted in descending order by size. 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 shaded blue.

The sampling team generally followed a systematic PPS sampling process within an explicit stratum (see Figure 5.1). In certain cases, the team systematically 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 42 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 (further details can be found in Appendix B).

Most ICCS countries conducted an extensive field trial of the study instruments prior to the main data-collection phase. Had a school been selected both for the field trial and for 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 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.

Within-school sampling design

Within-school sampling constituted the second stage of the sampling process in ICCS. 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 target-grade classes 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. The 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 so-called pseudo-class. This was done to avoid fluctuations in the total student sample size and to ensure efficient use of study resources.

Teacher sampling

WinW3S software 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 software to sample teachers.

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

References

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 (2011). *International standard classification of education: ISCED 2011*. Montreal, Quebec, Canada: UNESCO Institute for Statistics.

Zuehlke, O. (2011). Sampling design and implementation. In: W. Schulz, J. Ainley, & J. Fraillon, (Eds.) *ICCS 2009 technical report*. Amsterdam (pp. 59–68), The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

CHAPTER 6:

Field operation procedures and data preparation

Juliane Kobelt and Hannah Köhler

Overview

Successful administration of the ICCS 2016 assessment depended heavily on the contributions of the study's national research coordinators (NRCs) and national center staff. As for all large-scale cross-national surveys, administration of the assessment, along with the overall coordination and logistical aspects of the study, presented a set of significant challenges for each participating country.

The ICCS international study center (ISC) at the Australian Council for Educational Research (ACER) worked together with the IEA to develop internationally standardized field operation procedures, to assist the NRCs and to promote uniform instrument administration activities. The ICCS operation procedures were designed to be flexible enough to simultaneously meet the needs of individual participants and IEA's high quality survey standards.

The international project team first referred to the ICCS 2009 study procedures and those used in other previous IEA studies, such as IEA's Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS), and International Computer and Information Literacy Study (ICILS). The procedures were then tailored to suit the specific requirements of ICCS 2016.

All national centers received guidelines on survey operation procedures for each stage of the assessment. These guidelines provided advice on contacting schools, listing and sampling students, preparing materials for data collection, administering the assessment, scoring the assessment, and creating the data files. National centers were also supplied with procedures for quality control and provided feedback on survey activities, gathered via online questionnaires.

Field operation personnel

The role of the national research coordinators

One of the first steps in establishing the ICCS study for a certain country or education system[5] was the appointment of a national research coordinator (NRC). The NRC acted as the contact person for all those involved in ICCS within the country, and was the country representative at the international level.

NRCs were in charge of the overall implementation of the study, had responsibilities for the national decisions regarding ICCS and, where necessary, implemented and adapted internationally agreed-upon procedures for the national context under the guidance of the international project staff and national experts.

5 The majority of the entities that participated in ICCS were countries. Further, some subunits of countries featuring a distinct education system also participated in ICCS, for example Hong Kong, a Special Administrative Region of China. For reasons of simplicity, the text will refer to both participating countries and education systems as "countries".

The role of the school coordinators and test administrators

In order to facilitate successful administration of ICCS, the international team required the establishment of two roles within countries: the school coordinator and the test administrator. Their work involved preparing for the test administration in schools and carrying out the data collection in a standardized way.

In cooperation with school principals, national centers identified and trained school coordinators for all participating schools. The school coordinator could be a teacher or other staff member in the school. The school coordinator could also be the test administrator at the school, but was not to be a teacher of any of the sampled students. In some cases, national centers appointed external school coordinators from their own staff, for example. The coordinators' responsibilities included the following major tasks:

- Identifying eligible students and teachers belonging to the target population to allow the national center to perform within-school sampling;
- Arranging the date(s) and modalities of the test administration with the national center;
- Distributing instruments and related materials needed for test administration and making sure they were kept in a secure place and confidential at all times;
- Working with school principals, the test administrators, and the affected teachers to plan and administer the student testing;
- Ensuring that the test administrators return all testing materials after the testing session.

Test administrators were mainly responsible for administering the student test and questionnaires. They were employed either by the national center or directly by the schools. Training sessions, run by the national center centrally or by the schools, ensured that test administrators were adequately prepared to run the assessment sessions.

Field operation resources

Manuals and documentation

The ICCS survey operation procedures were sent to the NRCs in five units, each accompanied by additional materials, including more specialized manuals and software packages. The units and materials were organized and distributed chronologically according to the different stages of the study.

The five units were as follows:

- *Unit 1: Sampling Schools* defined the ICCS target populations and sampling goals, and described the procedures for the sampling of schools;
- *Unit 2: Working With Schools* guided the national centers through the process of contacting and working with schools (such as obtaining permission from relevant school authorities, establishing contact with schools, and identifying and training school coordinators);
- *Unit 3: Preparing Survey Material* described the processes involved in preparing the survey materials (such as translating and adapting assessment materials, documenting national adaptations made to the instruments, preparing for the international verification of translated materials and adaptations, preparing for the international verification of the instrument layout, and duplicating materials for the assessment);
- *Unit 4: Data Collection and Quality Monitoring Procedures* described the activities involved in working with the sampled schools and guided the NRCs through the data collection procedures, including the within-school sampling process, the tasks related to test administration, and the quality monitoring procedures;

- *Unit 5: Data Capture, Coding and Scoring* described the processes involved in using the online data collection system, and all activities related to data entry and scoring.

Accompanying manuals were as follows:

- The *School Coordinator Manual* described the role and responsibilities of the school coordinator as the main contact person within each participating school. The responsibilities included assisting the national center in the identification of classes, teachers and students, supporting the administration of the test and questionnaires, and ensuring that test materials in the school are secure and confidential at all times.
- The *Test Administrator Manual* described the role and responsibilities of the test administrator, including the distribution of the student test instruments according to the student tracking forms, the supervision of test sessions, ensuring that the timing of the test sessions was correct, and recording student participation.
- The *International and National Quality Control Monitor Manuals* provided international quality observers (IQOs) and national quality observers with information about ICCS, and described their role and responsibilities in the project. The manuals specified the timelines, actions, and procedures required to carry out the international and national quality assurance programs.
- The *Scoring Guides for Constructed-response Items* provided detailed and explicit guidance on scoring each constructed-response item.

Software

The international project team also supplied NRCs with software packages to assist with sampling and data collection:

- The *Windows® Within-school Sampling Software* (WinW3S): This enabled ICCS national centers to select students and teachers in each sampled school in agreement with sample design specifications and mandatory sampling algorithms. National centers also used the software to track school, teacher and student information, prepare the survey tracking forms, and to assign test instruments to students; the software included facilities to print labels for all the test booklets and questionnaires.
- The *IEA Online Survey System* (OSS): This software was used to transform the verified paper questionnaires to the online mode for those countries that used online data collection for teacher and school questionnaires. It enabled text passages on the paper questionnaires to be transferred to online questionnaires, while taking national adaptations to the questionnaires into account. The software also made it possible to deliver these online versions to respondents.
- The *IEA Data Management Expert* (DME): This software was used to enter information from any paper-based test booklets and questionnaires into computer data files. The DME included an international codebook, providing specific information on the variables in each of the ICCS 2016 civic knowledge test and questionnaire data files. National centers were required to adapt the international codebook structure to reflect any approved adaptations made to the national questionnaire.
- In addition to the software and manuals, the IEA conducted a field operation and scoring training seminar designed to train national center staff on all procedures and familiarize them with the supporting software, namely the IEA WinW3S, IEA Online Survey System, and IEA Data Management Expert software.

Field operation processes

Survey listing and tracking forms

ICCS relied on a series of survey listing and tracking forms to sample classes and teachers, assign booklets and questionnaires, and track the participation status of students and teachers. These forms facilitated the data collection and data verification process, and provided the information needed to compute sampling weights and evaluate the quality of the sampling process.

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

- *School tracking form:* The IEA Sampling Team provided national centers with this form, which listed the sampled schools and their replacements, information about the school identification code (ID), a school measure of size (MOS), the school name, and school contact information.
- *Class listing form:* A class listing form was created in WinW3S for each sampled school and sent to the school coordinators for completion. The school coordinators listed the eligible classes from the target grade in the participating schools and provided details about the classes, such as the class streams (if applicable), class names, and number of students. The number of teachers teaching students from the target grade was also specified on the class listing form.
- *Student listing form:* This form was created in WinW3S for each sampled class and sent to the schools for completion by the school coordinator prior to test administration. The school coordinators listed the eligible target grade students in the sampled classes of the participating schools and provided details about the students, such as the name of the student (if country regulations allowed for names to be provided to the national center), date of birth, gender and exclusion status (if applicable).
- *Student tracking form:* This form was created in WinW3S and 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 instruments to students and indicate student participation and the use of spare instruments.
- *Teacher listing form:* This form was created in WinW3S for each sampled school 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 the teachers, such as their name (if country regulations allowed for names to be provided to the national center), birth year, and gender, and indicated whether the teacher needed to be sampled with certainty (this was an option for national centers to increase the teacher sample by selecting all teachers of civic and citizenship education related subjects at the target grade).
- *Teacher tracking form:* This form was created in WinW3S and sent to the school coordinators together with the teacher questionnaires. The school coordinators used this form to indicate the completion and return status of the teacher questionnaires.

Linking students to classes and schools and teachers to schools

The international project staff established a system assigning unique hierarchical identification codes (IDs) to the sampled schools, teachers, and students (see Table 6.1), enabling them to be tracked throughout the study.

Every sampled student was assigned an eight-digit identification number unique within each country. Each number consisted of the four-digit identification 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 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.

Table 6.1: 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	CCCCTT	100101

Contacting schools and sampling classes

Once NRCs had obtained a list of the schools that had been sampled for ICCS, (for more information on all sampling procedures, please refer to Chapter 5 of this report), it was important for the ongoing success of the study that good working relationships with these schools were established. NRCs were responsible for contacting these schools and encouraging them to take part in the assessment, which 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 was provided with an *ICCS school coordinator manual*, which described their responsibilities in detail and encouraged them to contact the NRC if they had any questions.

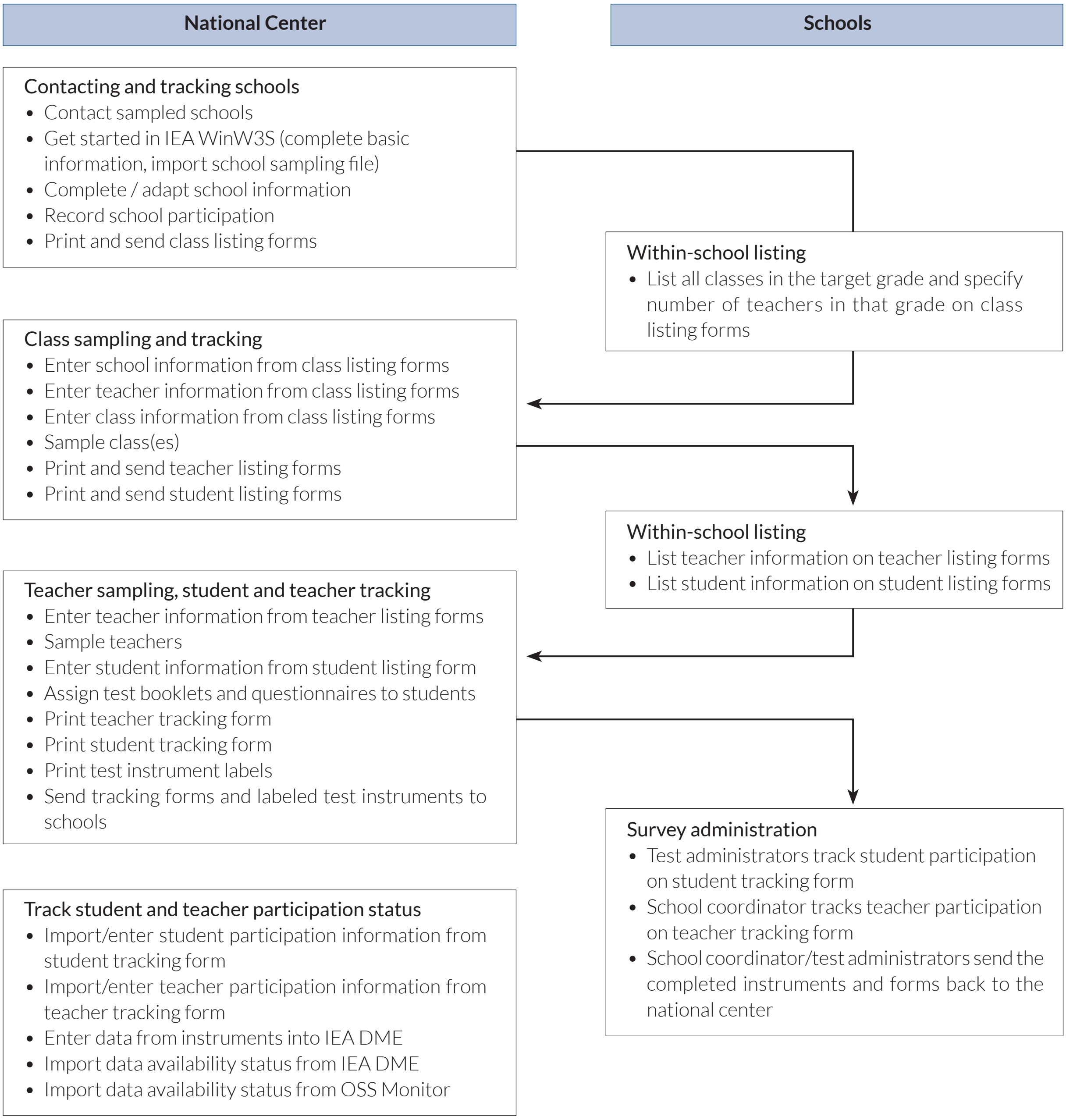
The responsibilities of the school coordinator included providing the necessary information about their school, coordinating the date, time and place for testing, obtaining parental permission (if necessary), liaising with the test administrator to coordinate the test session, distributing teacher and school questionnaires, and coordinating the completion of the student tracking forms and teacher tracking forms. School coordinators also ensured that all 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 them 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 had to be sampled, ensuring that every student in the school was in only one class (course), and 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 for all target grade students at the school. Once the class sampling was finalized, the national centers sent the *teacher listing form* and *student listing form* to the school coordinator. The school coordinator was asked to provide information for each single teacher that taught students in the target grade at the time of the assessment. Finally, school coordinators were asked to complete the *student listing form*, recording the student's name, date of birth, gender, and whether the student was excluded from testing.[6]

The national centers worked with schools to sample classes, track schools, teachers and students, and prepare for test administration (Figure 6.1).

6 There were strict criteria for the exclusion of students. Excluded students may include students with functional disabilities, students with intellectual disabilities, and non-native language speakers (as described in the ICCS 2016 school coordinator manual).

Figure 6.1: Procedures for working with schools to prepare for test administration



Notes:
WinW3S = Windows® Within-school Sampling Software, DME = Data Management Expert, OSS = Online Survey System.

Although all students enrolled in the sampled classes were part of the target population, ICCS recognized that some student exclusions were necessary because of either some functional 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 5). Countries were required to track and account for all excluded students and were cautioned that excluding more than five percent of students would require an annotation of their results in the ICCS reports. Conditions under which countries excluded students were carefully documented, because the definition of being disabled could vary from country to country.

Preparing the instruments for data collection

As outlined in Chapter 4, NRCs were required to document any national adaptations to the international instruments in national adaptation forms (NAFs) and submit them to the ISC for review and further discussion. These NAFs were provided in Microsoft Excel format and included all question texts (e.g., question stem, response options, and answer categories) and variable names.

ICCS 2016 Survey Operation Procedures Unit 3 provided countries with guidelines and explicit instructions on how to prepare the survey materials and thus produce good quality instruments. Subsequently, national centers managed the translation of the instruments from English into the language(s) used in the participating countries and later submitted them for independent verification (see Chapter 4 of this report for details).

Following the translation verification and revision of the instruments, national center staff assembled the final instruments. The ISC undertook the 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, the preparation of instruments comprised an additional verification step. The paper-based instruments had to be verified before national center staff were allowed to set up their online questionnaires using the IEA Online Survey System (OSS). The IEA undertook a final layout verification of the online instruments and national centers were asked to revise them, where necessary, prior to use (see Chapter 4 of this report for details).

Online data collection for school and teacher questionnaires

ICCS offered participating countries the option of administering the school and teacher questionnaires online instead of using paper-based questionnaires. To ensure the comparability of the data from the online mode, only those countries that had previously tested the online data collection (ODC) during the ICCS field trial were allowed to use the online option during the main survey. Sixteen countries administered school and teacher questionnaires online during the ICCS main survey: Belgium (Flemish), Bulgaria, Chile, Colombia, Croatia, Dominican Republic, Estonia, Finland, North Rhine-Westphalia (Germany), Lithuania, Malta, Norway, Peru, Russian Federation, Slovenia and Sweden.

After the school and teacher questionnaires had gone through translation and layout verification processes, they were prepared for delivery online using the IEA OSS software, as described in more detail in Chapter 4 of this report.

The online questionnaires were prepared and administered using the IEA OSS. The IEA OSS 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 information for data management. The IEA OSS allowed the consolidation of metadata into a single set of files that the ICCS national and international centers could 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 OSS comprises three distinct components. The *Designer* component was used to create, delete, disable, and edit survey components (e.g., questions and categories) and their properties. It enabled translation of all text passages in the existing national paper questionnaires and additional system texts, and it included a complete web server to verify and preview the survey exactly as if under live conditions. The *Designer* also supported the export of codebooks to IEA's generic data entry software, the IEA Data Management Expert (DME), to enable isomorphic data entry of online and paper questionnaires. The *Web* component was a compiled application that provided questionnaires in HTML format to the respondents for completion within standard internet browsers. Finally, the web-based *Monitor* component allowed national centers to audit participation in real-time. It also allowed the centers to follow up with schools when questionnaires were incomplete or not returned in a similar way to that used in the 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.

The electronic versions of the ICCS school and teacher questionnaires could only be completed via the internet. Accordingly, the design ensured that online respondents needed only an internet connection and a standard internet browser. No additional software or particular operating system was required. 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 to respondents as the default. This decision was based on the centers' and the schools' prior experience of participation in similar surveys and the ICCS field trial experience.

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 able to answer online.

The majority of countries used the online option for the school and teacher questionnaire as the default administration mode in ICCS 2016 (Table 6.2).

To ensure confidentiality and separation, 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 the administration, support was given by the national center, which, in turn, could contact the IEA if unable to solve the problem locally.

The navigational structure of the online questionnaire had to be as similar as possible to that of the paper questionnaires. Respondents could use "next" and "previous" buttons to navigate to an adjacent page, as if they were turning physical pages. In addition, a hypertext "table of contents" mirrored the experience of opening a specific page or question of a paper questionnaire. While most respondents followed the sequence of questions directly, these two features allowed respondents to skip or omit questions, just as if they were answering a self-administered paper questionnaire.

Table 6.2 Weighted percentages of online mode administration for school and teacher questionnaires

Country	School questionnaire (%)	Teacher questionnaire (%)
Belgium (Flemish)	93.5 (2.0)	95.8 (0.7)
Bulgaria	100.0 (0.0)	100.0 (0.0)
Chile	89.3 (3.4)	92.0 (2.4)
Colombia	100.0 (0.0)	100.0 (0.0)
Croatia	100.0 (0.0)	100.0 (0.0)
Dominican Republic	93.7 (3.1)	93.8 (3.0)
Estonia	100.0 (0.0)	100.0 (0.0)
Finland	100.0 (0.0)	99.0 (0.2)
Lithuania	100.0 (0.0)	100.0 (0.0)
Malta	100.0 (0.0)	97.3 (0.4)
North Rhine-Westphalia (Germany)	97.3 (2.7)	– *
Norway	100.0 (0.0)	100.0 (0.0)
Peru	23.6 (2.9)	28.0 (2.9)
Russian Federation	100.0 (0.0)	100.0 (0.0)
Slovenia	100.0 (0.0)	99.7 (0.2)
Sweden	100.0 (0.0)	100.0 (0.0)

Notes:
Standard deviations provided in parentheses.
* Concerns about the extremely low response rates (less than 10%) for the teacher surveys in North Rhine-Westphalia (Germany) led to a decision not to include the corresponding data in the international database.

To further ensure the similarity of the two sets of questionnaires, responses to the online questionnaires were not made mandatory, evaluated, or enforced in detail (e.g., using hard validations). Instead, some questions used soft validation; for example, respondents were asked to give numerical responses to questions that had a minimum and maximum value, such as the total number of students enrolled in a school. In some instances, respondents' answers to this type of question led to the response being updated according to the individual respondent's entries even if that response was outside the minimum or maximum value, but with the caveat that the response still needed to be within the specified width.

Certain differences in the representation of the two modes remained, however. To reduce response burden and complexity, the online survey automatically skipped questions not applicable to the respondent, in contrast to the paper questionnaire, which instructed respondents to proceed to the next applicable question. Rather than presenting multiple questions per page, the online questionnaire proceeded question by question.

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. Answers to online questionnaires were already in electronic format, and responses were stored on servers maintained by the IEA, thus there was no need for separate data entry.

Online data collection for survey activities questionnaires

In order to collect feedback about survey operations from NRCs, the international project team set up an online survey activities questionnaire (SAQ). The questionnaire was prepared and administered using the IEA Online Survey System.

Because the SAQ, unlike the other ICCS questionnaires, did not require any national adaptations and were completed in English, they were well suited for an online data collection.

The purpose of the SAQ was to gather opinions and information about the strength and weaknesses of the ICCS assessment materials (e.g., instruments, manuals, scoring guides, and software) and countries' experiences with the ICCS survey operation procedures. NRCs were asked to complete these questionnaires with assistance of their national data managers and the rest of the national center staff. The information was used to evaluate the survey operations and will also be used to improve the quality of the survey activities and materials for future ICCS cycles (see chapter 7 for more information on the answers to the SAQ).

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

Administering the ICCS assessment

The distribution of the printed materials to the schools required careful organization and planning by national centers. Using labels and the student tracking form produced by WinW3S, each sampled student was 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 also was assigned an international student questionnaire that was labeled in a way that linked it to the corresponding test booklet. Depending on the country's participation in one of the regional student questionnaires, 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 for each teacher listed on the teacher tracking form, and a school questionnaire was sent for the school principal.

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

According to procedures described in the test administrator manual, national centers assigned a test administrator to each sampled class; their role was to administer the test along with the international 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 undertook the test administrator role. The test administrator was responsible for distributing materials to the appropriate students, leading students through the assessment, and an accurate timing of the sessions. Following the test, they also administered the international student questionnaire and a regional instrument (where applicable).

The administration of the ICCS instruments consisted of either two or three parts. The first part concerned the civic knowledge test booklets. This was followed by the completion of the international student questionnaire. Countries in Europe and Latin America administered either the European student questionnaire or the Latin American student questionnaire afterwards. The time allotted for each of these sections was standardized across countries. To complete each part of the cognitive test, target grade students were allowed 45 minutes. If a student completed the test before the allotted time was over, he or she was allowed to review his or her answers or read quietly, but was not allowed to leave the testing room. To complete the international student questionnaire, students were given at least 40 minutes and were allowed to continue if extra time was necessary. In European and Latin American countries additional time was allocated to complete the regional student questionnaire. Each part of the student assessment was allocated an approximate time, and test administrators were required to document the starting and ending time of each section on the test administration form (Table 6.3).

Table 6.3 Time allowed for administering the ICCS student instruments

Instrument	Length of time allowed for test
Student test booklet	45 minutes
International student questionnaire	40+ minutes
European student questionnaire (where applicable)	20+ minutes
Latin American student questionnaire (where applicable)	15+ 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 to calculate the participation rate. If this was below 90 percent in any class, it was the school coordinator's responsibility 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.

Scoring the ICCS assessment and checking scorer reliability

Scoring the assessment

The success of assessments containing constructed-response questions depends on the degree to which student responses are scored reliably. Nine of the ICCS assessment items were constructed-response items (five of these items were newly developed for ICCS 2016 and four were secure trend items from ICCS 2009), and it was critical to the quality of the ICCS results that they were scored in a reliable manner. Reliability 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.

During the scoring training, which was conducted at the international level, national center staff members were trained how to score the constructed-response items in the ICCS assessment. Scoring training took place 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 for the new ICCS 2016 items. These guides were revised on the basis of this feedback. The training conducted before the main survey enabled national center staff to give additional feedback on the scoring guides for the five new ICCS 2016 items in light of their experiences of scoring the items in the field trial, and the guides were revised accordingly. The scoring guides for the four ICCS 2009 trend items were not revised and were identical to those used in ICCS 2009. Further detail of the development and revision of the *ICCS Main Survey Scoring Guide for Open-ended Response Items* is provided in Chapter 2.

Scorer training for the ICCS main survey was undertaken using a sample set of student responses created for use in the training, and taken or adapted from those collected in the ICCS 2016 field trial. As part of the field trial, country representatives had the opportunity to send scorer queries to the ISC. Countries submitted their queries in English (with non-English language student responses translated into English) and a selection of student responses submitted as part of the country query process was included in the set of responses used for the main survey scorer training.

The example responses used during the 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 borderline between 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.

Following the training, 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 to the constructed-response items. In some cases, national centers created their own example responses from student responses collected in their country.

To prepare for this task, the ISC provided national centers with suggestions on how to organize staff, materials, procedures, and the scoring process. National centers were encouraged to hire scorers who were attentive to detail and familiar with education, particularly those with 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 civic knowledge items.

Documenting scoring reliability

Documenting the reliability of the scoring process within countries was a highly important aspect of monitoring and maintaining the quality of the ICCS scored data. Scoring reliability within each country was established using two different scorers who independently scored a random sample of approximately 67 booklets per booklet type (booklets 1-8). The booklets to be doubled scored were selected by the WinW3S software. Reliability booklets were assigned so that about 200 responses per item were double scored in each participating country.

The degree of agreement between the scores, assigned by the two scorers, is a measure of the reliability of the scoring process. Items with low inter-rater reliability within a given country were not used in the estimation of student achievement for that country. Chapter 10 outlines the adjudication process relating to inter-item reliability of items.

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 and the two reliability scorers were unaware of each other's scores.

Field trial procedures

The ICCS field trial was a smaller administration of the ICCS assessment; on average, approximately 1000 students were tested in each participating country.

The field trial was crucial to the development of the ICCS assessment instruments, particularly the civic knowledge test. Items were tried out in the field trial in order to investigate the psychometric characteristics of the test items and make well-informed decisions about further use. The field trial also served the purpose of testing the ICCS survey operation 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 also allowed the NRCs and their staff to become acquainted with the activities, refine their national operations, and provide feedback that was used to improve the procedures for the main data collection. The field trial resulted in some small modifications to survey operation procedures and contributed significantly to the successful execution of ICCS. In almost all participating countries, the international field trial was conducted during October–November 2014.

CHAPTER 7:

Quality assurance procedures for the ICCS data collection

Gabriela Nausica Noveanu, Juliane Kobelt and Hannah Köhler

Considerable effort has been made to develop standardized materials and survey operations procedures that ensure the ICCS 2016 data meets the highest standards. Quality assurance encompassed internal mechanisms built into each stage of the data collection process.

The international quality assurance program

Overview

In order to document data collection activities and verify that the standardized ICCS procedures were followed, the IEA developed the quality assurance program, which was an integral part of the study both nationally and internationally. Two independent quality assurance programs were implemented in each participating country; these were designed to offer evidence of the procedures employed in data collection in order to advocate for data comparability, and included an international quality assurance program, conducted by the IEA, and a national quality assurance program, managed by each national center and based on the guidelines and manuals provided internationally.

This section provides an overview of the international quality assurance program and the data collected by the program. The international quality assurance program was implemented by independent international quality observers (IQOs) appointed by the IEA.

The major task of the IQOs was to conduct on-site visits during the main survey data collection. In each country, the IQO and the IQO's assistant(s) visited a sample of 10% of the participating schools during the ICCS administration. When there was a benchmarking participant from the same country, and only one centrally organized national center was responsible for all aspects of data collection, five additional school visits were required for the benchmarking entity.

For each school visit, IQOs observed the administration sessions and recorded their observations, noting any deviations from the standardized administration script, timing, and procedures. In addition, IQOs interviewed the school coordinators about their experiences coordinating the ICCS 2016 assessment. IQOs also checked whether the suggestions made by the international translation and layout verifiers had been integrated into the final assessment instruments, as documented in the national adaptation forms.

Prior to beginning their assignments, the IQOs were mandated to attend a training session conducted by the IEA. There were two training sessions, one for Southern Hemisphere countries and one for Northern Hemisphere countries. During the training, IQOs were introduced to the ICCS operation procedures and the design of the ICCS 2016 test booklets and context questionnaires, and familiarized with the tasks associated with their role. IQOs were also supplied with a manual detailing all their responsibilities and the necessary materials for completing the tasks.

In most participating countries and benchmarking entities, the national research coordinators (NRCs) assisted the IEA in nominating the IQO. There were also cases where the IEA recruited IQOs who had served the same role in previous IEA studies. An important aspect of the international quality assurance program is the independence of the IQOs from the national centers. The nominated person cannot be a member of the national center, a family member or personal friend of the NRC. Often, this person was a school inspector, a retired school teacher, or a researcher. The IQO was required to be fluent in both English and the language(s) spoken in the country. When necessary, the IQOs were asked to recruit assistants in order to effectively cover the territory or the schedule of ICCS administration. For ICCS 2016, a total of 24 IQOs were trained across the 23 participating countries and one benchmarking participant. In turn, the IQOs trained 69 assistants. Altogether, observers observed 371 administration sessions. The results of the ICCS 2016 IQO observations are reported in the following sections.

Findings from the ICCS 2016 administration

The main responsibilities for each IQO involved were:

- Visiting the national center for the selection of schools to be observed and collecting the required information and documentation;
- Observing selected administration sessions, conducting interviews with the school coordinators and reporting on specific forms; and
- Reviewing the assessment instruments to ensure that the results of the international translation and layout verifications had been properly addressed.

The IQOs were required to:

- Collect from the national center the complete set of ICCS 2016 national instruments and manuals, the national adaptation form(s), and tracking forms for the selected schools;
- Select, in consultation with the NRC and according to specified guidelines, the schools where the administration sessions are to be observed;
- Contact the school coordinator and test administrator of each selected school to organize the visit and arrange the interview;
- Observe the administration sessions for their level of adherence to the administration guidelines, documenting each session's activities on the 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 record the responses on the classroom observation record;
- Review the national assessment instruments and use of the international translation and layout verification feedback, and document whether the verifiers' comments were addressed appropriately; and
- Submit all collected national materials and completed observation records to IEA.

The IEA received documentation of the international quality assurance program from all 24 countries and benchmarking participants in ICCS 2016.

Observing the ICCS administration sessions

The observation records were organized into four sections:

- Section A: Preliminary activities;
- Section B: Administration process;
- Section C: Summary observations; and
- Section D: Interview with the school coordinator.

In section A of the classroom observation record, IQOs registered their observations of the condition of the survey materials. The second section of the form addressed activities that took place during the actual administration sessions of the student cognitive test, international questionnaire and regional questionnaire (if applicable). The cognitive test was administered in one session of 45 minutes with a break of 5 to 10 minutes before the international questionnaire administration. Typically, a break of 10-15 minutes followed for countries that administered the regional questionnaire.

During test administration, IQOs were asked to observe the activities of the test administrator, specifically the following:

- Distributing, collecting, and securing the test booklets and questionnaires;
- Following the ICCS 2016 administration script; and
- Making time announcements during the administration sessions.

IQOs registered their observations of the condition of the survey materials, the test administrator's level of preparation, and the suitability of the room where the administration took place.

In general, the reported information confirmed the very good quality of the preparations for ICCS administration (Table 7.1). The survey materials were usually stored in a secure location at the school or kept with the test administrator. There have been reports of very few deviations: in some cases, materials were not sealed properly or the observer could not verify whether they were stored properly as the administration started earlier than scheduled. According to the observers' reports, four percent of the test administrators did not check the assessment materials before the assessment started, and two percent of the observers could not judge about the test administrators' actions in this regard.

There were no recorded deviations from the within-sampling procedure. In general, discrepancies between the student identification information on the instruments and the student tracking form were attributed to new students who had joined the class after the completion of the listing forms or to typographical errors in the lists that were then corrected. In two of the visited classrooms, the student tracking form was reported as not being updated.

IQOs judged most test administrators as being familiar with the administration scripts. In general, observers noticed no procedural deviations in ICCS assessment preparations that they deemed severe enough to jeopardize the integrity of the administration. Nearly all of the observed administration sessions took place under favorable room conditions that were suitable for students to work without distraction (92%). The most commonly reported inconveniences were: insufficient workspace for the number of students sitting the test and noise from other students during school break times.

Table 7.1: IQOs' observations on ICCS 2016 administration preparation

Question	Response category (%)		
	Yes	No	Not answered
Were the assessment materials safely stored and securely sealed?	97	3	0
Did the test administrator verify adequate supplies of the test booklets and questionnaires prior to the students' arrival?	94	6	0
Does the student identification information on the test booklets and student questionnaires correspond with the student tracking form?	98	2	0
Did the test administrator familiarize himself or herself with the script prior to the administration of the assessment instruments?	97	3	0
Were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	92	8	0
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	96	4	0

Notes:
Percentages derived from responses to a total of 371 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

Administration of the student cognitive test

IQOs reported that the assessments were conducted in accordance with the international procedures, in particular on booklet distribution and adherence to time limits (Table 7.2). In very rare instances, IQOs observed and carefully documented significant deviations from the administration procedures. In one classroom, it was noted that the survey instruments were mistakenly given in a different order than prescribed.

In about one-third of sessions, the time for test administration was not equal to the time allowed. In about 82 percent of the observed deviations, students completed the test before the allotted time had elapsed. Consequently, for the observed sessions, the average test time was 44 minutes (one minute under the allocated time). The test administrators observed students working faster than expected, so remaining time announcements were made before the planned 10 minutes warning to inform the students that they still had ample time to complete their work or to persuade them to wait until the end of the session. Eighty-three percent of all IQO records showed that the break between test and questionnaire administration sessions did not exceed 10 minutes. In almost 14 percent of cases, there was no significant break between sessions, as only one or two minutes were recorded as having elapsed between them.

In accordance with the ICCS procedures, at the end of the testing session, test administrators were asked to collect and secure the test booklets. The IQOs reported that this occurred in 97 percent of sessions. The majority of the reported exceptions were because, in one country, according to the policy established at national level, the questionnaires were delivered with the test booklet in one envelope, so the students retained their test booklets until they completed the questionnaires.

Table 7.2: IQOs' observations on test administration sessions

Question	Response category (%)		
	Yes	No	Not answered
Does the test administrator distribute the test booklets according to the booklet assignment on the student tracking form and booklet labels?	99	1	0
Does the test administrator announce, "you have 10 minutes left" prior to the end of testing session?	94	6	0
Are there any other "time remaining" announcements made during the testing session?	20	80	0
Are the test booklets collected and secured at the end of testing session?	97	3	0

Notes:
Percentages derived from responses to a total of 371 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

One of the most important methods for standardizing the assessment administration was to have all test administrators follow the script in the test administrator manual. IQOs reported that in more than three-quarters of the observations, the test administrators exactly followed the script when preparing the students, distributing the materials, and giving directions and examples (Table 7.3). When test administrators deviated from the script, nearly all modifications were reported as "minor" and characterized most frequently as additions.

Table 7.3: IQOs' evaluation of the test administrators' adherence to the test administration script

Question	Response category (%)			
	Yes	No	Not answered	
Did the test administrator follow the test administration script in the *ICCS 2016 Test Administrator Manual*?	78	22	0	
Question	**Yes**	**No**	**Not applicable**	**Not answered**
If the test administrator made changes to the script, how would you describe them?				
• Additions	12	15	73	0
• Revisions	9	19	72	0
• Deletions	6	19	75	0
Question	**Yes**	**No**	**Not answered**	
Did the test administrator address students' questions appropriately?	98	2	0	

Notes:
Percentages derived from responses to a total of 371 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

IQOs observed students' compliance with instructions and overall cooperation during the test administration (Table 7.4). According to the IQOs' observations, in almost all sessions, students complied very well or fairly well with the instruction to stop work, the test administrator making sure their booklets were closed. In addition, IQOs described students as being extremely orderly and cooperative during two-thirds of the administration sessions.

Table 7.4: IQOs' evaluation of student cooperation during test administration

Question	Response category (%)			
	Very well	Fairly well	Not applicable	Not answered
When the test administrator ends testing session, how well do the students comply with the instruction to stop work?	90	9	1	0

Question	Extremely	Moderately	Somewhat	Hardly	Not answered
To what extent would you describe the students as orderly and cooperative?	61	30	8	2	1

Notes:
Percentages derived from responses to a total of 371 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

Administration of the student questionnaires

IQOs reported that the international student questionnaires were distributed according to the student tracking forms and questionnaire labels. The deviation from the procedure occurred in one country because the questionnaires were distributed together with the tests at the beginning of the administration. In most cases (77%), test administrators followed the student questionnaire administration script exactly. The changes, additions, revisions or deletions were reported almost equally (Table 7.5).

The international student questionnaire was administered in 40 minutes in one third of the observed sessions. As instructed in the *Test Administrator Manual*, students received extra time to complete the questionnaire (19% of cases). The allotted extra time varied from few minutes to a maximum of 20 minutes. The administration time was 42 minutes, in average, for the observed sessions.

The IQOs also provided observations on the administration of the two regional instruments for ICCS 2016 (Table 7.6).

Regional student questionnaires were distributed according to the student tracking forms and questionnaire labels in 75 percent of cases. Deviations from the procedure were reported in two countries: in one country, both questionnaires were combined in one booklet and in the other country all instruments were distributed in the beginning of the ICCS administration. IQOs judged the test administrators to have followed the administration script exactly in 69 percent of the sessions. Changes were generally considered minor, and were most commonly characterized as deletions.

In about half of the observed sessions, the observers reported that 20 minutes were needed to fill in the European student questionnaire and 15 minutes to fill in the Latin American student questionnaire. About 25 percent of students finished earlier. On average, for the completion of European student questionnaire, 20 minutes were needed, while for the Latin American student questionnaire, 14 minutes were needed.

Table 7.5: IQOs' evaluation of the international student questionnaire administration

Question	Response category (%)			
	Yes	No	Not answered	
Does the test administrator distribute questionnaires according to the student tracking form and questionnaire labels?	98	2	0	
Did the test administrator follow the questionnaire administration script in the *ICCS 2016 Test Administrator Manual*?	77	23	0	
Question	**Yes**	**No**	**Not applicable**	**Not answered**
If the test administrator made changes to the script, how would you describe them?				
• Additions	12	16	72	0
• Revisions	10	21	69	0
• Deletions	9	18	73	0

Notes:
Percentages derived from responses to a total of 371 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

Table 7.6: IQOs' evaluation of the regional student questionnaire administration

Question	Yes	No	Not applicable	Not answered
Does the test administrator distribute questionnaires according to the student tracking form and questionnaire labels?	75	8	17	0
Did the test administrator follow the regional questionnaire administration script in the *ICCS 2016 Test Administrator Manual*?	69	14	17	0
If the test administrator made changes to the script, how would you describe them?				
• Additions	5	14	81	0
• Revisions	4	13	83	0
• Deletions	10	9	82	0

Notes:
For the European student questionnaire there were 15 participating countries, and for the Latin American student questionnaire there were five participating countries. Percentages derived from responses to a total of 306 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

Summary observations of the ICCS 2016 administration

General impressions of how the administration was conducted, how well the test administrator monitored students, and any unusual circumstances that arose during the session were included in Section C of the classroom observation record.

Overall, IQOs reported that the quality of ICCS administration was "excellent" (54%), "very good" (30%), "good" (12%) or "fair" (3%); only one percent of sessions were recorded as "poor". High quality instruments were delivered to schools; there were only a few reports of defective booklets, and these were properly replaced. In most of the attended sessions, no problems were observed around taking the test; two percent of students (in general, one student per classroom) refused to take the test citing parental permission. Late students were admitted to the test rooms, but this occurred before the beginning of administration. In 13 percent of the observed sessions, students left the room for an "emergency" (usually a bathroom visit). In such cases, test administrators were instructed to collect the student's test booklet and return it only when the student reentered the assessment room. In some sessions, the booklet was left closed on the student's desk until the student returned to class. However, in some cases, students had already completed the test and, therefore, it was not necessary to give back the test booklets when they returned to the classroom. In four percent of the sessions, a different room was provided for students requiring special accommodation where all questions or some of them were read aloud. One student with a visual impairment received a test printed specially for him, while one with a hearing impairment was provided with a laptop that gave him/her instructions. There were reports of students accessing belongings during the test sessions. IQOs reported that, in some cases, mobile phones or tablets were left on tables, and that students could listen to music after they completed the test or questionnaire(s); some students also read books after finishing their test. In some instances, IQOs noted that the test administrators were not walking around the room but that they were still able to monitor students from the front of the room. In seven percent of cases, IQOs reported evidence of students attempting to cheat. However, in some of these instances, IQOs observed that students tried to communicate between themselves until they found out that the tests were different. It was also mentioned that test administrators intervened when necessary (Table 7.7).

Interview with the school coordinator

The IQOs conducted interviews with the school coordinators according to the guidelines included in the Section D of the classroom observation record. The purpose of this activity was to solicit an evaluation about the ICCS administration from the responsible persons, to collect suggestions for improvement, and to obtain additional background information on survey-related activities such as:

- Shipment of assessment materials;
- Arrangements for ICCS administration;
- Responsiveness of the national center to queries;
- Necessity for make-up sessions; and
- Organization of classes in the school (to validate within-school sampling procedures).

A large majority of school coordinators expressed a favorable impression of the ICCS survey, and 87 percent reported that the ICCS 2016 administration in their school went very well overall (Table 7.8). In general, when school coordinators reported problems these included: students' absenteeism, parents' refusals, tiredness, or waiting times for those students who finished earlier. School coordinators reported that 68 percent of staff in the school demonstrated positive attitudes towards the administration, and only one percent exhibited negative attitudes. There were complaints in some schools that ended with a refusal to complete the questionnaires; some of the teachers considered that questionnaires were not anonymous or the subject under evaluation was

Table 7.7: IQOs' general observations on the ICCS 2016 administration

Question	Response category (%)			
	Yes	No	Not answered	
Were any defective test booklets and/or questionnaires detected and replaced before the administration session began?	2	98	0	
Were any defective test booklets and/or questionnaires detected and replaced after the administration session began?	1	99	0	
Question	**Yes**	**No**	**Not applicable**	**Not answered**
If any defective test booklets and/or questionnaires were replaced, did the test administrator replace them appropriately?	6	2	91	0
Did any students refuse to take the test either prior to the testing or during the testing?	2	98	0	
Question	**Yes**	**No**	**Not applicable**	**Not answered**
If a student refused, did the test administrator accurately follow the instructions for excusing the student?	7	0.5	93	0
Were any late students admitted to the room?	17	83	0	
Did any students leave the room for an "emergency" during the testing session?	13	87	0	
Question	**Yes**	**No**	**Not applicable**	**Not answered**
If a student left the room for an "emergency", did the test administrator address the situation appropriately (collect the test booklet, and if readmitted, return the test booklet)?	16	2	82	0
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, dyslexia)?	4	96	0	
Did students store away everything, including all electronic devices, such as cell phones, portable computers, and photo or video cameras, having only a pen or a pencil and the test booklet for the duration of the test administration?	87	13	0	
During the administration sessions, did the test administrator walk around the room to be sure students were working on the test or completing the questionnaires and/or behaving properly?	97	3	0	
Did you see any evidence of students attempting to cheat on the test (e.g., by copying from a neighbor)?	7	93	0	

Notes:
Percentages derived from responses to a total of 371 ICCS administration sessions. Percentages may not add to 100%, due to rounding.

not relevant for them or that the administration time consumed lesson time. School coordinators were also asked about the quality of the school coordinator manual. Ninety-one percent were satisfied with the manual, while eight percent of them said that the manual needed improvement because it was voluminous and some information was repetitive.

Table 7.8: IQO reported interview responses of the school coordinators to questions about their perceptions of the ICCS 2016 administration

Question	Response category (%)			
	Very well, no problems	Satisfactorily, few problems	Unsatisfactorily, many problems	Not answered
Overall, how would you say the assessment administration went?	87	12	1	0.5
Question	**Positive**	**Neutral**	**Negative**	**Not answered**
Overall, how would you rate the attitude of the other school staff members towards the ICCS 2016 assessment?	68	31	1	0.5
Question	**Yes**		**No**	**Not answered**
Overall, do you feel the *ICCS 2016 School Coordinator Manual* worked well or does it need improvement?	91		8	1

Notes:
Percentages derived from responses to a total of 369 interviews. Percentages may not add to 100%, due to rounding.

There were only a small number of cases where components were missing from the shipments of assessment materials (Table 7.9). The time available for checking the materials ranged from weeks to days. In most cases where the school coordinator reported that there was no time to check the materials, this happened because the external test administrators brought the ICCS materials to the school on the administration day. In two countries, the school coordinator manuals were reported as missing. They were replaced by the national centers with an information sheet sent to the schools. Letters that were to be delivered to the parents were reported as missing materials in another country. Before the administration began, for all 371 sessions, a total of three defective booklets were found and replaced. Teacher questionnaire labels were incorrect in one school, but these were changed manually.

In more than half of the cases, school coordinators indicated that principals, civic and citizenship education teachers or classroom teachers gave students special instructions, motivational talks, or incentives prior to administration. In one country, students received a leaflet and watched a film provided by the NRC. Sixteen percent of school coordinators anticipated needing a make-up session, and almost 75 percent of them intended to conduct one. IQOs were also asked to verify that the class lists really included all classes because class sampling required a complete list of all classes in the school at the target grade. The majority of the school coordinators confirmed that the complete list of classes had been well documented. There were some schools where IQOs reported either one missing class or one extra class included in the list. In one country, school coordinators explained that new classes of refugees were created after the class lists were sent to the national center. Recent changes in school enrollment during the school year or students with severe learning disabilities were considered as grounds for missing students in the initial lists sent to the national centers. In two percent of cases, students were enrolled in more than one class in the school. School principals reported some reasons as follows: students with special educational needs or refugees are taking classes with different groups in the school.

Table 7.9: IQO reported interview responses of the school coordinators to questions evaluating the ICCS 2016 administration

Question	Response category (%)			
	Yes	No	Not applicable	Not answered
Prior to the testing day, did you have time to check your shipment of materials from the national study center?	80	16	0.5	4
Did you receive the correct shipment of the materials as listed in your school coordinator manual and according to the tracking forms?	88	8	0.5	4
If no, did the national center provide the missing materials in time for the testing?	4	1	0.5	94
Was the national center responsive to your questions or concerns?	94	4	0.5	1
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the ICCS 2016 assessment?	54	46	0.5	0
Was the teacher questionnaire administered online?	63	37	0.5	0
If the teacher questionnaire was administered online, did the teachers encounter any problems?	3	52	0.5	44
Was the school questionnaire administered online?	62	38	0.5	0
If the school questionnaire was administered online, did the person completing it encounter any problems?	1	51	0.5	48
Do you anticipate that a makeup session will be required at your school?	16	83	0.5	0
If yes, do you intend to conduct one?	15	5	0.5	79
Is this a complete list of the classes in this grade in this school?	91	9	0.5	0
To the best of your knowledge, are there any students in this grade level who are not in any of these classes?	10	89	0.5	0
To the best of your knowledge, are there any students in this grade level in more than one of these classes?	98	2	0.5	0
If there was another international assessment, would you be willing to serve as a school coordinator?	84	14	0.5	1

Notes:
Percentages derived from responses to a total of 369 interviews. Percentages may not add to 100%, due to rounding.

As a reflection of the planning and implementation of ICCS 2016, 85 percent of interviewed people said they would be willing to serve as a school coordinator in future international assessments with remarks that the experience was enjoyable and positive overall. Finally, it is notable that the response rate for the classroom observation records, section D, was close to 100 percent for all questions. The only exception recorded was because two school coordinators did not take part in the interviews.

Findings from the survey activities questionnaire

The survey activities questionnaire gathered information from the national research coordinators about whether the implementation of the ICCS procedures accorded with the standards outlined in the survey operations procedures manuals, and solicited feedback on the strengths and weaknesses of the approaches and materials used. It covered topics including 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. One major purpose of the survey was to gather information to improve the quality of future IEA studies.

Data were collected online from each NRC personally, with data manager and/or other national center staff assisting this process where necessary. The results presented in this section, reflect the quality of the ICCS procedures and materials in all 24 participating countries.

Sampling

The first part of the survey activities questionnaire collected information on the sampling procedures and manuals (Table 7.10), and results indicated that the sampling process worked very well overall. In most countries, the NRCs reported no difficulties in 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's sampling team. All countries indicated that the *Sampling Manual* sufficiently described the relevant processes and procedures. The Within-School Sampling Software (WinW3S), provided by the IEA to select classes and teachers, was used by all countries who participated in the main study.

Three NRCs reported encountering organizational constraints that required deviations from the standard ICCS within-school sampling design. In one country, within-school sampling was unnecessary because the ICCS sample included the entire target population. Other reported deviations from standard sampling protocol included administering the assessment to Grade 9 students at the beginning of the school year (instead of the end of the Grade 8 school year) and selecting two classes from some schools. In each of these cases, the deviation was documented and the sampling expert was consulted to ensure that the altered design met all sampling requirements. In about half of the countries, data protection laws meant that numbers or other codes (rather than names) were used on student and/or teacher lists.

Table 7.10: NRC responses to questions related to sampling

Question	NRC responses (*n*)			
	Very difficult	Somewhat difficult	Not difficult at all	Not answered
How difficult was it to...				
• Adapt the international sampling design to your national specifications?	0	1	22	1
• Compile a list of all eligible schools?	0	3	19	2
Question	**Yes**		**No**	**Not answered**
Did the sampling manual (SOP 1) sufficiently describe the following procedures?				
• Defining and identifying the target populations of the survey	23		0	1
• Developing national sampling plans (exclusions, stratification, sample sizes)	23		0	1
• Creating a sampling frame	22		0	2
Were there any conditions or organizational constraints that required deviations from the basic ICCS 2016 within-school sampling design (of sampling classes and teachers)?	3		20	1
Did you use other means instead of names to identify students and/or teachers on the forms and labels?	11		13	0
Did you use the Within-School Sampling Software (WinW3S), provided by the IEA, to sample classes and teachers?	24		0	0

Contacting schools and recruiting school coordinators

The NRCs were also questioned about contacting schools for participation and recruiting school coordinators (see Table 7.11). Generally, the 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 ministries. A number of NRCs reported difficulties in convincing the selected schools to participate. Common reasons cited were logistical issues (timing, and availability of students and staff), concerns about overburdening students and teachers, and most importantly the sensitive nature of civics 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 more than one-third of the countries, formal training sessions were held.

Table 7.11 NRC responses to questions about contacting schools and recruiting school coordinators

Question	NRC responses (*n*)		
Which materials did you use to request school participation?			
• Letters based on the examples provided in Appendix A of the Survey Operations Procedures Unit 2 (MS)	6		
• Letters based on examples from other national projects	10		
• Other	8		
Question	**Yes**	**No**	**Not answered**
Did you use incentives to convince schools to participate (for example, vouchers, money, books, lotteries, school-specific feedback)?	8	13	3
Question	**NRC responses (*n*)**		
How did you train the school coordinators? (Check all that apply)			
• Formal face-to-face training sessions	9		
• Through telephone, email or online meetings	18		
• Written instructions	23		
• Other	1		

Adapting and translating the ICCS assessment materials

As part of the survey activities questionnaire, NRCs provided information about the process of adapting and translating the ICCS student cognitive test and questionnaires addressed to students, teachers and principals into national languages. Most NRCs identified national center staff as individuals responsible for adapting and/or translating the assessment instruments; in half of all participating countries, outside specialists were responsible for translation, often working in conjunction with the national center staff. The process of translation and adaptation was generally not considered to be difficult (Table 7.12). Some countries reported general difficulties in adapting, and particular issues with the teacher questionnaire or ISCED levels. Adaptation and translation of questions from the previous cycle was considered challenging, as some terminology had already become outdated. When asked about their experiences with the external adaptation negotiation and translation verification processes, organized by the ISC and IEA, respectively, a small number of NRCs indicated that they had experienced difficulties, mainly with respect to reaching agreement on the adaptation of certain specialist concepts, and in harmonizing ICCS 2009 questions with ICCS 2016 questions. Twenty-two of the countries used the international version as the source for the adaptation and translation process. Four countries adapted the instruments provided by another national center to their national context. The countries participating in the 2009 cycle also used the national versions of the instruments used in 2009, as the items designated for comparison across cycles had to be administered in the same way in both cycles. In two of the countries, the translation used in the previous cycle was considered inappropriate in relation to the actual evolution of the language. Almost all countries reported that they did not encounter any problems in translating or adapting the *School Coordinator Manual*, *Test Administrator Manual or Scoring Guide*. Four countries reported that using the national adaptation form was somewhat difficult because of the large file size.

Table 7.12: NRC responses to questions about adapting and translating the ICCS assessment materials

Question	NRC responses (*n*)			
	Adaptation		Translation	
	Yes	No/not applicable	Yes	No/not applicable
Who adapted and translated the international version of the student test booklets?				
• Own staff	24	0	20	4
• Outside translator(s)	4	20	12	12
Who adapted and translated the international version of the student, teacher, and school questionnaires?				
• Own staff	24	0	20	4
• Outside translator(s)	4	20	12	12
Did you experience difficulties with the adaptation and/or translation of the international version of the following instruments to your national context?				
• Student test booklets	3	21	4	20
• Student questionnaire	6	18	3	21
• Teacher questionnaire	6	18	2	22
• School questionnaire	4	20	2	22
• Regional student questionnaire (where applicable)	1	23	1	23
Question	**Yes**		**No**	
Did you have major problems regarding the processes of external verification for any of the instruments (test booklets and questionnaires)?				
• Adaptation verification	1		23	
• Translation verification	3		21	
Question	**NRC responses (*n*)**			
How did you adapt and translate the international version of the instruments (test booklets and questionnaires)?				
• Directly from English source version	22		2	
• Directly from the translation provided by another national center	4		20	
• 2009 national version for trend items	21		3	
• 2009 national version for other than trend items	7		17	
• Other	2		22	

Assembling and printing the ICCS assessment materials

All countries completed the third step of the verification process (layout verification, conducted by the ISC) with no difficulties (Table 7.13). In general, the NRCs did not find it difficult to assemble the various survey instruments. A few printing errors (such as poor print quality, missing pages, or double printed pages) were detected, but these errors were minor and NRCs reported these were easily resolved.

Table 7.13: NRC responses to questions about assembling and printing the ICCS assessment materials

Question	NRC responses (*n*)				
	Not at all	A little	Somewhat	Very	Not answered/Not applicable
How difficult was it to assemble the test booklets and questionnaires?					
• Student test booklets	18	6	0	0	0
• Student questionnaire	22	2	0	0	0
• Teacher questionnaire	23	1	0	0	0
• School questionnaire	23	1	0	0	0
• Regional student questionnaire (where applicable)	19	1	0	0	4

Question	Yes	No	Not answered
Did you have major problems or do you have any comments regarding the process of external layout verification by the International Study Centre at ACER?	0	24	0
Did you detect any of the following problems during the printing process: problems with print quality, missing pages, page disorder, not intended upside down pages?	4	20	0
Did you discover any potential breaches of security (for example with printing, shipping, administration, submission to the national center)?	0	24	0

Administering the booklets and questionnaires

The NRCs were also asked about aspects of the administration of the booklets and questionnaires during the data collection (Table 7.14). Some 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 of the sampled students), or teachers from other schools.

As with the training of school coordinators, most countries used more than one approach to train their test administrators, relying most heavily on written instructions (18 countries), supplemented by formal training sessions (14 countries) or email/telephone correspondence (13 countries).

Table 7.14: NRC responses to questions about booklet and questionnaire administration

Question	NRC responses (*n*)		
Who were the test administrators or the main survey? (Check all that apply)			
• National center staff	8		
• Regional or district government staff	3		
• External contractor staff	6		
• Teachers from other schools	24		
• Teachers from the sampled schools but not of the sampled students	12		
• Teachers of the sampled students	6		
• School coordinators	12		
• Other	5		
How did you train the test administrators? (Check all that apply)			
• Formal face-to-face training sessions	14		
• Through telephone, email or video-link	13		
• Written instructions	18		
• Other	2		
Question	Yes	No	Not answered
For the administration of (paper-based) instruments, did you experience any difficulties to reach a high participation of students, teachers and school principals within the school?			
• Students	2	22	0
• Teachers	2	22	0
• School principals	2	22	0

Preparing and administering the online questionnaire

All countries were given the option of administering their school and/or teacher questionnaires online, and two-thirds used this delivery method (16 countries; see Table 7.15). Almost all countries reported that the adaptation of the online questionnaires was not difficult at all. The overall picture for the use of online administration is similar for both teachers and school principals. Reasons given for non-responses were lack of time or interest shown mostly by the teachers teaching subjects that were not related to civic and citizenship, although national centers put a lot of effort into supporting participation in schools. Overall, the countries experienced only minor challenges in achieving high participation rates for the online administration.

Table 7.15: NRC responses to questions about preparing and administering the online questionnaire

Question	NRC responses (*n*)		
	Yes	No	Not answered
Did you use the online data collection mode for administering the school and the teacher questionnaires?	16	8	0

Question	Not difficult at all	Somewhat difficult	Very difficult	Not applicable
How difficult was it to adapt the online questionnaires using the IEA SurveySystem Designer software?	15	1	0	8

Question	Yes	No	Not answered
For the online administration, did you experience any difficulties to reach a high participation of teachers or school principals within the school?			
• Teachers	8	8	8
• School principals	6	10	8

National quality assurance programs

The international project team provided materials to assist national centers in conducting their own national quality assurance program. The *ICCS 2016 Survey Operations Procedures Unit 4* provided guidelines for selecting the national quality observers (NQOs), and basic information about the observers' duties. In addition, the *ICCS 2016 National Quality Observer Manual* was developed for use in training NQOs. The NQOs' responsibilities were similar to those of the IQOs, and involved visiting selected schools to observe and document the data collection sessions and interview school coordinators.

Twenty NRCs confirmed that they had implemented a national quality assurance program, and that NQOs and their assistants had visited 360 schools in the participating countries. In three countries NRCs did not conduct national programs because they considered that information collected through the international program was sufficient or because the test administrators employed were considered very well trained. The NQOs confirmed the quality of the surveying process was generally good. There were only two issues reported: student absenteeism issues in a few countries, and online teacher participation rates in one country. Make-up sessions were organized in all countries except for one, where a new administration process for the absent students was denied.

Scoring open-ended response items and coding occupation data

NRCs were also asked to comment on the staff responsible for scoring open-ended response items, the procedures used to verify scoring reliability, and the coding of occupation data (see Table 7.16). The participating countries used scorers from a variety of different backgrounds, including national center staff, university students, and teachers. Some countries reported difficulties in scoring less detailed student responses, and some suggested that a greater variety of appropriate example responses might improve scoring.

There were two possible ways for coding occupations in the ICCS 2016 main survey. The first method was to enter the occupation information from the student questionnaire in the IEA Data Management Expert (DME), then export the occupation data to a specially designed Excel file, and code in the Excel file. An alternative way was to code the occupations directly in the instruments (for subsequent scanning or data entry using the DME). The majority of countries used the occupation data Excel file exported from the IEA DME to code occupation data. For both methods, the countries used experienced coders in the area that were trained using the translated version of the recommended coding scheme and examples from field trial.

Table 7.16 NRC responses to questions about scoring open-ended response items and coding occupational data

Question	NRC responses (*n*)			
	Yes	No	Not answered	Not applicable
Who scored your open-ended response items? (Check all that apply.)				
National center staff	14	0	0	0
Teachers/professional educators	8	0	0	0
University students	14	0	0	0
Other, please specify	3	0	0	0
Did you have any difficulties with the procedures for reliability scoring (i.e., double scoring of booklets 1, 2, and 3)?	3	20	1	0
Did you use the occupation data Excel file for coding occupations?	16	8	0	0
Did you have any difficulties with entering the occupation data to the occupation data Excel file?	0	16	0	8

Entering and submitting data

When asked who was responsible for entering questionnaire data, the NRCs reported using national center staff most often, followed by a combination of staff from the national center and an external data entry company (Table 7.17). In some cases, data entry was done by university students or other external assistants. Most countries used the IEA Data Management Expert (DME) software to enter data manually, and reported that they did not experience any problems with the software. In most cases, the rest of the countries used scanning procedures and subsequently imported data into the IEA DME.

Table 7.17 NRC responses to questions about entering and submitting data

Question	NRC responses (*n*)		
Did you use your own staff or external staff to enter the data from the test booklets and questionnaires into computer files?			
• Own staff	9		
• External data entry company	5		
• Combination of the above	5		
• Other	5		
Question	**Yes**	**No**	**Not answered**
Did you use the IEA Data Management Expert (DME) to manually enter your data?	20	4	0

The NRCs' assessment of the quality of the ICCS manuals

The survey activities questionnaire also 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 (Table 7.18). When asked for their suggestions for improvement, some NRCs suggested that the information in the manuals could be more concise, and that it would then be easier to locate the relevant instructions.

Table 7.18 NRC responses to questions about their impressions of the manuals provided

Question	NRC responses (*n*)			
	Very helpful	Somewhat helpful	A little helpful	Not helpful at all
Did you find the manuals helpful for carrying out the ICCS 2016 Main Survey?				
Survey Operations Procedures Unit 1	22	1	1	0
Survey Operations Procedures Unit 2	22	1	1	0
Survey Operations Procedures Unit 3	21	2	1	0
Survey Operations Procedures Unit 4	22	1	1	0
Survey Operations Procedures Unit 5	21	2	1	0
School Coordinator Manual	18	5	1	0
Test Administrator Manual	17	7	0	0
National Quality Observer Manual	18	5	1	0
Scoring Guide	22	2	0	0

CHAPTER 8:

Data management, cleaning and creation of the ICCS international database

Hannah Köhler and Christine Busch

Introduction

This chapter describes the procedures for verifying the ICCS 2016 data and creating the ICCS 2016 international database (IDB) implemented by the International Association for the Evaluation of Educational Achievement (IEA), the ICCS International Study Center (ISC) at the Australian Council for Educational Research (ACER), and the national centers of the participating countries.

Preparing the ICCS 2016 international database and ensuring its integrity was a complex endeavor requiring extensive collaboration between the IEA, the ISC, and the national centers of participating countries. Once the countries had created their data files and submitted them to the IEA, an exhaustive process of verification and editing known as 'data cleaning' began.

Data cleaning is the process of checking data for inconsistencies and formatting the data to create a standardized output. The overriding concerns of the data cleaning process 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.

Data sources

Data entry and verification of paper booklets and questionnaires

Each national center was responsible for transcribing the information from any paper-based test booklets and questionnaires into computer data files using the IEA Data Management Expert (DME) software. The DME is a software system developed by the IEA that facilitates data entry and incorporates validation checks to identify inconsistencies. As a general principle, national centers were instructed to enter data for any booklet or questionnaire that contained at least one valid response, discarding unused or empty instruments.

National centers entered responses from the paper instruments into data files created from an internationally predefined codebook. The codebook contained information about the names, lengths, labels, valid ranges for continuous measures or counts or valid values for nominal or ordinal questions, and missing codes for each variable.

Before data entry commenced, national centers were required to adapt the international codebook structure to reflect any approved adaptations made to the national questionnaire versions (e.g., a nationally added response category).

To ensure consistency across participating countries, the basic rule for data entry in the DME required national staff to enter data "as is" without any interpretation, correction, truncation, imputation, or cleaning. Resolution of any inconsistencies remaining after this data entry stage was delayed until data cleaning (see the section later in this chapter on "Cleaning the international and national databases").

The guiding principles for data entry included the following:

- Responses to categorical questions to be generally coded as "1" if the first option was used, "2" if the second option was marked, and so on.
- Responses to "check-all-that-apply" questions to be coded as either "1" (checked) or "9" (not checked/omitted).
- Responses to numerical or scale questions (e.g., school enrollment) to be entered "as is", that is, without any correction or truncation, even if the value was outside the originally expected range.
- Likewise, responses to filter questions and filter-dependent questions to be entered exactly as filled in by the respondent, even if the information provided was logically inconsistent.
- Non-responses to be coded as "omitted".
- Ambiguous responses, responses given outside of the expected format, or conflicting responses (e.g., selection of two options in a multiple-choice question), to be coded as "invalid".
- Misprinted questions or items to be entered as "not administered".

Data entered with the DME were automatically validated. As each respondent ID number was entered it was checked by the DME software for alignment with a five-digit checksum generated by the IEA WinW3S. A mistype in either the ID or the checksum resulted in an error message prompting the data entry person to check the entered values. The data-verification module of the DME also checked for a range of other issues such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. When such issues were flagged by the software, the individuals entering the data were prompted to resolve or to confirm the inconsistencies before resuming data entry.

Double data entry

To check data entry reliability in participating countries, national centers were required to enter a random sample of 30 units of each non-blank survey instrument (test booklets and questionnaires) twice by two different data entry persons (punchers). The IEA recommended that countries begin the double data entry process as early as possible during the data capture period in order to identify possible systematic incidental misunderstandings or mishandlings of data entry rules and to initiate appropriate remedial actions, for example, retraining national center staff. Those entering the data were required to resolve identified discrepancies between the first and second data entries by consulting the original instruments and applying the international rules in a uniform way.

While it was desirable that each and every discrepancy be resolved before submission of the complete dataset, the acceptable level of disagreement between the originally entered and double-entered data was established at one percent or less for questionnaire data and at the 0.1 percent or less level for the student test data. Values above this level required a complete re-entry of data. This restriction guaranteed that the margin of error observed for processed data remained well below the required threshold.

The level of disagreement between the originally entered and double-entered data was evaluated by the IEA, and it was found that in general the margin of error observed for processed data was well below the required threshold.

Online data collection of school and teacher questionnaires

As documented in Chapter 6 of this report, ICCS offered online collection of school and teacher questionnaire data as an international option conducted according to a mixed-mode design. Participating countries could adopt the online option as a default data-collection mode for some or all respondents (that is, school principals and teachers). National centers had to ensure that individual respondents who refused to participate in the online mode or who did not have access to the required infrastructure for online participation were provided with a paper questionnaire, thereby ruling out unit nonresponse as a result of a forced administration mode.

To ensure confidentiality, national centers provided every respondent with a letter that contained individual login information along with information on how to access the online questionnaire. This login information corresponded to the ID and checksum provided from the IEA WinW3S, meaning that the identity validation step occurring at the national centers for paper-based questionnaires occurred when the respondents' logged-in to the survey.

As respondents completed their online questionnaires, their data were automatically stored in one central international server and, therefore, no manual data entry was needed. Data for each country-language combination were stored in a separate table on the server. The different language versions within countries were then merged (at the IEA) with the data from the paper-based questionnaires and with data collected as part of the within-school sampling process.

Potential sources of error originating from the use of the two parallel modes had to be kept to the absolute minimum to ensure uniform and comparable conditions across modes and countries. To achieve this, ICCS questionnaires in both modes were self-administered, had identical contents and comparable layout and appearance, and required the data collection for both modes to take place over the same period of time.

Data verification at the national centers

Before sending the data to the IEA for further processing, national centers carried out mandatory validation and verification steps on all entered data and undertook corrections as necessary.

The corresponding routines were included in the DME software, which automatically and systematically checked data files for duplicate identification codes and data outside the defined valid ranges or value schemes. Data managers reviewed the corresponding reports, resolved any inconsistencies, and (where possible) corrected problems by looking up the original survey questionnaires. In addition, national centers were requested to perform several checks that identified inconsistent records across datasets. Data managers verified all findings prior to data submission and documented all changes or edits applied, as well as any verified findings that could not or did not need to be changed and therefore could remain.

While the instrument data were being entered, the data manager or other staff at each national center used the information from the survey tracking forms to verify the completeness of the materials. Participation information (e.g., whether a student participated in the assessment or was absent) was entered via the IEA WinW3S. The validation process was supported by an option in WinW3S to generate an inconsistency report. This report listed all types of discrepancies between variables recorded during the within-school sampling and test administration process and made it possible to cross-check these data against the actual availability of data entered in the IEA DME and the database for online respondents. Data managers were requested to resolve these problems before final data submission to the IEA. If inconsistencies remained or the national center could not solve them, the IEA asked the center to provide documentation on these problems. The IEA used this documentation when processing the data at a later stage.

Cleaning the international and national databases

Overview

As described earlier in this chapter, 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. Furthermore, the data from the online questionnaires were automatically stored in one central international server. Staff at the IEA then subjected these data to a comprehensive process of checking and editing. To facilitate the data cleaning process, the IEA 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, the national versions of test booklets and questionnaires, as well as information from the survey activities questionnaire (see details in Chapter 6). National centers also submitted their final national adaptation forms (NAFs) in order to provide and confirm complete documentation on all national adaptations. In addition, national centers were asked to provide documentation on all changes or edits applied to the data prior to submission, as well as any verified findings that could remain.

In order to ensure the integrity of the international database, a uniform data cleaning process was followed, involving regular consultation between the IEA and the NRCs. After each country had submitted its data and required documentation, the IEA, in collaboration with the NRCs, conducted a four-step cleaning procedure upon the submitted data and documentation:

(1) Documentation and structure check;

(2) Identification variable (ID) cleaning;

(3) Linkage cleaning;

(4) Background cleaning (resolving inconsistencies in questionnaire data).

The cleaning process was an iterative process. Numerous iterations of the four-step cleaning procedure were completed on each national data set. This repetition ensured that all data were properly cleaned and that any new errors that could have been introduced during the data cleaning were rectified. The cleaning process was repeated as many times as necessary until all data were made consistent and comparable. Any inconsistencies detected during the cleaning process were resolved in collaboration with national centers, and all corrections made during the cleaning process were documented in a cleaning report, produced for each country.

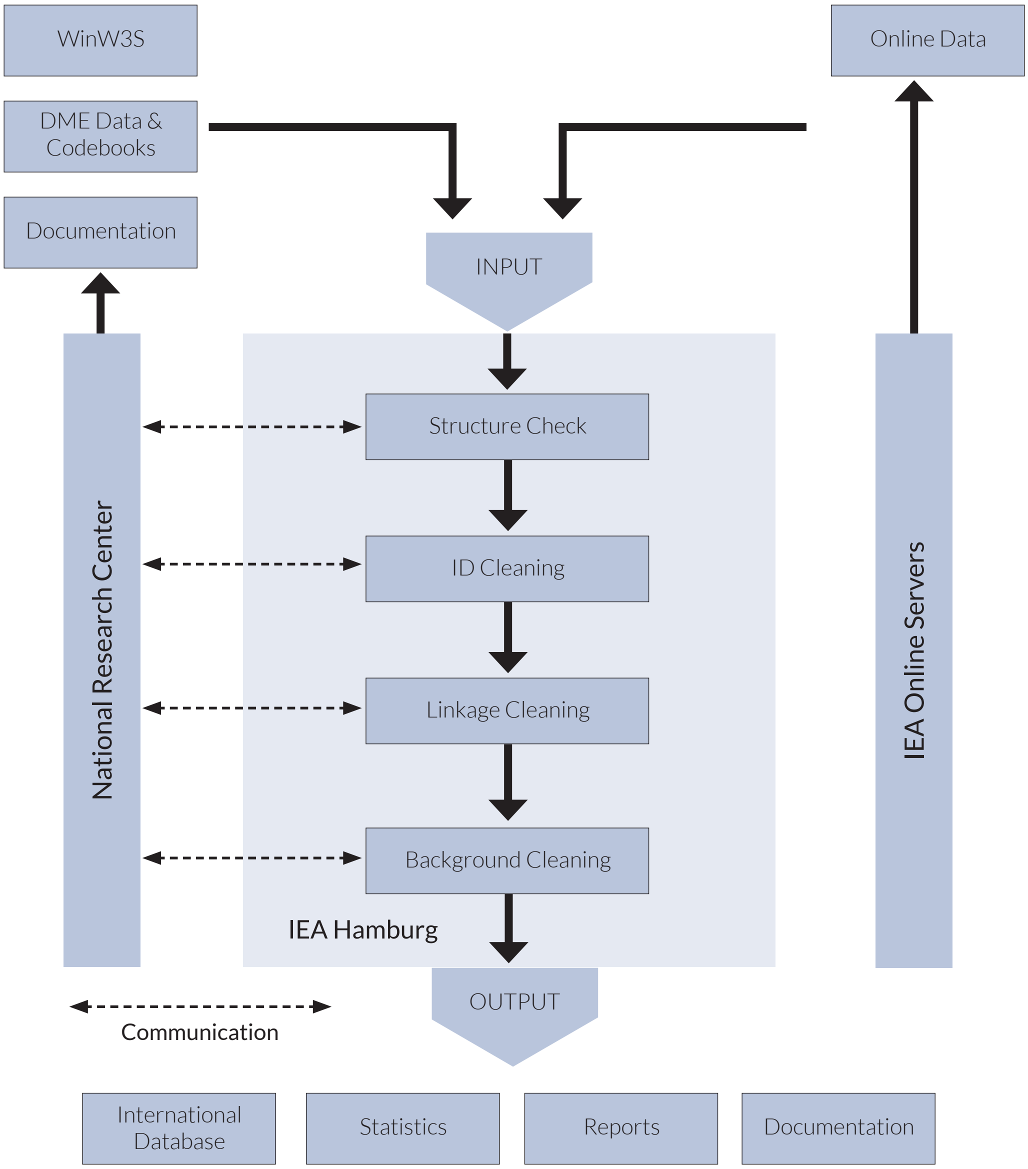
After the final cleaning iteration, the databases with information about student participation and exclusion were passed to the IEA Sampling Unit, which used this information to calculate students' participation rates, exclusion rates, and student sampling weights (see Chapter 9 for details). Afterwards the data, including sampling weights, were sent to the ICCS ISC so that scaling could be performed (see Chapter 10 and Chapter 11 for details). The NRCs were provided with interim data products to review at different points in the process.

Preparing national data files for analysis

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 a set of activities explained in the following subsections (Figure 8.1). The IEA carried out all of these activities in close communication with the national centers.

Figure 8.1: Overview of the data cleaning process



Checking documentation, import and structure check

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

The IEA first merged the tracking information and sampling information captured in the IEA WinW3S database with the student-level databases containing the corresponding student instrument data. During this step, IEA staff also merged the data from the school and teacher questionnaires for both the online and paper modes of administration. At this stage, data from the different sources were transformed and imported into one structured query language (SQL) database so that this information would be available during all further data-processing stages.

The first checks identified differences between the international and the national file structure. Some countries made adaptations to their questionnaires, such as adding national variables or omitting or modifying international variables. 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 cultural or language-specific terms), whereas other countries inserted response categories within existing international variables or added national variables.

To keep track of adaptations, the IEA asked the national centers to complete NAFs while they were adapting and translating the international version of the survey instruments. Where necessary, the IEA modified the structure of the national data files to ensure that the resulting data remained comparable across countries. Details about country-specific adaptations to the international instruments can be found in Appendix B of the *ICCS 2016 user guide for the international database* (Köhler, Weber, Brese, Schulz, & Carstens, 2018).

The IEA then discarded variables created purely for verification purposes during data entry, and made provisions for adding new variables necessary for analysis and reporting, including reporting variables, derived variables, sampling weights, and scale scores.

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

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

Cleaning identification variables

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. Some countries administered the school and teacher questionnaire online in addition to the paper mode. This could yield the possibility that a respondent completed both the paper and the online versions of the questionnaire. If two records in an ICCS database shared the same ID number and contained exactly the same data, the IEA deleted one of the records and kept the other one in the database. In the rare case that both records contained different data and IEA staff found it impossible to identify which record contained the "true data," national centers were asked to confirm which record should be kept.

Although the ID cleaning covered all data from all instruments, it focused mainly on the student questionnaire file. In addition to checking the unique student ID number, it was crucial to check variables pertaining to student participation and exclusion status, as well as students' dates of testing in order to calculate student age at the time of testing. The student tracking forms provided an important tool for resolving anomalies in the database.

Checking linkage

As data on students, their schools, and teachers appeared in a number of different data files, a process of linkage cleaning was implemented to ensure that the data files would correctly link together. The linking of the data files followed a hierarchical system of identification codes that included school, class, teacher and student components[7]. These codes linked the students with their class and/or school membership, as well as teachers with their school.

Linkage cleaning consisted of a number of checks to verify that student entries matched between student test files, student questionnaire files and scoring reliability files. In addition, at this stage, checks were conducted to ensure that teacher and student records linked correctly with their corresponding schools. The student tracking forms and teacher tracking forms were crucial in resolving any anomalies. The IEA also liaised with NRCs about any problematic cases, and the national centers were provided with standardized reports listing all inconsistencies identified within the data.

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 determined the treatment of inconsistent responses on a question-by-question basis using all available documentation to make an informed decision. IEA staff also checked all questionnaire data for consistency across the responses given.

For example, Question 18 in the school questionnaire asked for the total school enrollment (number of students) in all grades, while Question 19 asked for the enrollment in the target grade only. Clearly, the number given as a response to Question 19 could not possibly exceed the number provided by school principals in Question 18. The IEA flagged inconsistencies of this kind and then asked the national centers to review these issues.

A filter question, which appeared in the teacher questionnaires, directed respondents to a further section of the questionnaire. The IEA applied the following cleaning rules to the filter question and the dependent questions that followed:

- If the answer to the filter question was "No" the IEA recoded any responses to the dependent questions as "logically not applicable";
- If the response to the filter question was omitted but at least one valid response was found in the dependent questions then the IEA recoded the filter question to "Yes". This of course is only possible for dichotomous filter questions (e.g., with response options such "Yes/No").

7 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). The teacher ID includes the school ID (e.g., teacher 100103 may be described as teacher 3 in school 1012).

The IEA also applied what are known as *split variable checks* to questions where the answer was coded into several variables. For example, Question 3b 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 seven variables, each one coded as "Yes" if the corresponding "Yes" option was checked and "No" if the "No" option was filled in. 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," these responses were recoded accordingly, provided that the students had given affirmative responses in the other categories.

Resolving inconsistencies between tracking information and questionnaire data

Two different sets of ICCS 2016 data indicated the age and gender of students. The first set was the tracking information provided by the school coordinator or test administrator throughout the within-school sampling and test/questionnaire administration process. The second set comprised the actual responses given by students in the student questionnaires. In some cases, data across these two sets did not match and resolution was needed. If the information on gender or birth year and month was missing in the student questionnaire, but the student participated, this information, when available, was copied over from the tracking data to the questionnaire. If discrepancies were found between existing tracking and questionnaire gender and age data for students, the IEA queried the case with the national center, and the national center investigated which source of information was correct. If unresolved, tracking data were trusted over questionnaire data for students.

The teacher questionnaire did not ask teachers to provide birth year and month, but rather to choose between six age-ranges. Year of birth, which was indicated in the tracking forms, was then recoded into respective age groups and cross-checked against the range indicated by the questionnaire responses. If gender and/or age-range information was missing from the teacher questionnaire but the teacher participated, this data were copied over from the tracking information to the questionnaire. If discrepancies were found between existing tracking and questionnaire gender and age data, the questionnaire information replaced the tracking information.

Handling of missing data

Two types of entries were possible during the ICCS 2016 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 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:* this signified that the item or question was not administered to the respondent, which meant that the respondent could not read and answer the question. The not administered missing code was used for those student test items that were not in the set of assessment blocks administered to a student either deliberately (due to the rotation of assessment blocks) or, in a very few cases, due to incorrect translations. This missing code was also used for those records that were included in the international database but did not contain a single response to one of the assigned questionnaires. This situation applied to students who participated in the student test but did not answer the questionnaire, or vice versa. In addition, the not administered code was used for individual questionnaire items that a national center decided not to include in the country-specific version of the questionnaire. Furthermore, if a particular question or item (or a whole page) was misprinted or for other reasons not available to the respondent, then the corresponding variable(s) is coded as not administered.

- *Invalid:* this code was used in both the questionnaire and the test 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 filter question in a way that made the dependent questions following it not applicable to him or her.
- *Not reached:* this applied only to the individual items of the student test and indicated those items that students did not attempt due to a lack of time. 'Not reached' codes were derived as follows: First, the last answer given by a student in a session is identified. This could be either a valid or invalid response to an item. The first omitted response after this last answer is coded as 'omitted', but all following responses to these items in the session are then coded as 'not reached'. For example, the response pattern '1 9 4 2 9 9 9 9 9 9' (where '9' represents 'omitted') is recoded to '1 9 4 2 9 R R R R R' (where 'R' represents 'not reached').

Data cleaning quality control

Because ICCS 2016 was a large and highly complex study with very high standards for data quality, maintaining these standards required an extensive set of interrelated 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 applied them to simulation datasets containing all possible problems and inconsistencies.
- *Registering all incoming data and documents in a specific database:* the IEA 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 recorded these in the ICCS main survey general cleaning documentation that was provided to NRCs and NDMs.
- *Logging every "manual" correction to a country's data files in a recoding script:* logging these changes, which occurred only occasionally, allowed IEA 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 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 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 compared national adaptations recorded in the documentation for the national datasets against the structure of the submitted national data files. IEA staff then recorded any identified deviations from the international data structure in the national adaptation database and in the *ICCS 2016 user guide for the international database* (Köhler et al., 2018). Whenever possible, the IEA recoded national deviations to ensure consistency with the international data structure, however, if international comparability could not be guaranteed, the IEA removed the corresponding data from the international database.

Interim data products

Building the ICCS 2016 international database was an iterative process. On completion of each major data-processing step, the IEA 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 release of the draft and final versions of the international database. All interim data were made available in full to the ISC at ACER, whereas, as mentioned, each participating country received only its own data. Before the international databases were finalized, three major interim versions of the data files were sent to each country.

The first version was sent as soon as the data could be considered "clean" with regard to the four-step cleaning procedures described earlier in this chapter. The goal of this data verification and validation phase was to provide the participating countries with an opportunity to review a version of their national data as originally collected and/or derived after the IEA implemented per-country or general edits after data were submitted. The intention was to identify residual errors of systematic or incidental nature before the national databases were finalized with regard to weighting and scaling.

A second version of the data files was sent to countries when preliminary weights and scale scores were available and had been merged with the data files. The weights already reflected decisions made at the adjudication stage. Corrections due to feedback from countries, and general corrections or updates were implemented in this version. The goal of this data verification phase was to provide the participating countries with an opportunity to review their national data after additional country-specific and general corrections, and after weighting and scaling took place, before the national databases were finalized, in order to create the draft international database for analysis and reporting.

A third version, the draft international database, was made available to all participating countries under an affidavit of non-disclosure, reflecting all residual corrections, full ID scrambling, and general confidentiality edits. This version contained only records that satisfied the sampling standards and allowed the NRCs to replicate the results presented in the international reports. This version could also be used to prepare for national reporting.

Interim data products were accompanied by univariate statistics, containing unweighted summary statistics and frequencies of each variable. To enable the participating countries to thoroughly review the cleaning process, three further documents were supplied: a cleaning report, listing remaining case-level findings; a report documenting recodings for all country-specific data edits applied by the IEA; and a general cleaning documentation, describing in detail the data processing done at the IEA. An international codebook documenting the structure of all data files and a report documenting all national adaptations were also appended to each data version.

Final product: the ICCS 2016 international database

The ICCS international database incorporated all national data files from participating countries. The data processing and cleaning effort implemented at the IEA ensured that the ICCS 2016 international database contained high-quality data. More specifically, the process 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 successfully linked within and across levels;

- Only those records adjudicated as participating remained in the international database files; and
- Sampling weights and scale scores were available for international comparisons.

The indirect identification of individuals was prevented by applying general confidentiality edits, such as scrambling of identification variables and jackknife zone information. Furthermore, some of the personal data variables needed during field operations and data processing only were removed, and variables that were identified as highly identifying were suppressed or categorized. Two versions of the international database are available. The Public Use Files (PUF) are available to the general public, whereas the Restricted Use Files (RUF) are available to researchers only upon request (from the IEA). In the Public Use File some variables are removed or categorized to minimize the risk of disclosing confidential information; the Restricted Use File is an extended version for scientific use.

More information about the ICCS international database is provided in the *ICCS 2016 user guide for the international database* (Köhler et al., 2018).

Reference

Köhler, H., Weber, S., Brese, F., Schulz, W., & Carstens, R. (Eds.). (2018). *ICCS 2016 user guide for the international database*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

CHAPTER 9:

Weighting procedures

Sabine Weber, Sabine Tieck and Duygu Savaşcı

Overview

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 is largely based on Zuehlke (2011), and 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 in ICCS 2016. Not every student or teacher who completed a survey instrument was automatically regarded as 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 5, 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 calculated weights for all participating units in ICCS. All findings presented in ICCS reports are based on weighted data. Researchers 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 for 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 "participating school" if, in its sampled class, at least 50 percent of its students had 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 "participating class" if at least 50 percent of its students participated.
- A sampled school was regarded as "participating school" if all sampled classes participated.

In one ICCS country (Malta) all of the schools in the population and in another ICCS country (Croatia) all schools in one explicit stratum were selected for the study and all classes were selected. The class participation requirement also applied in these countries; 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 followed properly, 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 "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 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 $h=1,\ldots,H$. 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_{hi} 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, $WGTFAC1_{hi}$, was given by:

$$WGTFAC1_{hi} = \frac{M_h}{n_h^s \times M_{hi}}$$

where n_h^s is the number of sampled schools in stratum h, M_h is the total number of students enrolled in the schools of explicit stratum h, and M_{hi} 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, $WGTADJ1S_{hi}$, was calculated for each participating school i in stratum h as:

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

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

The number $n_h^{s,e}$ in this section is not necessarily equal to n_h^s in the preceding section, as $n_h^{s,e}$ is restricted to schools deemed as eligible to participate 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), developed by the IEA, 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, $WGTFAC2S_{hij}$, was given by:

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

where C_{hi} is the total number of classes with eligible students enrolled in the target grade and c_{hi}^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 is 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, $WGTADJ2S_{hij}$, for each participating class j was calculated as:

$$WGTADJ2S_{hij} = \frac{c_{hi}^{s}}{c_{hi}^{p}}$$

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

Student non-response adjustment (WGTADJ3S)

For all schools, the adjustment for student non-response inside each class for each participating student k, $WGTADJ3S_{hijk}$, was calculated as follows:

$$WGTADJ3S_{hijk} = \frac{s_{hij}^{e}}{s_{hij}^{p}}$$

where s_{hij}^{e} is the number of eligible students and s_{hij}^{p} is the number of participating students in class j in school i in stratum h. 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.

Final student weight (TOTWGTS)

The final student weight, $TOTWGTS_{hijk}$, of each student k in class j of school i in stratum h is the product of the five student-weight components:

$$TOTWGTS_{hijk} = WGTFAC1_{hi} \times WGTADJ1S_{hi} \times WGTFAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}$$

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 teacher data from the same school. To account for non-responding schools in the sample, the school weight adjustment for the teacher survey, $WGTADJ1T_{hi}$, was calculated as follows for each school i:

$$WGTADJ1T_{hi} = \frac{n_{h}^{s,e}}{n_{h}^{p-tch}}$$

where $n_{h}^{s,e}$ is again the number of sampled eligible schools and n_{h}^{p-tch} is the number of participating schools in the teacher survey in stratum h.

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*, $WGTFAC2T_{hil,}$ was calculated as:

$$WGTFAC2T_{hil} = \begin{cases} 1 & \text{for certainty teachers,} \\ \dfrac{T_{hi} - T_{hi}^{cert}}{t_{hi}^{s} - T_{hi}^{cert}} & \text{for non-certainty teachers,} \end{cases}$$

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

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, $WGTADJ2T_{hil}$, for each teacher *l* as:

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

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 $t_{hi}^{p-noncert}$ the number of participating non-certainty teachers in school *i* in stratum *h*. 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:

$$WGTADJ2T_{hil} = \frac{t_{hi}^{s,e-noncert} \times WGTFAC2T_{hi} + T_{hi}^{e-cert}}{t_{hi}^{p} \times WGTFAC2T_{hi}}$$

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 *h*. In the standard case, where all sampled teachers within a school were eligible for ICCS, this formula was simplified as follows:

$$WGTADJ2T_{hil} = \frac{T_{hi}}{t_{hi}^{p} \times WGTFAC2T_{hi}}$$

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

Teacher multiplicity adjustment (WGTADJ3T)

Some teachers in ICCS were teaching at the target grade in more than one school (the teacher questionnaire provided information about the number of schools a teacher was working in) and therefore had a larger selection probability. In order to account for this, a teacher multiplicity adjustment, $WGTADJ3T_{hil}$, was calculated as the inverse of the number of schools in which the teacher was teaching:

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

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 of each teacher *l* of school *i* in stratum *h* was the product of the five teacher-weight components:

$$TOTWGTT_{hil} = WGTFAC1_{hi} \times WGTADJ1T_{hi} \times WGTFAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}$$

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 school questionnaires.

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 student 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, $WGTADJ1C_{hi}$, was calculated as follows for each participating school *i*:

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

Here, n_h represents the number of sampled schools and n_h^{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* was the product of the two weight components:

$$TOTWGTC_{hil} = WGTFAC1_{hi} \times WGTADJ1C_{hi}$$

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, with only two exceptions: In Malta, one of the 199 sampled classes did not participate, so that the unweighted class participation rate was 198/199 = 99.5 percent. In one explicit stratum in Croatia, all schools and all classes were sampled. As one of the 82 classes in this stratum did not participate, the unweighted class participation rate was 99.9 percent.

Let *sfp* be the set of eligible and participating students in all participating schools, that is, in 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, $UPRS_{students}$, 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, $UPRS_{overall\_BR}$, was calculated as:

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

The unweighted overall participation rate in the student survey after replacement, $UPRS_{overall\_AR}$, is then given by:

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

Weighted participation rates in the student survey

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

$$WPRS_{schools\_BR} = \frac{\sum_{h}\sum_{i \in op}\sum_{j}\sum_{k \in sfp} WGTFAC1_{hi} \times WGTFAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}}{\sum_{h}\sum_{i \in op}\sum_{j}\sum_{k \in sfp} WGTFAC1_{hi} \times WGTADJ1S_{hi} \times WGTFAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}}$$

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

The weighted school participation rate in the student survey after replacement, $WPRS_{schools\_AR}$, was calculated as:

$$WPRS_{schools\_AR} = \frac{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGTFAC1_{hi} \times WGTFAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGTFAC1_{hi} \times WGTADJ1S_{hi} \times WGTFAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}}$$

The weighted class participation rate, $WPRS_{classes}$, was calculated as:

$$WPRS_{classes} = \frac{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGTFAC1_{hi} \times WGTFAC2S_{hij} \times WGTADJ33S_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGTFAC1_{hi} \times WGTFAC2S_{hij} \times WGTADJ2S_{hij} \times WGTADJ3S_{hijk}}$$

and the weighted student participation rate, $WPRS_{students}$, as:

$$WPRS_{students} = \frac{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGTFAC1_{hi} \times WGTFAC2S_{hij} \times 1_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGTFAC1_{hi} \times WGTFAC2S_{hij} \times WGTADJ3S_{hijk}}$$

The weighted overall participation rate in the student survey before replacement, $WPRS_{overall\_BR}$, was calculated as:

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

and the weighted overall participation rate in the student survey after replacement, $WPRS_{overall\_AR}$, was:

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

Overview of participation rates in the student survey

For all countries in the ICCS student survey, the unweighted participation rates (Table 9.1) and weighted participation rates (Table 9.2) for students and schools were calculated.

Table 9.1: 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
Belgium (Flemish)	80.0	98.2	94.8	75.8	93.1
Bulgaria	100.0	100.0	94.6	94.6	94.6
Chile	92.1	100.0	94.8	87.4	94.8
Chinese Taipei	93.3	94.0	97.7	91.2	91.9
Colombia	96.7	100.0	96.5	93.3	96.5
Croatia*	96.6	98.3	91.1	86.9	88.4
Denmark	52.5	84.8	93.0	48.9	78.9
Dominican Republic	96.5	100.0	96.3	92.8	96.3
Estonia	95.8	98.2	90.0	86.2	88.3
Finland	87.9	98.4	91.5	80.4	90.0
Hong Kong SAR	56.1	61.5	95.9	53.8	59.0
Italy	92.4	100.0	96.0	88.6	96.0
Korea Republic of	59.7	62.4	98.0	58.5	61.2
Latvia	92.3	94.2	88.8	82.0	83.7
Lithuania	99.5	99.5	91.7	91.2	91.2
Malta**	100.0	100.0	95.5	95.1	95.1
Mexico	91.9	95.9	95.7	87.9	91.8
Netherlands	54.0	82.0	92.6	50.0	75.9
Norway	96.6	100.0	93.5	90.4	93.5
Peru	100.0	100.0	96.9	96.9	96.9
Russian Federation	100.0	100.0	97.7	97.7	97.7
Slovenia	90.0	96.7	92.2	83.0	89.1
Sweden	98.1	98.7	91.1	89.3	89.9
Benchmarking participant					
North Rhine-Westphalia (Germany)	16.3	38.6	91.0	14.9	35.1

Notes:
* The unweighted class participation rate in Croatia is 98.7 percent.
** The unweighted class participation rate in Malta is 99.5 percent.

Table 9.2: 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
Belgium (Flemish)	79.9	98.2	94.7	75.7	92.9
Bulgaria	100.0	100.0	94.4	94.4	94.4
Chile	93.9	100.0	94.8	89.0	94.8
Chinese Taipei	93.2	93.9	97.7	91.0	91.7
Colombia	96.2	100.0	95.9	92.3	95.9
Croatia*	96.2	98.0	91.7	88.1	89.8
Denmark	54.5	84.8	93.0	50.7	78.9
Dominican Republic	96.8	100.0	96.6	93.5	96.6
Estonia	96.2	98.3	90.5	87.0	88.9
Finland	88.0	98.3	91.7	80.7	90.1
Hong Kong SAR	56.3	61.7	95.9	54.0	59.2
Italy	92.4	100.0	96.0	88.7	96.0
Korea Republic of	58.6	61.8	98.0	57.5	60.6
Latvia	92.7	93.9	89.3	82.7	83.8
Lithuania	99.3	99.3	92.1	91.5	91.5
Malta**	100.0	100.0	95.5	95.1	95.1
Mexico	93.5	96.7	95.6	89.4	92.5
Netherlands	52.6	81.9	92.5	48.7	75.8
Norway	95.2	100.0	93.7	89.2	93.7
Peru	100.0	100.0	96.8	96.8	96.8
Russian Federation	100.0	100.0	97.0	97.0	97.0
Slovenia	90.5	96.7	92.1	83.4	89.0
Sweden	98.3	98.8	90.8	89.2	89.7
Benchmarking participant					
North Rhine-Westphalia (Germany)	16.7	40.5	90.8	15.1	36.8

Notes:

* The weighted class participation rate in Croatia is 99.8 percent.

** The weighted class participation rate in Malta is 99.5 percent.

Unweighted participation rates in the teacher survey

To calculate unweighted participation rates in the teacher survey (see Table 9.3), let *op*, *fp* and *np*, and n^{op}, n^{fp} and n^{np} be defined as previously; the participation status now refers to the teacher survey instead of the student survey. 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, UPRTteachers, can then be defined as:

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

The unweighted overall participation rate in the teacher survey before replacement, $UPRT_{overall\_BR}$, can then be computed as:

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

and the unweighted overall participation rate in the teacher survey after replacement, $UPRT_{overall\_AR}$, can then be calculated as:

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

Table 9.3: 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
Belgium (Flemish)	77.6	95.2	90.2	70.0	85.8
Bulgaria	95.2	95.2	91.0	86.6	86.6
Chile	87.1	94.9	84.6	73.6	80.3
Chinese Taipei	95.3	96.0	98.6	94.0	94.7
Colombia	88.7	90.7	91.6	81.3	83.1
Croatia	97.2	98.9	95.4	92.7	94.3
Denmark	17.8	27.6	80.3	14.3	22.1
Dominican Republic	87.9	90.8	93.2	82.0	84.6
Estonia	28.1	29.3	61.1	17.2	17.9
Finland	83.0	93.4	84.1	69.8	78.6
Hong Kong SAR	n/a	n/a	n/a	n/a	n/a
Italy	92.4	100.0	95.8	88.4	95.8
Korea Republic of	67.8	71.1	99.3	67.3	70.6
Latvia	90.4	92.3	93.7	84.7	86.5
Lithuania	99.5	100.0	96.2	95.6	96.2
Malta*	100.0	100.0	96.3	96.3	96.3
Mexico	91.0	94.6	93.3	84.9	88.3
Netherlands	49.3	74.7	83.3	41.1	62.2
Norway	93.2	96.6	87.2	81.3	84.3
Peru	100.0	100.0	99.5	99.5	99.5
Russian Federation	39.8	39.8	99.9	39.7	39.7
Slovenia	88.7	95.3	93.3	82.8	89.0
Sweden	85.4	86.0	85.4	72.9	73.5
Benchmarking participant					
North Rhine-Westphalia (Germany)	n/a	n/a	n/a	n/a	n/a

Note:
n/a = not applicable.

Weighted participation rates in the teacher survey

The weighted participation rates for the teacher survey were calculated (see Table 9.4). The weighted school participation rate in the teacher survey before replacement, $WPRT_{schools\_BR}$, was calculated as:

$$WPRT_{schools\_BR} = \frac{\sum_{h}\sum_{i \in op}\sum_{l \in tfp} WGTFAC1_{hi} \times WGTFAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}}{\sum_{h}\sum_{i \in fp}\sum_{l \in tfp} WGTFAC1_{hi} \times WGTADJ1T_{hi} \times WGTFAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}}$$

The weighted school participation rate in the teacher survey after replacement, $WPRT_{schools\_AR}$, was calculated as:

$$WPRT_{schools\_AR} = \frac{\sum_{h}\sum_{i \in fp}\sum_{l \in tfp} WGTFAC1_{hi} \times WGTFAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}}{\sum_{h}\sum_{i \in op}\sum_{l \in tfp} WGTFAC1_{hi} \times WGTADJ1T_{hi} \times WGTFAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}}$$

The weighted teacher participation rate, $WPRS_{teachers}$, was calculated as:

$$WPRS_{teachers} = \frac{\sum_{h}\sum_{i \in fp}\sum_{l \in tfp} WGTFAC1_{hi} \times WGFAC2T_{hil} \times WGTADJ3T_{hil}}{\sum_{h}\sum_{i \in fp}\sum_{l \in tfp} WGTFAC1_{hi} \times WGTFAC2T_{hil} \times WGTADJ2T_{hil} \times WGTADJ3T_{hil}}$$

The weighted overall participation rate in the teacher survey before replacement, $WPRT_{overall\_BR}$, was:

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

and the weighted overall participation rate in the teacher survey after replacement, $WPRT_{overall\_AR}$, was:

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

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 January 2017, sampling referee Marc Joncas made binding decisions as to which country would be grouped in which category.

Table 9.4: 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
Belgium (Flemish)	74.6	95.2	89.9	67.1	85.6
Bulgaria	96.9	96.9	93.0	90.1	90.1
Chile	88.1	95.1	85.1	75.0	80.9
Chinese Taipei	95.3	95.9	98.8	94.1	94.7
Colombia	89.5	91.3	92.0	82.3	84.0
Croatia	97.5	98.9	96.6	94.2	95.5
Denmark	17.7	27.6	83.0	14.7	22.9
Dominican Republic	89.9	91.7	93.1	83.7	85.4
Estonia	27.9	29.5	62.4	17.4	18.4
Finland	81.0	92.6	83.5	67.7	77.3
Hong Kong SAR	n/a	n/a	n/a	n/a	n/a
Italy	92.7	100.0	95.7	88.7	95.7
Korea Republic of	68.7	72.9	99.4	68.3	72.4
Latvia	92.6	93.7	94.9	87.8	88.9
Lithuania	98.0	100.0	96.2	94.3	96.2
Malta	100.0	100.0	96.5	96.5	96.5
Mexico	93.7	97.3	92.7	86.8	90.2
Netherlands	49.3	74.8	82.5	40.7	61.7
Norway	89.2	96.7	87.8	78.4	85.0
Peru	100.0	100.0	99.7	99.7	99.7
Russian Federation	43.1	43.1	99.8	43.0	43.0
Slovenia	89.1	95.3	93.3	83.2	89.0
Sweden	85.5	86.0	84.1	71.9	72.4
Benchmarking participant					
North Rhine-Westphalia (Germany)	n/a	n/a	n/a	n/a	n/a

Note:
n/a = not applicable.

Student survey participation standards

Categories for sampling participation in the ICCS student survey were defined according to strict criteria (Figure 9.1).

Figure 9.1: Categories into which countries were placed with respect to student sampling participation rates

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

A country was in this category if:

- It had 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 in this category if:

- It failed to meet the requirements for Category 1 but has either an unweighted or 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 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 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 are included.

If a country did not meet the requirements for Category 1 or Category 2 but could provide documentation showing that they had complied with ICCS sampling procedures, it was placed in Category 3.

Teacher survey participation standards

The sampling participation categories for the teacher survey were similar to those in 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 towards 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. Three categories for teacher sampling participation were defined (Figure 9.2).

Figure 9.2: 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.

A country was in this category if:

- It had 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 of 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 in category 2 if:

- It failed to meet the requirements for Category 1 but had either an unweighted or 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 of at least 75 percent (after rounding to the nearest whole percentage point).

Category 3: Unacceptable sampling response rate even when replacement schools are included.

If a country did not meet the requirements for Category 1 or Category 2 but could provide documentation showing that they had complied with ICCS sampling procedures, it was placed in Category 3.

Reporting data

In those instances where a participating country could not be placed in participation Category 1, the ICCS team considered it necessary to make readers of the international reports aware of the increased potential for bias.

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

Participation categories were defined for each country for both the student and the teacher surveys (Table 9.5). In North Rhine-Westphalia (Germany), 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 benchmarking participant was not included in the analysis of teacher data in the ICCS 2016 international report. In Hong Kong SAR and in the Russian Federation, unapproved teacher selection procedures were applied in the majority of schools, which made it also impossible to calculate sampling weights. Therefore, it was not possible to report these countries' teacher data together with results from the other countries.

Table 9.5: Participation by country in the student and teacher surveys

Country	Student survey	Teacher survey
Belgium (Flemish)	1	2
Bulgaria	1	1
Chile	1	1
Chinese Taipei	1	1
Colombia	1	1
Croatia	1	1
Denmark	2	3
Dominican Republic	1	1
Estonia	1	3
Finland	1	2
Hong Kong SAR	3	–
Italy	1	1
Korea, Republic of	3	3
Latvia	1	1
Lithuania	1	1
Malta	1	1
Mexico	1	1
Netherlands	2	3
Norway	1	1
Peru	1	1
Russian Federation	1	3
Slovenia	1	1
Sweden	1	1
Benchmarking participant		
North Rhine-Westphalia (Germany)	3	–

Reference

Zuehlke, O. (2011). Sampling design and implementation. In: W. Schulz, J. Ainley, & J. Fraillon, (Eds.) (2011). *ICCS 2009 technical report*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

CHAPTER 10:

Scaling procedures for ICCS test items

Eveline Gebhardt and Wolfram Schulz

Overview

This chapter describes the procedures used to analyze and scale the ICCS 2016 international test items 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); and
- Estimation of changes in civic content knowledge between 2009 and 2016.

The development of the ICCS 2016 test items is described in Chapter 2 and was guided by the ICCS 2016 assessment framework (see Schulz, Ainley, Fraillon, Losito, & Agrusti, 2016).

The scaling model

We used item response theory (IRT) scaling methodology 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, θ_n is the estimated ability of person n, and δ_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 θ_n.

In the case of items with more than two categories (as, for example, with constructed open-ended test items with partial and full scores), we can generalize this model to the partial credit model (Masters & Wright, 1997), which takes the form:

$$P_{x_i}(\theta_n) = \frac{\exp\sum_{j=0}^{x}(\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i}\exp\sum_{j=0}^{h}(\theta_n - \delta_i + \tau_{ij})} \quad x_i = 0,1,\ldots,m_i$$

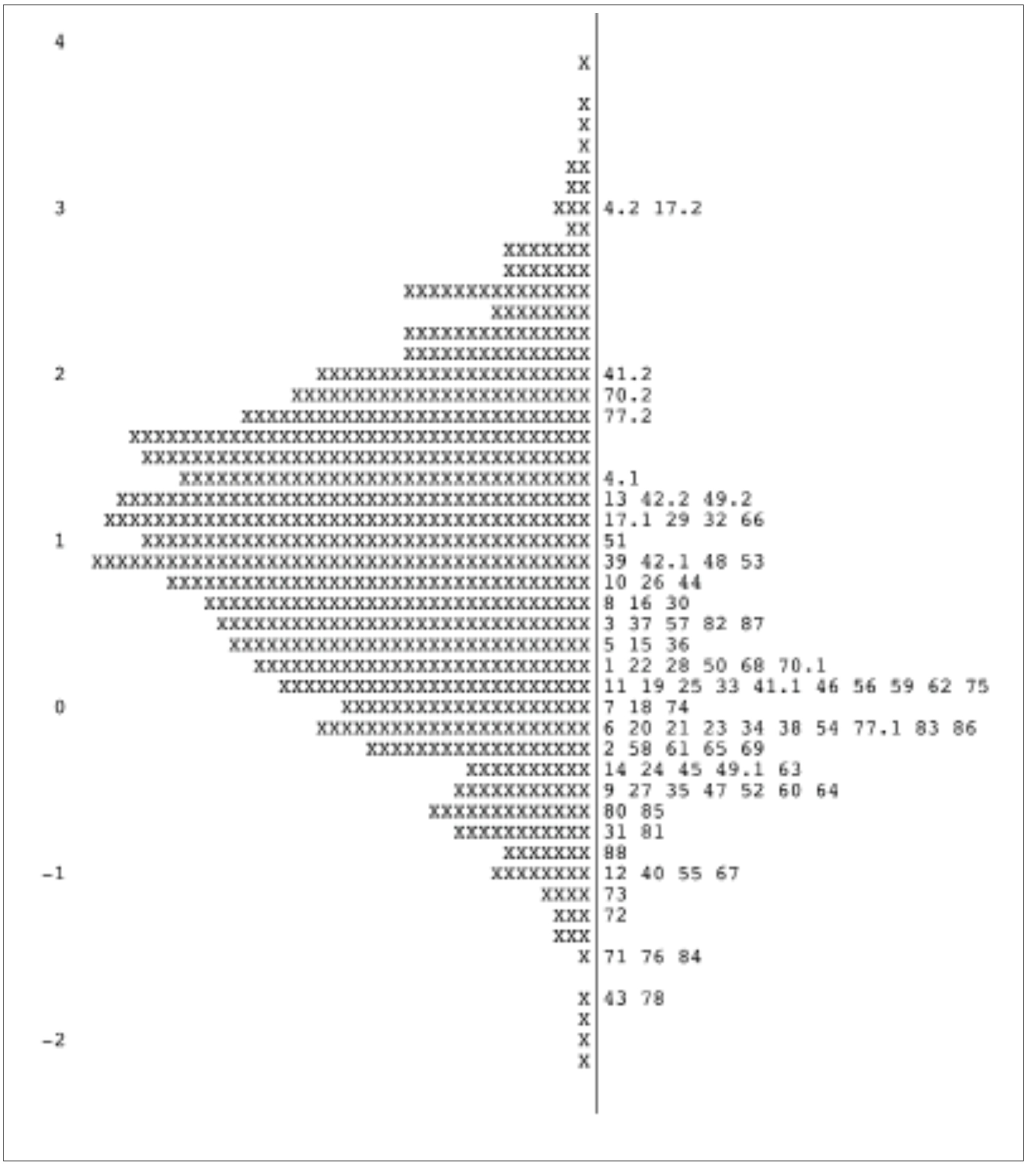
Here, $P_{x_i}(\theta_n)$ denotes the probability of person n scoring x on item i and θ_n denotes the person's ability. The item parameter δ_i gives the location of the item on the latent continuum; τ_{ij} denotes an additional step parameter for each step j between adjacent categories.

To scale the ICCS 2016 test data, we used the scaling software package ACER ConQuest, Version 4 (Adams, Wu, & Wilson, 2015).

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. As a start point, we estimated the distribution of cognitive abilities among ICCS 2016 students and the location of item thresholds (with a response probability, *rp*, of 0.5; see Figure 10.1). Item thresholds were equal to item difficulties of dichotomous items. For partial credit items, a difficulty threshold was estimated for each score.

Figure 10.1: Mapping of student abilities and item thresholds



The range of item difficulties broadly matched the abilities found in the student population. However, the average item difficulties were somewhat lower than the average student abilities. Overall, ICCS 2016 test items were better at targeting students in the lower than in the higher civic knowledge ranges. The nature of this targeting varied across countries according to the distribution of students' civic knowledge within each country.

We used multidimensional item response models to assess the dimensionality of items. Two of the possible item dimensions that we explored were based on the structure of the cognitive domains described in the ICCS 2016 assessment framework (Schulz et al., 2016).

We explored dimensionality with regard to the ICCS 2016 assessment framework content dimensions (civic society and systems versus others) and cognitive dimensions (knowing versus reasoning and applying). Multidimensional item response theory (IRT) models using the ACER ConQuest software typically showed latent correlations over 0.95, thus indicating a very high similarity between the item subgroups. Given these results, we decided not to report on any subscales for civic knowledge in the ICCS 2016 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 us 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). We consequently decided to assess item fit by using a range of item statistics.

We determined the item-rest correlations of correct responses (or partial credit responses) and the weighted item fit statistics (Table 10.1). There were no items with item-total correlations below 0.2 (indicating low discrimination), and only item CI2WFO2 showed relatively poor residual-based item fit (weighted mean square fit statistic of 1.24).

We also used ConQuest 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.

The ICC for item CI2WFO2 shows the discrimination was not entirely satisfactory and, although fewer than expected low performing students received a score of zero and fewer than expected high performing students received a score of 2, the curve still indicated that the students with higher levels of knowledge received higher scores for this item (Figure 10.2). Given this outcome, we decided to retain this item for scaling.

Figure 10.2: Item characteristic curve by category for item CI2WFO2

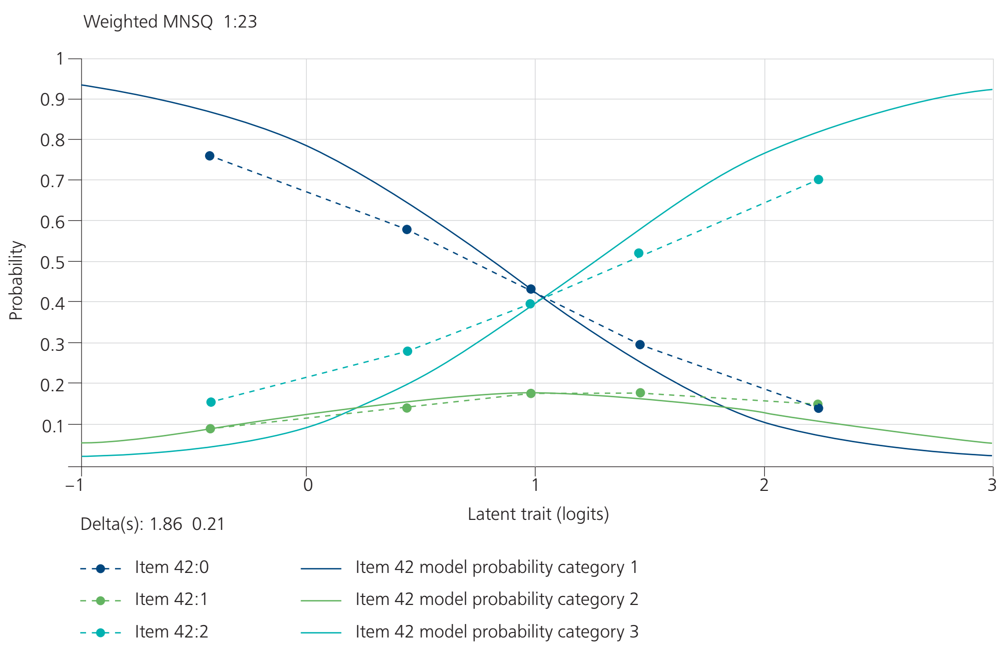


Table 10.1: Item total-score correlations and weighted item fit for international sample

Item no.	Item name	Item-rest correlation	Weighted fit	Item no.	Item name	Item-rest correlation	Weighted fit
1	CI2ASM1	0.36	1.00	45	CI308M1	0.40	0.96
2	CI2ASM2	0.44	0.92	46	CI312M1	0.42	0.95
3	CI2BCM1	0.37	0.99	47	CI314M1	0.29	1.03
4	CI2BIO1	0.33	1.17	48	CI3CAM1	0.27	1.08
5	CI2BPM1	0.28	1.09	49	CI3CBO1	0.47	1.01
6	CI2BPM2	0.42	0.95	50	CI3CPO1	0.38	0.97
7	CI2CCM1	0.36	1.00	51	CI3CPO2	0.41	0.96
8	CI2CCM2	0.34	1.02	52	CI3CRM1	0.43	0.92
9	CI2CEM1	0.39	0.96	53	CI3CRM2	0.35	1.01
10	CI2CEM2	0.22	1.12	54	CI3CSM1	0.36	0.99
11	CI2CNM1	0.35	1.00	55	CI3DBM1	0.38	0.95
12	CI2CNM2	0.46	0.90	56	CI3DBM2	0.44	0.94
13	CI2DLM1	0.37	0.98	57	CI3DDM1	0.38	0.99
14	CI2ECM1	0.46	0.91	58	CI3EPM1	0.40	0.94
15	CI2ECM2	0.27	1.09	59	CI3GMM1	0.32	1.04
16	CI2ETM2	0.24	1.12	60	CI3GTM1	0.37	0.95
17	CI2ETO1	0.36	1.09	61	CI3ICM1	0.34	1.03
18	CI2FDM1	0.38	0.99	62	CI3IEM1	0.35	1.01
19	CI2FSM1	0.42	0.94	63	CI3IVM1	0.45	0.93
20	CI2GFM1	0.36	1.01	64	CI3LPM1	0.46	0.91
21	CI2GLM1	0.38	1.00	65	CI3LSM1	0.42	0.94
22	CI2GLM2	0.32	1.03	66	CI3LTM1	0.31	1.04
23	CI2HRM1	0.36	0.97	67	CI3MAM1	0.30	1.02
24	CI2JOM1	0.31	1.02	68	CI3MDM1	0.42	0.97
25	CI2ORM1	0.44	0.92	69	CI3MPM1	0.42	0.94
26	CI2PCM1	0.34	1.02	70	CI3MPO2	0.39	1.11
27	CI2PCM2	0.39	0.95	71	CI3NPM1	0.32	0.97
28	CI2PGM1	0.38	0.99	72	CI3NPM2	0.30	0.96
29	CI2PGM2	0.33	1.04	73	CI3NWM1	0.36	0.96
30	CI2PJM1	0.23	1.11	74	CI3PAM1	0.50	0.90
31	CI2PJM2	0.41	0.91	75	CI3PEM1	0.34	1.01
32	CI2PRM1	0.32	1.03	76	CI3PLM1	0.36	0.92
33	CI2RCM1	0.38	1.00	77	CI3PRO1	0.43	1.07
34	CI2REM2	0.33	1.03	78	CI3REM1	0.37	0.90
35	CI2REM3	0.43	0.93	79	CI3RPM1	0.20	1.11
36	CI2SCM1	0.40	0.98	80	CI3RRM1	0.40	0.93
37	CI2SCM2	0.26	1.09	81	CI3SCM1	0.35	0.96
38	CI2VOM1	0.39	0.95	82	CI3SDM1	0.43	0.95
39	CI2VOM2	0.34	1.02	83	CI3SMM1	0.40	0.98
40	CI2VOM3	0.40	0.95	84	CI3SPM1	0.33	0.97
41	CI2WFO1	0.41	1.09	85	CI3SWM1	0.40	0.95
42	CI2WFO2	0.36	1.24	86	CI3UHM1	0.42	0.94
43	CI303M1	0.38	0.94	87	CI3ULM1	0.33	1.03
44	CI307M1	0.34	1.02	88	CI3VGM1	0.42	0.89

We analyzed the functioning of the partial credit scoring guides by reviewing the proportion of responses in each response category and the correct ordering of mean abilities of students across response categories. This analysis persuaded us that the scaling properties of all nine partial credit items could be satisfactorily included in the scaling of student test data.

Assessment of scorer reliabilities

In order to score the open-ended items in the ICCS 2016 cognitive test, we followed the scoring guides that were refined as an outcome of experiences in the international field trial of new test items. Within countries, for each of the eight booklets, subsamples of about 200 student records were scored twice by different scorers. This double- scoring procedure allowed us to assess scorer reliabilities and the percentages of scorer agreement, which ranged between 62 and 100 percent (Table 10.2). Internationally, scorer agreement for the six items was between 89 and 95 percent.

Table 10.2: Percentages of scorer agreement for open-ended ICCS test items

Country	Scorer agreement for ICCS test items (%)								
	CI2BIO1	CI2ETO1	CI2WFO1	CI2WFO2	CI3CBO1	CI3CPO1	CI3CPO2	CI3MPO2	CI3PRO1
Belgium (Flemish)	95	96	97	96	95	94	95	94	93
Bulgaria	82	82	81	81	87	90	97	85	77
Chile	92	96	95	86	93	93	92	94	91
Chinese Taipei	80	81	79	86	92	93	91	80	86
Colombia	100	100	100	100	100	100	100	100	100
Croatia	100	100	100	100	100	100	100	100	100
Denmark	92	92	96	96	93	97	98	93	98
Dominican Republic	100	100	100	100	100	100	100	100	100
Estonia	78	78	67	76	63	85	82	62	72
Finland	90	93	90	82	91	93	93	90	86
Hong Kong SAR	98	96	97	95	95	96	97	91	96
Italy	86	91	89	80	89	86	92	85	82
Korea, Republic of	83	94	88	89	86	99	95	93	86
Latvia	91	95	88	92	91	98	95	95	89
Lithuania	99	98	100	99	99	98	99	99	98
Malta	98	88	81	87	89	97	93	80	93
Mexico	94	96	96	93	96	95	92	95	95
Netherlands	85	84	79	75	87	88	90	78	90
North Rhine-Westphalia (Germany)	95	97	96	96	95	97	97	95	94
Norway	83	89	84	81	85	87	91	88	88
Peru	100	100	100	100	100	100	100	100	100
Russian Federation	98	97	98	97	97	99	100	97	97
Slovenia	92	92	86	88	84	89	94	92	91
Sweden	64	83	76	65	75	86	86	81	70
International average	90	92	90	89	90	94	94	90	90

We retained, for scaling and inclusion in the international database, data from items scored with a minimum of 60 percent scorer agreement. We made this adjudication for each open-response item scored in each country. None of the items needed to be removed because of low scorer reliability.

Differential item functioning by gender

We further explored the quality of the items by assessing 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.

We derived estimates of gender DIF by including interaction terms in the item response model. We could then model gender DIF for dichotomous items as:

$$P_i(\theta_n) = \frac{\exp(\theta_n - (\delta_i - \eta_g - \lambda_{ig}))}{1+\exp(\theta_n - (\delta_i - \eta_g - \lambda_{ig}))}$$

Parameter θ_n is the estimated ability of person *n* and δ_i is the estimated location of item *i*. For the purpose of measuring parameter equivalence across the two gender groups, we included two extra terms in the scaling model. The additional parameter for gender effects is denoted by λ_{ig}. To obtain proper estimates, we also needed to include the overall gender effect (η_g) in the model. We constrained both item-by-gender interaction estimates (λ_{ig}) and overall gender effects (η_g) to have a sum of 0.

Gender DIF estimates for a partial credit model for items with more than two categories (here, constructed items), with τ_{ij} denoting an additional step parameter, could then be modeled as:

$$P_{x_i}(\theta_n) = \frac{\exp\sum_{j=0}(\theta_n - (\delta_i - \eta_g - \lambda_{ig} + \tau_{ij}))}{\sum_{h=0}^{m_i}\exp\sum_{j=0}^{h}(\theta_n - (\delta_i - \eta_g - \lambda_{ig} + \tau_{ij}))} \quad x_i = 0,1,2,\ldots,m_i$$

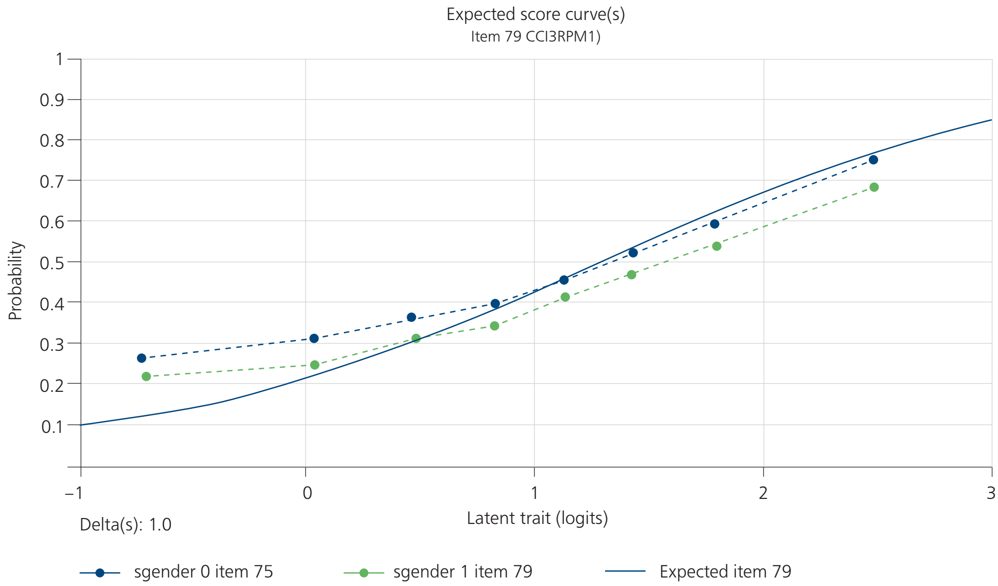
,

We thus estimated the gender DIF for each item (Table 10.3). One item (C13RPM1) showed relative large gender DIF (more than 0.30) and was, on average, 0.44 of a logit easier for male students than for female students with the same ability. When we plotted the gender DIF of this item (Figure 10.3), we found that male students (coded as 0 and indicated by the blue line) were more likely to choose the correct answer than female students (coded as 1 and indicated by the green line) at any ability level.

Table 10.3: Gender DIF estimates for test items

Item no.	Item name	Gender DIF estimate	Item no.	Item name	Gender DIF estimate
1	CI2ASM1	–0.30	45	CI308M1	0.00
2	CI2ASM2	–0.10	46	CI312M1	0.29
3	CI2BCM1	–0.02	47	CI314M1	–0.10
4	CI2BIO1	0.16	48	CI3CAM1	0.04
5	CI2BPM1	–0.23	49	CI3CBO1	0.12
6	CI2BPM2	0.20	50	CI3CPO1	0.33
7	CI2CCM1	–0.08	51	CI3CPO2	0.17
8	CI2CCM2	0.00	52	CI3CRM1	–0.01
9	CI2CEM1	–0.10	53	CI3CRM2	0.10
10	CI2CEM2	–0.08	54	CI3CSM1	0.17
11	CI2CNM1	–0.03	55	CI3DBM1	0.08
12	CI2CNM2	0.24	56	CI3DBM2	–0.04
13	CI2DLM1	–0.24	57	CI3DDM1	–0.02
14	CI2ECM1	0.06	58	CI3EPM1	–0.03
15	CI2ECM2	–0.04	59	CI3GMM1	–0.14
16	CI2ETM2	–0.26	60	CI3GTM1	–0.14
17	CI2ETO1	–0.03	61	CI3ICM1	–0.37
18	CI2FDM1	–0.05	62	CI3IEM1	–0.13
19	CI2FSM1	–0.11	63	CI3IVM1	0.21
20	CI2GFM1	0.04	64	CI3LPM1	0.24
21	CI2GLM1	–0.24	65	CI3LSM1	0.12
22	CI2GLM2	–0.02	66	CI3LTM1	–0.22
23	CI2HRM1	0.01	67	CI3MAM1	–0.17
24	CI2JOM1	–0.04	68	CI3MDM1	–0.09
25	CI2ORM1	0.10	69	CI3MPM1	–0.10
26	CI2PCM1	0.01	70	CI3MPO2	–0.14
27	CI2PCM2	0.14	71	CI3NPM1	0.11
28	CI2PGM1	0.08	72	CI3NPM2	0.31
29	CI2PGM2	0.02	73	CI3NWM1	0.20
30	CI2PJM1	–0.06	74	CI3PAM1	–0.01
31	CI2PJM2	0.19	75	CI3PEM1	–0.19
32	CI2PRM1	–0.09	76	CI3PLM1	0.25
33	CI2RCM1	–0.13	77	CI3PRO1	0.04
34	CI2REM2	0.01	78	CI3REM1	0.30
35	CI2REM3	0.26	79	CI3RPM1	–0.44
36	CI2SCM1	–0.04	80	CI3RRM1	0.13
37	CI2SCM2	–0.28	81	CI3SCM1	–0.16
38	CI2VOM1	0.02	82	CI3SDM1	–0.18
39	CI2VOM2	–0.05	83	CI3SMM1	0.14
40	CI2VOM3	0.39	84	CI3SPM1	–0.06
41	CI2WFO1	–0.03	85	CI3SWM1	0.24
42	CI2WFO2	0.06	86	CI3UHM1	0.01
43	CI303M1	0.07	87	CI3ULM1	0.03
44	CI307M1	–0.24	88	CI3VGM1	–0.04

Figure 10.3: Gender DIF plot for item CI3RPM1



Notes:
The expected average score is indicated by the solid blue line. Male students are indicated by the blue dashed line; female students are indicated by the green dashed line.

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 with the same ability, but from different countries, 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.

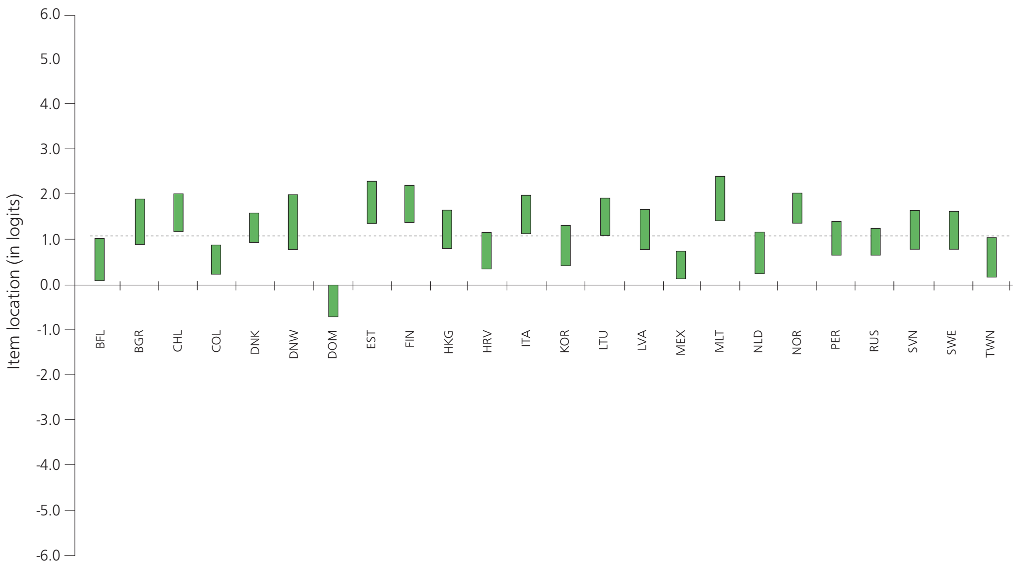
For the main survey analyses of ICCS 2016 test items, we compared national calibrations with international item parameters in order to assess the occurrence of item-by-country interaction. We then computed confidence intervals for each national item parameter, basing the computation on the respective standard errors and adjusting them for possible design effects and for multiple comparisons.

As an example, item CI3MPO2 showed significant item-by-country interaction in 12 countries (Figure 10.4). However, the size of the differences was below a predefined threshold that we had set for inclusion (i.e., less than 1.3 logits) and therefore this item was retained. We produced similar graphs for the test-item adjudication process at the international and national levels; information about occurrence of cross-national DIF was used to identify items for post-verification checks after completion of the main data collection.

Although the ICCS 2016 test items generally showed only limited item-by-country interactions, there were some national item difficulties that deviated quite considerably from the international item difficulty. In these cases, we omitted these items from scaling for those national samples where larger deviations had been observed (1.3 logits or more).

We also examined item-by-country interaction by item type, in particular by reviewing this for open-ended items. With these items, item-by-country interaction can be evidence of differences in the relative harshness of markers across countries. When comparing the relative difficulties of open-ended items to multiple-choice items across all countries, we found that all open-ended items

Figure 10.4: Example of an item-by-country interaction graph for item CI3MPO2



had been easier for students from the Dominican Republic than for students from other countries. This suggested there had been problems with the way scoring procedures had been conducted in the Dominican Republic, and we removed all nine open-ended items for this country from scaling and from 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). We used the omitted response category when a student provided no response at all to an item administered to him or her. Not-administered items were those that, although present in the whole item pool, were not in the 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.

The percentages of omitted and invalid responses for the international calibration sample (Table 10.4) indicated that there were considerably more omissions for open-ended items than for multiple-choice items and generally low percentages of invalid responses.

We created a separate missing category called "not reached" (coded as 6) for analysis purposes. We gave an item 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. The extent of occurrence of Code 6 items provided us with information about the appropriateness of test length and overall test difficulty. The average number of not reached items per student was 0.2 over all countries. There were two countries with considerably higher proportions of this type of missing values, with on average 1.2 "not reached" items per student.

Item adjudication

We conducted the adjudication of test items in two phases: first at the international level for the ICCS international calibration sample and then separately for each national subsample.

At the international level, we assessed item characteristics for the calibration sample. Here, we reviewed item-fit statistics, item-score correlations, item characteristic curves, general measurement equivalence across countries (item-by-country interaction), and gender DIF. For

Table 10.4: Percentages of omitted and invalid responses for test items

Item no.	Item name	Percentage omitted	Percentage invalid	Item no.	Item name	Percentage omitted	Percentage invalid
1	CI2ASM1	0.97	0.36	45	CI308M1	0.76	0.34
2	CI2ASM2	1.04	0.29	46	CI312M1	1.17	0.29
3	CI2BCM1	0.76	0.54	47	CI314M1	0.95	0.32
4	CI2BIO1	8.25	0.01	48	CI3CAM1	0.97	0.22
5	CI2BPM1	0.61	0.41	49	CI3CBO1	8.61	0.01
6	CI2BPM2	0.84	0.34	50	CI3CPO1	8.21	0.00
7	CI2CCM1	1.13	0.17	51	CI3CPO2	8.01	0.01
8	CI2CCM2	1.06	0.23	52	CI3CRM1	1.00	0.24
9	CI2CEM1	1.15	0.27	53	CI3CRM2	0.99	0.30
10	CI2CEM2	1.45	0.44	54	CI3CSM1	0.91	0.61
11	CI2CNM1	0.74	0.63	55	CI3DBM1	1.13	0.27
12	CI2CNM2	0.82	0.18	56	CI3DBM2	1.06	0.18
13	CI2DLM1	1.56	0.18	57	CI3DDM1	0.80	0.22
14	CI2ECM1	0.83	0.26	58	CI3EPM1	0.98	0.32
15	CI2ECM2	1.08	0.31	59	CI3GMM1	0.94	0.30
16	CI2ETM2	1.15	0.33	60	CI3GTM1	1.07	0.17
17	CI2ETO1	12.54	0.00	61	CI3ICM1	1.05	0.32
18	CI2FDM1	1.13	0.26	62	CI3IEM1	1.12	0.23
19	CI2FSM1	1.27	0.30	63	CI3IVM1	0.88	0.33
20	CI2GFM1	1.75	0.29	64	CI3LPM1	0.77	0.38
21	CI2GLM1	0.77	0.48	65	CI3LSM1	0.81	0.30
22	CI2GLM2	0.94	0.30	66	CI3LTM1	1.01	0.47
23	CI2HRM1	1.30	0.20	67	CI3MAM1	1.36	0.26
24	CI2JOM1	0.83	0.24	68	CI3MDM1	1.00	0.25
25	CI2ORM1	1.01	0.24	69	CI3MPM1	1.05	0.18
26	CI2PCM1	0.86	0.51	70	CI3MPO2	13.99	0.00
27	CI2PCM2	0.77	0.21	71	CI3NPM1	0.54	0.33
28	CI2PGM1	1.35	0.27	72	CI3NPM2	0.67	0.25
29	CI2PGM2	1.91	0.22	73	CI3NWM1	0.68	0.20
30	CI2PJM1	0.91	0.31	74	CI3PAM1	1.09	0.19
31	CI2PJM2	0.79	0.17	75	CI3PEM1	1.26	0.16
32	CI2PRM1	1.02	0.33	76	CI3PLM1	0.66	0.15
33	CI2RCM1	1.49	0.21	77	CI3PRO1	7.92	0.01
34	CI2REM2	1.13	0.59	78	CI3REM1	0.58	0.19
35	CI2REM3	0.73	1.13	79	CI3RPM1	1.11	0.38
36	CI2SCM1	1.14	0.30	80	CI3RRM1	0.91	0.17
37	CI2SCM2	1.05	0.32	81	CI3SCM1	0.75	0.19
38	CI2VOM1	1.48	0.16	82	CI3SDM1	2.00	0.43
39	CI2VOM2	1.33	0.33	83	CI3SMM1	1.06	0.39
40	CI2VOM3	1.26	0.16	84	CI3SPM1	0.63	0.37
41	CI2WFO1	8.32	0.02	85	CI3SWM1	0.63	0.29
42	CI2WFO2	9.65	0.00	86	CI3UHM1	0.72	0.32
43	CI303M1	0.57	0.16	87	CI3ULM1	1.06	0.45
44	CI307M1	1.75	0.26	88	CI3VGM1	0.98	0.17

open-ended items, we also considered scorer reliabilities and the correct ordering of average ability estimates per category. Only one of the 88 test items (CI3RPM1) had unsatisfactory scaling properties and showed considerable gender DIF. Therefore, we removed this item from the international scaling of civic knowledge.

When discussing this item further, it was decided that the item was somewhat confusing and more difficult than expected. Based on these observations and the large gender DIF, we decided to remove the item from the international scale.

At the national level, we reviewed test items by comparing national item-fit statistics with international item-fit statistics. We also flagged test items for individual countries that showed large item-by-country interactions (see section on cross-national measurement equivalence), and removed open-ended national items when the scorer agreement for the item fell below 60 percent. As mentioned above, we also omitted all open-ended items for one country, the Dominican Republic, because it was evident that students received relatively higher scores on these items than their international counterparts.

National centers were provided with item statistics (see example in Table 10.5) and were asked to review flagged test items. Item-level information 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. It also included highlighting open-ended items where the category-total correlations were disordered. In some cases, national centers informed the international study center of translation problems that had not been detected during verification. In these cases, we categorized the items as "not administered" in the international database and excluded them when scaling the corresponding national data.

Working independently from those conducting the national item reviews, the international study center 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, we identified additional national items that needed to be set to "not administered" in the international database and excluded from scaling of the corresponding national data.

In cases where items were translated correctly but showed 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), we excluded national items from scaling of the national data but retained them in the international database. We also excluded items from scaling across the various national samples because of translation/printing errors or large item-by-country interactions (Table 10.6).

Table 10.5: Example of item statistics provided to national centers

Parameter		Item												
Item number		1	2	3	4	5	6	7	8	9	10	11	12	13
Item name		CI2ASM1	CI2ASM2	CI2ASM3	CI2BIO1	CI2BPM1	CI2BPM2	CI2CCM1	CI2CCM2	CI2CEM1	CI2CEM2	CI2CNM1	CI2CNM2	CI2DLM1
Total cases (n)		2332	2332	2334	2293	2316	2316	2332	2332	2329	2329	2297	2297	2334
Facility 2016 (%)		74	83	79	38	91	79	80	86	81	51	76	94	62
Facility 2009 (%)		68	83	78	32	88	68	81	84	77	47	65	93	61
Discrimination		0.40	0.39	0.43	0.36	0.35	0.30	0.36	0.45	0.25	0.25	0.25	0.39	0.40
Percentages by response (%)	0				35									
	1	4	10	4	48	1	1	1	7	2	2	76 *	1	7
	2	7	83 *	4	14 *	91 *	13	80 *	86 *	4	36	2	94 *	18
	3	14	3	79 *		6	79 *	12	4	81 *	10	16	2	11
	4	74 *	3	10		1	4	6	1	13	51 *	6	3	62 *
	Omitted	0	0	1		0	1	0	0	0	1	0	0	0
	Not reached	1	1	0	3	0	1	0	0	0	0	0	0	1
	Invalid	0	0	1	0	0	0	0	1	0	0	0	0	0
Point bi-serial correlations	0				–0.29									
	1	–0.16	–0.23	–0.25	0.20	–0.19	–0.22	–0.12	–0.30	–0.17	–0.18	0.25*	–0.17	–0.17
	2	–0.24	0.39*	–0.26	0.21*	0.35*	–0.13	0.36*	0.45*	–0.25	–0.11	–0.24	0.39*	–0.19
	3	–0.18	–0.15	0.43*		–0.23	0.30*	–0.22	–0.21	0.25*	–0.13	–0.10	–0.26	–0.20
	4	0.40*	–0.20	–0.17		–0.18	–0.20	–0.20	–0.14	–0.05	0.25*	–0.14	–0.19	0.40*
	Omitted	–0.06	–0.06	–0.09		–0.06	–0.04	–0.03	–0.05	–0.08	–0.01	–0.03	–0.04	–0.04
	Not reached	–0.09	–0.09	–0.04	–0.18	–0.04	–0.03	–0.07	–0.08	–0.07	–0.08	–0.08	–0.09	–0.08
	Invalid	–0.13	–0.14	–0.11	–0.11	–0.11	–0.12	–0.14	–0.14	–0.10	–0.10	–0.08	–0.08	–0.09

Table 10.5: Example of item statistics provided to national centers (contd.)

Parameter		Item												
Item number		1	2	3	4	5	6	7	8	9	10	11	12	13
Item name		CI2ASM1	CI2ASM2	CI2ASM3	CI2BIO1	CI2BPM1	CI2BPM2	CI2CCM1	CI2CCM2	CI2CEM1	CI2CEM2	CI2CNM1	CI2CNM2	CI2DLM1
Rasch analysis	Rasch difficulty	0.39	-0.19	0.01	2.27	-1.06	0.07	0.04	-0.55	-0.01	1.65	0.34	-1.45	1.09
	Std error	0.05	0.06	0.06	0.04	0.08	0.06	0.06	0.07	0.06	0.05	0.05	0.09	0.05
	Step 1 difficulty				1.12									
	Step 2 difficulty				3.42									
	Mean square fit	0.98	0.97	0.93	1.04	0.93	1.03	0.98	0.93	1.04	1.07	1.07	0.87	0.96
	DIF girl	0.13	0.06	0.00	–0.22	–0.07	–0.09	0.15	0.01	0.37	0.15	0.11	–0.35	0.01
	DIF boy	–0.13	–0.06	0.00	0.22	0.07	0.09	–0.15	–0.01	–0.37	–0.15	–0.11	0.35	–0.01
International parameters	Rasch difficulty	0.26	–0.26	0.53	2.24	0.42	–0.12	–0.08	0.66	–0.48	0.78	–0.10	–1.00	1.22
	Point bi-serial	0.37	0.45	0.39	0.16	0.29	0.43	0.38	0.35	0.40	0.23	0.36	0.47	0.38
	Facility (%)	64	74	57	26	60	70	69	56	77	54	71	83	45

Note:
* Full credit response

Table 10.6: National items excluded from scaling

Country	Item set to "not administered" in database	Reason for deletion	Item excluded from scaling	Reason for exclusion
Belgium (Flemish)			CI3NPM1	Large DIF (+1.4)
Bulgaria			CI3CPO2	Large DIF (+1.3)
Chinese Taipei			CI2DLM1 CI2PCM1 CI2PCM2	CI2DLM1: Large DIF (–1.4) CI2PCM1: Large DIF (+1.4) CI2PCM2: Large DIF (+1.5)
Colombia	CI3CBO1 CI3CPO2 CI3EPM1 CI3SWM1	CI3CBO1: Translation error CI3CPO2: Translation error CI3EPM1: Distractor order error CI3SWM1: Distractor order error	CI3CPO2 CI3NPM1	CI3NPM1: Large DIF (–1.3) CI2BPM1: Large DIF(–1.4)
Croatia	CI303M1	Translation error	CI3GTM1 CI3RRM1 CI3EPM1	CI3GTM1: Large DIF (+1.4) CI3RRM1: Large DIF (–1.6) CI3EPM1: Large DIF (–1.3)
Denmark			CI2BPM1	Large DIF (–1.4)
Dominican Republic	CI2BIO1 CI2ETO1 CI2WFO1 CI2WFO2 CI3CBO1 CI3CPO1 CI3CPO2 CI3MPO2 CI3PRO1	All open-ended response items deleted due to issues relating to scoring accuracy		
Estonia			CI3IEM1	Large DIF(+1.7)
Hong Kong SAR			CI2PCM1 CI303M1	CI2PCM1: Large DIF (+1.6) CI303M1: Large DIF (+1.6)
Italy			CI2DLM1 CI2ETO1	CI2DLM1: Large DIF (+1.4) CI2ETO1: Large DIF (+1.3)
Korea, Republic of	CI3EPM1 CI3LPM1 CI2DLM1	CI3EPM1: Distractor order error CI3LPM1: Distractor order error CI2DLM1: Translation error	CI2ORM1 CI2PGM2 CI314M1 CI3DBM2 CI2REM3 CI312M1	CI2ORM1: Large DIF (–1.4) CI2PGM2 Large DIF (–1.4) CI314M1: Large DIF (+1.6) CI3DBM2: Large DIF (+1.4) CI2REM3: Large DIF (+1.3) CI312M1: Large DIF (–1.3)

Table 10.6: National items excluded from scaling (contd.)

Country	Item set to "not administered" in database	Reason for deletion	Item excluded from scaling	Reason for exclusion
Latvia	CI3VGM1 CI303M1	CI3VGM1: Distractor order error CI303M1: Translation error	CI307M1	Large DIF (+1.3)
Malta			CI307M1	Large DIF (+1.9)
Mexico			CI2PCM1	Large DIF (+1.6)
Netherlands	CI2PJM1 CI308M1	CI2PJM1: Translation error CI308M1: Distractor order error		
North Rhine-Westphalia (Germany)	CI2FDM1 CI303M1 CI3REM1	CI2FDM1: Translation error CI303M1: Translation error CI3REM1: Translation error		
Norway			CI3RPM1	Large DIF (+1.8)
Peru	CI3SWM1	Distractor order error	CI3CPO2 CI2BIO1	CI3CPO2: Large DIF (+2.0) CI2BIO1: Large DIF (+1.3)
Russian Federation	CI2CNM1	Translation error	CI2BIO1	Large DIF (–1.8)
Slovenia	CI3DDM1 CI3RPM1	CI3DDM1: Translation error CI3RPM1: Translation error		
Sweden			CI3SCM1	Large DIF (+1.3)

International item calibration and test reliability

Item parameter estimates were obtained from a joint data file that included response data from both ICCS 2016 and ICCS 2009. We included the ICCS 2009 data to improve the estimation of link items and for the purpose of equating, using the preferred IEA methodology also applied for the equating of TIMSS and PIRLS data. We included ICCS 2016 data from all 21 countries that met the sampling requirements and ICCS 2009 data from 18 countries that participated in both 2009 and 2016 and had met sample participation requirements. Countries were equally weighted within each ICCS cycle for the calibration and all items were included (except for items that were deleted nationally or internationally following the adjudication process).

We omitted missing student responses that were likely to be due to problems with test length ("not reached items") from the calibration of item parameters, but treated them as "incorrect" when scaling the student responses.

From this, we identified a set of item parameters that we used to scale the ICCS 2016 test data (Table 10.7).

Table 10.7: Final set of ICCS 2016 item parameters used to scale international test items

Item no.	Item name	Item parameter	Step 1	Item no.	Item name	Item parameter	Step 1
1	CI2ASM1	0.084		45	CI308M1	–0.504	
2	CI2ASM2	–0.455		46	CI312M1	0.019	
3	CI2BCM1	0.380		47	CI314M1	–0.672	
4	CI2BIO1	2.350	–0.938	48	CI3CAM1	0.808	
5	CI2BPM1	0.220		49	CI3CBO1	0.381	–0.600
6	CI2BPM2	–0.056		50	CI3CPO1	0.297	
7	CI2CCM1	–0.151		51	CI3CPO2	0.847	
8	CI2CCM2	0.480		52	CI3CRM1	–0.681	
9	CI2CEM1	–0.567		53	CI3CRM2	0.782	
10	CI2CEM2	0.698		54	CI3CSM1	–0.280	
11	CI2CNM1	0.023		55	CI3DBM1	–1.166	
12	CI2CNM2	–1.047		56	CI3DBM2	–0.014	
13	CI2DLM1	1.185		57	CI3DDM1	0.254	
14	CI2ECM1	–0.567		58	CI3EPM1	–0.541	
15	CI2ECM2	0.148		59	CI3GMM1	0.015	
16	CI2ETM2	0.434		60	CI3GTM1	–0.721	
17	CI2ETO1	2.078	–0.867	61	CI3ICM1	–0.301	
18	CI2FDM1	–0.011		62	CI3IEM1	–0.129	
19	CI2FSM1	0.048		63	CI3IVM1	–0.434	
20	CI2GFM1	–0.368		64	CI3LPM1	–0.854	
21	CI2GLM1	–0.258		65	CI3LSM1	–0.350	
22	CI2GLM2	0.073		66	CI3LTM1	1.012	
23	CI2HRM1	–0.225		67	CI3MAM1	–1.124	
24	CI2JOM1	–0.434		68	CI3MDM1	0.200	
25	CI2ORM1	0.086		69	CI3MPM1	–0.358	
26	CI2PCM1	0.489		70	CI3MPO2	1.033	–0.587
27	CI2PCM2	–0.744		71	CI3NPM1	–1.861	
28	CI2PGM1	0.101		72	CI3NPM2	–1.330	
29	CI2PGM2	1.054		73	CI3NWM1	–1.244	
30	CI2PJM1	0.555		74	CI3PAM1	–0.117	
31	CI2PJM2	–0.800		75	CI3PEM1	–0.044	
32	CI2PRM1	0.899		76	CI3PLM1	–1.658	
33	CI2RCM1	0.132		77	CI3PRO1	0.782	–0.840
34	CI2REM2	–0.102		78	CI3REM1	–2.036	
35	CI2REM3	–0.572		79	CI3RPM1	Removed	
36	CI2SCM1	0.393		80	CI3RRM1	–0.730	
37	CI2SCM2	0.321		81	CI3SCM1	–0.914	
38	CI2VOM1	-0.232		82	CI3SDM1	0.499	
39	CI2VOM2	0.574		83	CI3SMM1	–0.168	
40	CI2VOM3	–1.167		84	CI3SPM1	–1.655	
41	CI2WFO1	1.022	–0.850	85	CI3SWM1	–0.801	
42	CI2WFO2	1.040	1.026	86	CI3UHM1	–0.305	
43	CI303M1	–1.814		87	CI3ULM1	0.394	
44	CI307M1	0.540		88	CI3VGM1	–1.306	

Note:
Standard errors are not included because they are very small due to the large number of student responses and the assumption of a simple random sample during the estimation.

The overall reliability of the international test, as obtained from the scaling model, was 0.84 (ACER ConQuest estimate).

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 θ can be improved by using a larger number of test items. However, in large-scale surveys such as ICCS 2016, the purpose is to obtain accurate population estimates through use of instruments that also 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. 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 international calibration 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 more detailed description of the underlying methodology, see Adams, Wu, & Macaskill, 1997; also Adams, 2002.) In order to obtain estimates of students' civic knowledge, we used the ACER ConQuest software, which allowed us to draw plausible values (see Adams et al., 2015).

We used all available international student questionnaire variables and also those derived from regional instruments for conditioning of students. To deal with missing responses, we substituted all missing responses in a variable with either the mode or the mean and added, as extra variables, additional indicators for missing values. In Appendix D, we list all the international and regional student-level variables (along with their respective coding) that we used to condition the plausible values of civic knowledge (Table D.1).

Because of the large number of variables, we used the principal components of all student-level variables (single items or scale indices) as conditioning variables that reflected 99 percent of the variance. At the student level, we used only gender and its missing indicator as a direct conditioning variable. At the school level we included stratum indicators and classroom average ability estimates (WLEs), adjusted for a student's own ability estimate.

Equating ICCS 2016 to ICCS 2009

For ICCS 2009, we transformed the civic knowledge scale 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 was computed by applying the formula

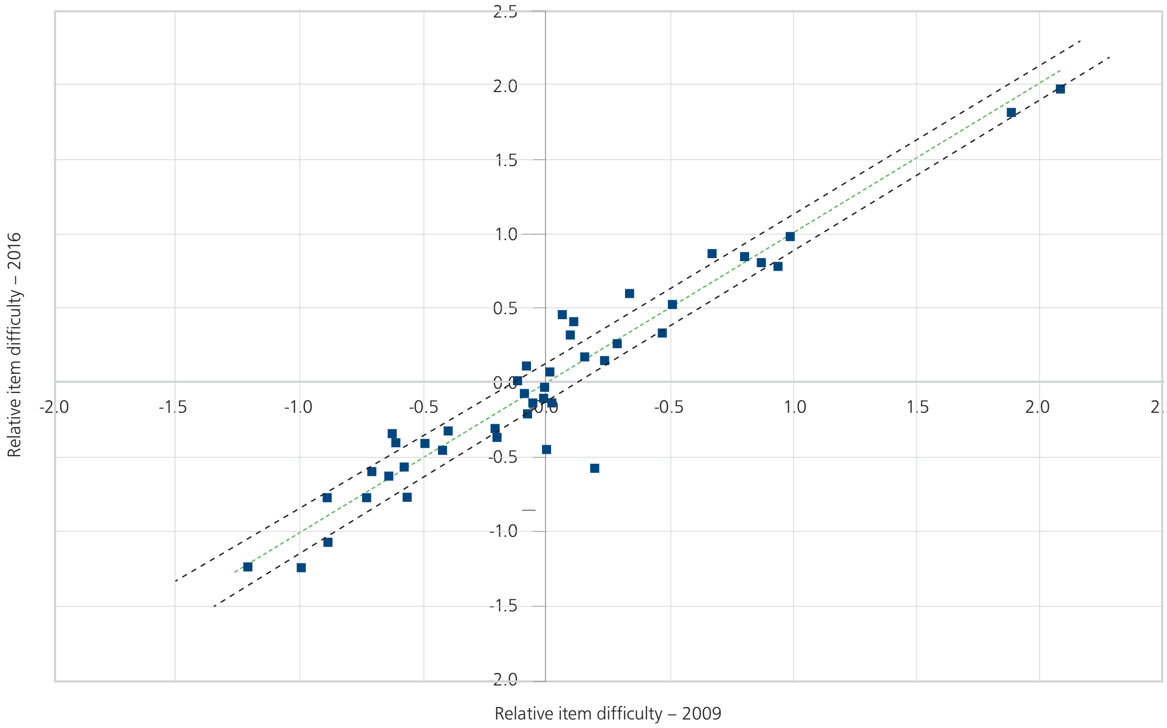
$$\theta_n' = 500 + 100\left(\frac{\theta_n^{09} - \bar{\theta}}{\sigma_\theta}\right)$$

where θ_n' were the student scores in the international metric, θ_n^{09} were the original logit scores, $\bar{\theta}$ was the international mean of student logit scores (-0.01) with equally weighted country subsamples, and σ_θ was its corresponding international standard deviation (0.95). This transformation was applied to each of the five plausible values.

In 2016, we computed the equating transformation to this historical by rescaling the ICCS 2009 data on the ICCS 2016 scale. As mentioned before, all ICCS 2009 item parameters were re-estimated concurrently during the ICCS 2016 calibration process.

Before joining the data from the two assessment cycles, we reviewed the relative difficulties of the common items to evaluate the quality of the link. A scatter plot of the relative difficulties of the 42 common items (Figure 10.5) revealed that the majority of items were within the confidence interval of complete measurement invariance. None of the differences between relative items difficulties was more than half a logit and there were no obvious outliers. Therefore, we decided to keep all the common items as link items.

Figure 10.5: Scatterplot for link item parameter estimates from ICCS 2009 and ICCS 2016



For equating purposes, the new item parameter estimates were used to redraw plausible values for the ICCS 2009 sample, using full conditioning. We only included the 18 countries that met the sampling requirements in both cycles. Subsequently, we computed the pooled mean and standard deviation of the plausible values on the 2009 scale and on the 2016 scale. Comparing those distributions resulted in the following linear transformation to equate the ICCS 2016 student abilities (θ_n^{16}) onto the historical ICCS 2009 scale in logits (θ_n^E):

$$\theta_n^E = 0.996 \times \theta_n^{16} - 0.634$$

These equated plausible values were subsequently placed on the ICCS 2016 international reporting scale by applying the transformation from ICCS 2009:

$$\theta_n' = 500 + 100\left(\frac{\theta_n^E - (-0.01)}{0.95}\right)$$

Uncertainty in the link

Because the transformation equating the ICCS 2016 data with the ICCS 2009 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 we had chosen an alternative set of link items. Uncertainty in the transformation thus relates to the sampling of the link items, 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. We can, however, estimate the likely range of magnitudes for this error and 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.

The following approach has been used to estimate the equating error. Suppose we have a total of L score points in the link items in K modules. Using i to index items in a unit and j to index units, $\hat{\delta}_{ij}^{y}$ is the estimated difficulty of item i in unit j for year y, with differences calculated as:

$$\hat{c}_{ij}^{2016} - \hat{\delta}_{ij}^{2009}$$

The size (number of score points) of unit j is m_j so that:

$$\sum_{j=1}^{K} m_j = L \text{ and } \bar{m} = \frac{1}{K} \sum_{j=1}^{K} m_j$$

Further let:

$$c_{\cdot j} = \frac{1}{m_j} \sum_{i=1}^{m_j} c_{ij} \text{, and } \bar{c} = \frac{1}{N} \sum_{j=1}^{K} \sum_{i=1}^{m_j} c_{ij}$$

and then the link error, taking into account the clustering was computed as follows:

$$LinkError_{2016,\,2009} = \sqrt{\frac{\sum_{j=1}^{K} m_j^2 (c_{\cdot j} - \bar{c})^2}{K\,(K-1)\bar{m}^2}} = \frac{\sum_{j=1}^{K} m_j^2 (c_{\cdot j} - \bar{c})^2}{L^2} \; \frac{K}{K-1}$$

The 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.

When developing proficiency levels, we applied a method that ensured that the notion of "being at a level" could be interpreted consistently and which recognized that the achievement scale is a continuum. We therefore attempted 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?

We adopted the following two parameters for defining proficiency level:

- *The response probability (rp) for reporting item parameters:* this was set at *rp* = 0.62;
- *The width of the proficiency levels:* this was set at 0.8 logits.

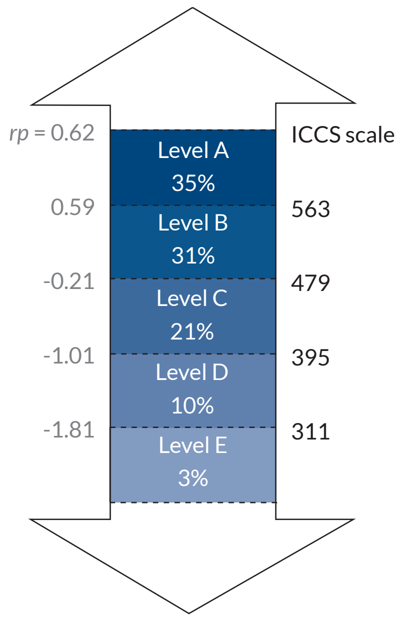
Using these parameters, we were able to infer the following about students' aptitude in relation to the proficiency levels:

- 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 that we chose 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.

We thus identified four proficiency levels that could be used when reporting student performances from the assessment (Figure 10.6 shows the cut-points for these levels) and the percentage of students at each proficiency level across the participating ICCS countries.

Figure 10.6: Proficiency level cut-points and percentage of students at each level



When reporting released test items and mapping them against proficiency levels, we had to transform location parameters of these items to a value that reflected a response probability of 62 percent. We achieved this 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 δ_i' for each item that we obtained was as follows:

$$\delta_i' = 500 + 100 \times \left(\frac{\delta_i + \ln(0.62/0.38) - \bar{\theta}}{\sigma_\theta} \right)$$

Here, δ_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 σ_θ is its corresponding international standard deviation (0.95).

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: 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.

Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). A*CER ConQuest: Generalised Item Response Modelling Software* [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.),. *Handbook of modern item response theory* (pp. 101–122). New York/Berlin/Heidelberg: Springer.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *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, Fraillon, J., J., Losito, B., & Agrusti, G. (2016). *IEA International Civic and Citizenship Education Study 2016 Assessment Framework*. 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.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: Mesa Press.

CHAPTER 11:

Scaling procedures for ICCS questionnaire items

Wolfram Schulz and Tim Friedman

Introduction

This chapter describes the procedures for the scaling of the ICCS questionnaire data (for students, teachers and schools) and the indices that were derived from these data.

Generally, it is possible to distinguish two general types of indices based on data from the ICCS 2016 questionnaires:

- *Simple indices* were constructed through arithmetical transformation or recoding; for example an index of immigration background based on information about the country of birth of students and their parents.
- *Scale indices* that were derived through the scaling of items; this was typically achieved by using item response modeling of items with two or more categories.

The first part of this chapter lists the simple indices that were derived from ICCS 2016 data and describes how they were created. The second part outlines the procedures used for the scaling of questionnaire data in ICCS 2016. Finally, the third part lists the scaled indices with statistical information on the factor structure of related item sets, scale reliabilities and parameters used for the IRT scaling.

Results from an analysis of cross-country validity of item dimensionality and constructs were already part of ICCS 2016 field trial analyses. At this time, the International Study Center at ACER used field trial data to conduct a review of the extent to which measurement models held across participating countries for draft item material. The review made use of exploratory and confirmatory factor analysis, and item response modeling to examine cross-national measurement equivalence before the final selection of main survey questionnaire items (see an overview of the different types of analyses in Schulz, 2009).

Simple indices

Student questionnaire

Student age (S_AGE) was calculated as the difference between the year and month of the testing and the year and month of a student's birth. Information from the student questionnaire (Question 1) was used to derive age, except for students where this information was missing. In these cases information from student tracking forms (see Chapter 8 for more details) provided data for the calculation of this index. The formula for computing S_AGE was:

$$S\_AGE = (T_y - S_y) + \frac{(T_m - S_m)}{12}$$

where T_y and S_y are, respectively, the year of the test and the year of birth of the tested student, in four-digit format (e.g., "2016" or "2001"), and where T_m and S_m are respectively the month of the test and the month of the student's birth. The result was rounded to two decimal places.

In Question 2, students were asked their sex. These were recorded as the *student gender* (S_GENDER) with a value of 1 assigned to females and value of 0 assigned to males. For students with omitted data for this question, we used the information from the tracking form instead.

Question 3 asked students indicate their expected highest level of educational attainment. The corresponding index of *students' expected educational attainment* (S_ISCED) had the following categories:

(0) Completion of ISCED level 2 (lower-secondary) or lower;

(1) Completion of ISCED level 3 (upper-secondary);

(2) Completion of ISCED level 4 (non-tertiary post-secondary) or ISCED level 5 (short-cycle tertiary education);

(3) Completion of ISCED level 6 (bachelor or equivalent), ISCED level 7 (masters or equivalent) or ISCED level 8 (doctoral or equivalent).

In Question 4, students were required to indicate the country of birth for themselves and for each of their parents. The index of *immigrant family background* (S_IMMIG) was created using these data, and had three categories:

(1) Native students (students who had at least one parent born in the country of assessment);

(2) First-generation students (students who were born in the country of assessment and whose parent(s) were born in another country);

(3) Non-native students (students who were born outside the country of assessment and whose parent(s) were born in another country).

We assigned missing values to students with missing responses for either their own place of birth, or that of their mother and father, or for all three questions. The analysis of immigrant background and civic knowledge was based on a dichotomous indicator variable that distinguished between students from immigrant families (Categories 2 and 3) and students from non-immigrant families (Category 1).

Question 5 of the ICCS 2016 student questionnaire asked students what language they speak at home most of the time, with each country providing a list of options relevant to their context. We used this information to derive an index on *test language used at home* (S_TLANG), in which responses were grouped into two categories:

- The language spoken at home most of the time differed from the language of assessment.
- The language spoken at home most of the time was the language of assessment.

Occupational data for each student's two parents were obtained by asking students to provide details as to what their jobs are, using open-ended questions (Questions 6 and 8). The responses to these questions were coded by national centers into four-digit codes using the International Standard Classification of Occupations (ISCO-08) framework (International Labour Organization, 2007). These codes are contained in the indices S_MISCO (student's mother) and S_FISCO (student's father). We then mapped these codes to the international socioeconomic index of occupational status (ISEI) (Ganzeboom, de Graaf, & Treiman, 1992). The three indices that we obtained from these scores were *mother's occupational status* (S_MISEI), *father's occupational status* (S_FISEI), and the *highest occupational status of both parents* (S_HISEI), with the last 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.

Questions 7 and 9 asked about the education attainment of students' mothers and fathers. 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). In ICCS 2016 we classified levels of parental education according to the International Standard Classification of Education (ISCED) (UNESCO, 2011).

Recoding educational qualifications into the following categories provided indices of highest parental educational attainment:

(0) Did not complete ISCED level 2;

(1) ISCED level 2 (lower-secondary);

(2) ISCED level 3 (upper-secondary);

(3) Completion of ISCED level 4 (non-tertiary post-secondary) or ISCED level 5 (short-cycle tertiary education);

(4) Completion of ISCED level 6 (bachelor or equivalent), ISCED level 7 (masters or equivalent) or ISCED level 8 (doctoral or equivalent).

For each student, indices with these categories are available for the *mother's highest educational attainment* (S_MISCED) and *father's highest educational attainment* (S_FISCED). The index for *highest educational level of parental education* (S_HISCED) corresponds to the higher educational attainment of either parent.

In Question 10, students were asked to indicate how interested they are and how interested their parent(s) are in political and social issues. The (recoded) indices for *student interest* (S_SINT) *mother interest* (S_MINT) and *father interest* (S_FINT) consisted of the following categories: (0) "not interested at all," (1) "not very interested," (2) "quite interested" and (3) "very interested". An index of the *highest level of parents' interest in political and social issues* (S_HINT) was created by computed the maximum value of both S_MINT and S_FINT.

Question 11 of the student questionnaire asked students how many books they had in their homes. Responses to this question formed the basis for an index of students' *home literacy resources* (S_HOMLIT) with the following categories:

(0) None or very few (0–10 books);

(1) Enough to fill one shelf (11–25 books);

(2) Enough to fill one bookcase (26–100 books);

(3) Enough to fill two bookcases (101–200 books);

(4) Enough to fill three or more bookcases (more than 200 books).

The religion of each student was captured in Question 33. This question was not mandatory internationally, and some countries opted not to administer it. Those countries that did participate provided a list of options relevant to their context. We used this information to derive an index on *student's religious affiliation* (S_RELIG) in which responses were grouped into two categories:

(0) The student indicated that they do not have a religion;

(1) The student indicated that they do have a religion.

Countries were also given the option to administer a question on the frequency that students attend religious services (Question 34). A simple index was formed capturing students' attendance of religious services (S_RELSER) based on the following categories:

(0) Never;

(1) Less than once a year;

(2) At least once a year;

(3) At least once a month;

(4) At least once a week.

Teacher questionnaire

Question 2 of the teacher questionnaire asked teachers to indicate what percentage of their classroom teaching time was at the target grade during current school year. Responses to this question were used to form the index teacher's teaching load at school (T_TIME). This index was coded so that a value of 0.1 reflects "less than 20%," 0.3 "20–39%," 0.5 "40–59%," 0.7 "60–79%," and 0.9 "80% and more."

The individual teacher age (T_AGE) consisted of the midpoint of the age ranges given in Question 3 of the teacher questionnaire. We assigned "less than 25" a value of 23 and coded "60 or over" as 63.

The sex of teacher (T_GENDER) was computed from the data captured from Question 5 of the teacher questionnaire. Female teaches were coded as 1, whereas male teachers were coded as 0.

School questionnaire

Question 17 collected information about whether the school was a public school or a private school. This information was used to form an indicator of private school management (C_PRIVATE), where public schools were coded as 0 and private schools were coded as 1.

Question 18 asked principals to provide the numbers of female and male students enrolled at their school. To calculate an indicator of school size (C_SCSIZE), we used a simple addition of the total number of boys enrolled on a suitable date at the beginning of the school year (IC3G18A) and the total number of girls enrolled on the same date (IC3G18B).

Question 19 asked principals to provide the numbers of female and male students at the target grade enrolled at their school. We used the same procedure to calculate school enrollment in the target grade (C_GRENROL) by adding the number of boys (IC3G19A) and the number of girls (IC3G19B) at each target grade. To calculate the percentage of target-grade students at each school (C_TGPERC), we divided the number of students enrolled in the target grade (C_GRENROL) by the total number of enrolled students at the school (C_SCSIZE). We then multiplied this value by 100.

Question 20 collected information about size of the immediate community where the school was located. This information was used to form an indicator of school urbanity (C_URBAN), where schools located in communities with more than 100,000 inhabitants were coded as 1 and other schools were coded as 0.

Question 21 asked principals to estimate the percentage of students at their school coming from economically affluent homes and those from economically disadvantage homes ("0–10%," "11–25%," "26–50%," and "more than 50%"). We used the responses to compute an indicator of *school composition by student background* where a value of 1 was assigned to "schools with more affluent than disadvantaged students," 2 to "schools with neither more affluent nor more disadvantaged students," and 3 to "schools with more disadvantaged than affluent students".

Scaling procedures

Review of item statistics

Before scaling questionnaire items, we reviewed reliabilities both overall and for national samples using Cronbach's alpha coefficient as an estimate of the internal consistency of each scale (Cronbach, 1951), for which values above 0.7 are typically regarded as satisfactory and values above 0.8 indicate high reliability (see for example, Nunnally & Bernstein, 1994, pp. 264–265). Apart from scale reliabilities, we also looked at the percentages of missing responses (which tended to be very low in most cases) as well as the correlations between individual items and the scale scores based on all other items in a scale (adjusted item-total correlations).

Confirmatory factor analysis

Structural equation modeling (SEM) (Kaplan, 2009) allows the confirmation of theoretically expected dimensions and, at the field-trial stage, the respecification of the expected dimensional structure.[8] When using confirmatory factor analysis, researchers 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:

$$x = \Lambda_x \xi + \delta,$$

where Λ_x is a q × k matrix of factor loadings, ξ denotes the latent variable(s), and δ is a q × 1 vector of unique error variables. The expected covariance matrix is fitted according to the theoretical factor structure.

When conducting the confirmatory factor analyses for ICCS 2016 questionnaire data, selected model-fit indices provided measures of the extent to which a particular model with an assumed a-priori structure "fitted the data." For the ICCS 2016 analysis, the assessment of model fit was primarily conducted through reviews 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).

For the purpose of assessing model fit, RMSEA values over 0.10 were rated as unacceptable, those between 0.08 and 0.1 as indicating marginally satisfactory model fit while values of 0.05 and lower indicate a close model fit (see MacCallum, Browne & Sugawara, 1996). As additional fit indices, CFI and NNFI are bound between 0 and 1. Values below 0.90 indicate a non-satisfactory model fit whereas values greater than 0.95 were interpreted as suggesting a close model fit (see Bentler & Bonnet, 1980; Hu & Bentler, 1999).

In addition to these fit indices, standardized factor loadings and the corresponding residual item variances provided further evidence of model fit for questionnaire data. Standardized factor loadings λ' 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 δ' (these provide an estimate of the unexplained proportion of variance) as:

$$\delta' = (1 - \lambda'^2)$$

The use of multidimensional models also allows an assessment of the estimated correlation(s) between latent factors which provide(s) information on the similarity of the different dimensions measured by related 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. Therefore, the ICCS analysis team relied on robust weighted least squares estimation (WLSMV) (see Muthén, du Toit, & Spisic, 1997; Flora & Curran, 2004) to estimate the confirmatory factor models. The software package used for estimation was Mplus 7 (Muthén & Muthén, 2012).

Confirmatory factor analyses were carried out for sets of conceptually related questionnaire items that measured between one or more different dimensions. This approach allowed an assessment of the measurement model and the associations between related latent factors. The scaling analyses were restricted to data from those countries that met sample participation requirements (see

8 In the initial stages of field trial analyses, we also employed exploratory factor analysis (Tucker & MacCallum, 1997; Fabrigar, Wegener, MacCallum, & Strahan, 1999) to determine item dimensionality of larger item pools.

Chapter 7 for further information). National samples of students, teachers, and schools, received weights that ensured equal representations of countries in the analyses.

In international studies, model parameters may vary across country and it may not be appropriate to assume the same factor structure for each population. To test parameter invariance, multiple-group modelling, as an extension of CFA, offers an approach to test the equivalence of measurement models across sub-samples (Little, 1997; Byrne, 2008).

When considering a model where respondents belong to different groups indexed as g = 1, 2, ..., G, the multiple-group factor model becomes:

$$x_g = \Lambda_{xg} \xi_g + \delta_g$$

A test of factorial invariance (H_Λ) where factor loadings are defined as being equal (often referred to as "metric equivalence" (Horn & McArdle, 1992) can be defined as:

$$H_\Lambda : \Lambda_1 = \Lambda_1 = \Lambda_2 = \ldots = \Lambda_g$$

Model-fit indices may then be compared across different multiple-group models, each with an increasing degree of constraints, from relaxed models with no constraints through to constrained models with largely invariant model parameters.

In this report, for all international and regional student questionnaire scales, three different multiple-group models for CFA were estimated with different levels of constraints on parameters:

(A) Unconstrained models where all parameters are estimated as country-specific (configural invariance);

(B) Models with constrained factor loadings across countries (metric invariance);

(C) Models with constraints on factor loadings and intercepts (scalar invariance).

The last model is the only one which ensures full comparability of measurement models across participating countries. When comparing model fit across the three conditions, it needs to be acknowledged that with data from large samples, as is typically the case in international large-scale assessments, even very small differences appear to be significant. This makes hypothesis testing using tests of statistical significance rather problematic. Therefore, when reviewing results it is more appropriate to focus on relative changes in model fit across the three models with different levels of constraints.

Item response modeling

In line with the scaling of test item data (see Chapter 10), item response modeling was used as an appropriate way of scaling questionnaire items. The one-parameter (Rasch) model (Rasch, 1960) for dichotomous items models the probability of selecting an item category 1 instead of 0 as:

$$P_i(\theta_n) = \frac{\exp(\theta_n - \delta_i)}{1+\exp(\theta_n - \delta_i)}$$

where $P_i(\theta_n)$ is the probability of person *n* scoring 1 on item *i*, θ_n is the estimated latent trait of person *n*, and δ_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 θ_n.

In the case of items with more than two categories (as, for example, with Likert-type items), this model can be generalized to the (Rasch) partial credit model (Masters & Wright, 1997), which takes the form of:

$$P_{x_i}(\theta_n) = \frac{\exp \sum_{j=0}^{x} (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^{h} (\theta_n - \delta_i + \tau_{ij})} \quad x_i = 0,1,\ldots,m_i$$

where $P_{xi}(\theta_n)$ denotes the probability of person *n* scoring *x* on item *i*, θ_n denotes the person's latent trait, the item parameter δ_i gives the location of the item on the latent continuum, and τ_{ij} denotes an additional step parameter for each step *j* between adjacent categories.

Weighted mean-square statistic (*infit*), statistics based on model residuals, provided a way of assessing general fit to the scaling model. The residual-based statistics in conjunction with a wide range of further item statistics provided the basis for an assessment of IRT model fit. ICCS 2016 used the ACER Conquest software package (Adams, Wu, & Wilson, 2015) for the analysis of item scaling properties and the estimation of item parameters.

The international item parameters were derived using equally weighted national datasets:

(A) *Calibration of item parameters for the student questionnaire:* This was done based on a pooled database with equally weighted national samples from 21 countries that met sample participation requirements for the student survey.

(B) *Calibration of item parameters for the teacher questionnaire:* This was done based on a pooled database with equally weighted national samples from 17 countries that met sample participation requirements for the student survey.

(C) *Calibration of item parameters for school questionnaire*: This was done based on a pooled database with equally weighted national samples from 21 countries that met sample participation requirements for the student survey.

Following the estimation of international item parameter from the calibration sample, we computed weighted likelihood estimation to obtain individual student scores. Weighted likelihood estimations are computed by minimizing the equation:

$$\sum_{i\in\Omega}\left[\left(r_x + \frac{J_n}{2I_n}\right) - \sum_{i=1}^{k} \frac{\exp(\sum_{j=0}^{x}\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i}\exp\sum_{j=0}^{h}(\theta_n - \delta_i + \tau_{ij})}\right] = 0$$

for each case *n*, where r_x is the sum score obtained from a set of *k* items with *j* steps between adjacent categories. This can be achieved by applying the Newton-Raphson method. The term $J_n / 2I_n$ (with I_n being the information function for student *n* and J_n being its derivative with respect to θ) is used as a weight function to account for the bias inherent in maximum likelihood estimation (see Warm, 1989). ACER ConQuest software allowed us to pre-calibrate item parameters in order to derive scale scores.

For scales that were new to ICCS 2016, the transformation of weighted likelihood estimates to an international metric resulted in reporting scales with an ICCS 2016 average of 50 and a standard deviation of 10 for equally weighted datasets from the countries that met sample participation requirements. This is achieved by applying the following formula:

$$\theta_n' = 50 + 10\ \frac{\theta_n - \eta_{\theta\ (ICCS16)}}{\sigma_{\theta(ICCS16)}}$$

where θ_n' are the scores in the international metric, θ_n are the original weighted likelihood estimates in logits, and $\mu_{\theta(\mathrm{ICCS16})}$ is the international mean of logit scores with equally weighted country subsamples. $\sigma_{\theta(\mathrm{ICCS16})}$ is the corresponding international standard deviation of the original weighted likelihood estimates. The means and standard deviations (Table 11.1) were used to transform the original scale scores for the international and regional student, teacher, and school questionnaires into the international metric.

Table 11.1: Transformation parameters for new ICCS 2016 questionnaire scales (means and standard deviations of original IRT logit scores)

Questionnaire	Scale	IRT logit score	
		Mean	SD
International student questionnaire	S_ABUSE	–1.82	1.26
	S_CITRESP	1.89	1.40
	S_CIVLRN	0.46	1.31
	S_CNTATT	2.08	1.79
	S_COMPART	–1.46	1.10
	S_ILLACT	–1.98	2.51
	S_INTACT	0.42	1.39
	S_LEGACT	0.04	1.78
	S_SCACT	0.53	1.67
	S_SCHPART	–0.42	1.03
	S_SOCMED	–1.39	1.14
Teacher questionnaire	T_BULSCH	–2.22	1.92
	T_CIVCLAS	–0.21	1.41
	T_PCCLIM	1.17	3.79
	T_PDACCE	0.29	2.25
	T_PDATCH	1.19	1.79
	T_PROBSC	–1.83	1.68
	T_PRPCCE	0.57	1.77
	T_STDCOM	–0.12	1.61
	T_STUDB	0.97	2.63
	T_TCHPRT	–0.20	1.65
School questionnaire	C_AVRESCOM	0.83	1.95
	C_BULACT	0.57	1.56
	C_BULSCH	–1.26	1.97
	C_COMCRI	–2.31	2.29
	C_COMETN	–2.50	1.78
	C_COMPOV	–0.93	2.51
	C_ENGAGE	1.45	1.65
	C_ENPRAC	1.01	1.53
	C_STDCOM	0.04	1.13
	C_STSBELS	2.58	2.18
	C_TCPART	–0.02	1.80
	C_TCSBELS	3.22	2.28
European student questionnaire	E_CCOOP	2.11	1.53
	E_EULRN	0.86	1.86
	E_EUNEG	–0.45	1.31
	E_EUPOS	0.78	1.40
	E_EURATT	1.74	2.12
	E_FREEMOVE	2.78	2.17
	E_INDFUT	2.54	1.96
	E_RESTMIG	–0.07	1.46
Latin American student questionnaire	L_ATTDIV	2.05	1.48
	L_ATTHS	0.91	2.23
	L_EMPCLAS	2.53	1.74
	L_DISCRIM	0.33	1.44

Linking questionnaire scale indices to ICCS 2009

The ICCS 2016 international, European and Latin American student questionnaires included a number of questions with unchanged question and item wording, and countries participating in both cycles were requested to use identical translations from the source version to those in the previous cycle. For scales based on this type of item sets, it was possible to equate the scale scores based on ICCS 2016 data to those from ICCS 2009. Equating was applied to 13 scales from the international questionnaire, two scales from the regional European questionnaire, and four scales from the Latin American regional questionnaire. For some of the scales individual items had been discarded or had been augmented by new items.[9] In these cases, the equating procedure was based only on those items which had remained unchanged.

Prior to conducting an equating of scale scores across cycles, item parameter estimates from calibrations from each cycles were reviewed with regard to differences. This review showed that differences were generally negligible across the two cycles. In one case, an item set measuring students' attitudes toward their country of residence, reviews of item parameters suggested that responses to the common (unchanged) items might have been influenced by the presence or absence of items related to students' perceptions of certain features of their country of residence (such as the functioning of the political system). In view of the changes in item composition and the differences between item parameters across the two cycles we decided not to equate this scale with the one established in ICCS 2009.

As for scales new to ICCS 2016, item parameters were calibrated using IRT (Rasch) partial credit model and we used the resulting item parameters to compute weighted likelihood estimates (as described previously in this chapter). To transform the scale scores to be on the same reporting metric as established in ICCS 2009, we used a two-step transformation process. The first step placed the ICCS 2016 logit scores on the ICCS 2009 logit metric (using linear constants denoted A1 and B1), while the second step (using linear constants denoted A2 and B2) transformed the latter to the same reporting metric as in ICCS 2009. Both transformations were combined into one set of final transformation constants, A and B, as follows:

$$B = B_1 \times B_2$$

$$A = A_2 + B_2 \times A_1$$

To compute the first set of transformation parameters we used the mean/sigma method suggested by Kolen and Brennan (2004). For both cycles, we calculated the combined item location and step parameters ($\delta_i + \tau_{ij}$) for each of the common items.[10] The transformation parameters in the first step were calculated on the means and standard deviations of these combined parameters as:

$$B_1 = \frac{SD_{ICCS09}}{SD_{ICCS16}}$$

$$A_1 = MN_{ICCS09} - \frac{SD_{ICCS09}}{SD_{ICCS16}} \times MN_{ICCS16}$$

MN_{ICCS16} and SD_{ICCS16} are the mean and standard deviation of the combined item location and step parameters from the calibration of ICCS 2016, respectively, while MN_{ICCS09} and SD_{ICCS09} are the corresponding (combined) parameters from the calibration based on ICCS 2009 data.

The second set of transformation parameters, A_2 and B_2 were based on the scale transformations established in the previous cycle for the reporting of ICCS 2009 data. These transformation

9 In all but one case, there was only one item exchanged or added. For one Latin American scale (L_ATTVIOL), two new items were added to the scale.

10 This is equivalent to using the differences between item location and set parameters ($\delta_i - \tau_{ij}$). The number of these (combined) parameters for each item is equal to the number of steps between item categories.

set the original logit scores from ICCS 2009 to their reporting metric, where 50 represents the average of equally weighted ICCS 2009 countries and 10 the corresponding standard deviation. The transformation parameters A_2 and B_2 were calculated as:

$$B_2 = 10/\boldsymbol{\delta}_{\theta(ICCS09)}$$

$$A_2 = 50 - (\boldsymbol{\mu}_{\theta(ICCS09)} \times B_2)$$

Where $\boldsymbol{\mu}_{\theta(ICCS2009)}$ is the mean and $\boldsymbol{\sigma}_{\theta(ICCS2009)}$ is the standard deviation of logit scores for equally weighted countries in ICCS 2009.

For all linked scales, the final transformation of ICCS 2016 was computed as follows:

$$\theta_n' = A + B \times \theta_n$$

Here, θ_n' are the scores in the international metric, and θ_n are the original weighted likelihood estimates in logits.

Given the (relatively minor) changes in item parameters between the two first cycles of ICCS, there is a degree of uncertainty associated with equating procedures applied to common sets of items (see Monseur & Berezner, 2007). To consider this lack of precision, we estimated the likely magnitude of error based on the differences between calibrated item location parameters in 2009 and 2016 using the following formula:

$$\sigma(linking\_error) = \sqrt{\frac{\sigma^2}{N}}$$

Here, σ^2 is the variance of the differences between item location parameters across the two cycles and N reflects the number of link items in each scale. The linking error estimates were then transformed to the reporting metric (see Table 11.2) and were taken into account when estimating changes in scale scores between the two cycles (see Chapter 12).

Table 11.2: Transformation parameters and linking error for equated questionnaire scales

Scale	Transformation components						Linking error	
	A_2	B_2	A_1	B_1	A	B	In logits	In reporting metric
S_CITCON	49.97	8.58	−0.62	0.96	44.67	8.28	0.048	0.411
S_CITEFF	50.04	6.98	−0.32	0.93	47.78	6.48	0.037	0.261
S_CITSOC	50.02	6.37	−1.31	0.88	41.66	5.62	0.067	0.428
S_ELECPART	50.85	5.08	−1.69	0.84	42.25	4.28	0.004	0.019
S_ETHRGHT	50.74	4.87	−2.30	1.00	39.53	4.87	0.051	0.250
S_GENEQL	50.46	6.60	−1.81	0.96	38.55	6.37	0.064	0.423
S_INTRUST	50.09	5.77	−0.36	0.89	47.99	5.12	0.050	0.291
S_OPDISC	50.04	8.59	−0.52	0.94	45.54	8.05	0.040	0.340
S_POLDISC	49.56	7.77	1.11	0.83	58.22	6.43	0.056	0.438
S_POLPART	49.89	5.31	0.80	0.88	54.14	4.70	0.066	0.351
S_RELINF	49.28	4.30	0.41	1.07	51.03	4.59	0.027	0.115
S_STUTREL	50.12	5.97	−1.14	0.89	43.28	5.33	0.076	0.454
S_VALPARTS	50.14	6.46	−1.52	0.83	40.30	5.37	0.036	0.235
E_IMMRGHT	50.21	5.90	−1.50	0.84	41.34	4.94	0.113	0.667
E_EUIDENT	49.87	6.49	−1.36	0.73	41.02	4.77	0.112	0.726
L_ATTCORR	50.18	6.92	0.55	0.82	53.96	5.67	0.039	0.269
L_ATTVIOL	50.05	6.08	0.88	0.92	55.39	5.58	0.056	0.341
L_AUTGOV	50.21	8.36	0.40	0.93	53.56	7.80	0.026	0.221
L_DISLAW	50.02	9.31	−0.06	0.82	49.44	7.68	0.028	0.259

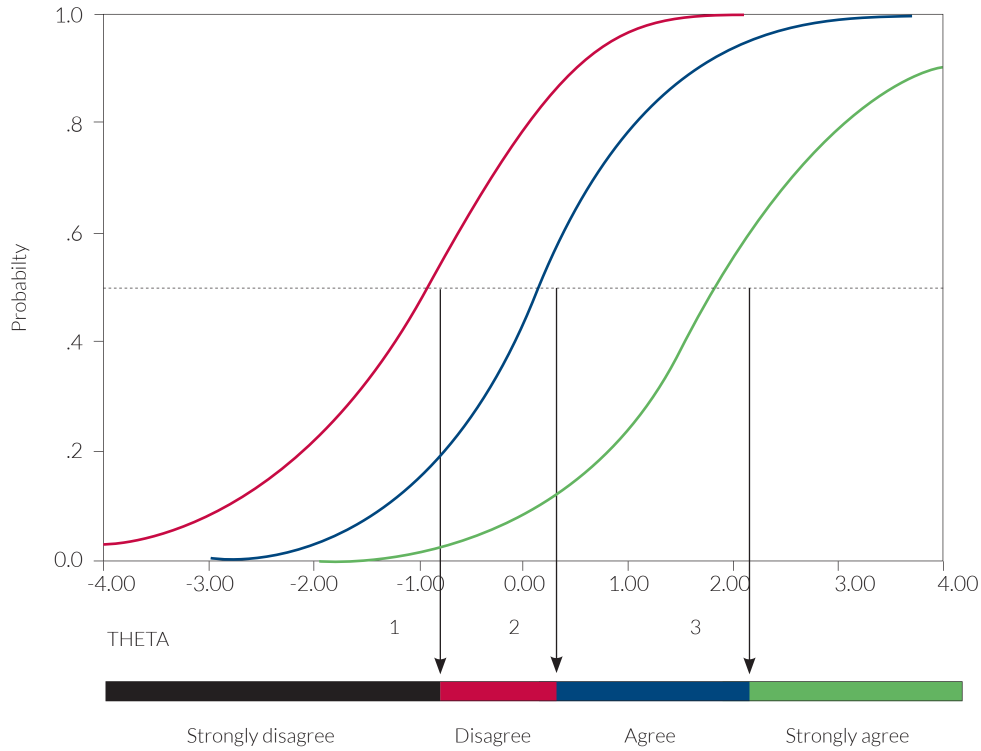
Describing questionnaire scale indices

Questionnaire scales derived from weighted likelihood estimates (logits) present values on a continuum with an ICCS 2016 (or, where equated, 2009) average of 50 and a standard deviation of 10 (for equally weighted national samples). This allows an interpretation of these scores by comparing individual scores or group average scores with the ICCS 2016 (or 2009) average. However, the individual scores do not reveal anything about the actual item responses and the extent to which respondents endorsed the items used to measure the latent variable. The scaling model used to derive individual scores allows the development of descriptions of these scales through a mapping of scale scores to (expected) item responses.[11]

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 relative to the probabilities of all lower categories. This process is equivalent to computing the odds of scoring higher than a particular category.

As an example, we plotted cumulative probabilities against scale scores for a fictitious item (Figure 11.1). The three vertical lines denote those points on the latent continuum where it becomes more likely to score > 0, > 1, or > 2. These locations Γ_k are Thurstonian thresholds which can be obtained through an iterative procedure that calculates summed probabilities for each category at each (decimal) point on the latent variable.

Figure 11.1: Summed category probabilities for a fictitious item

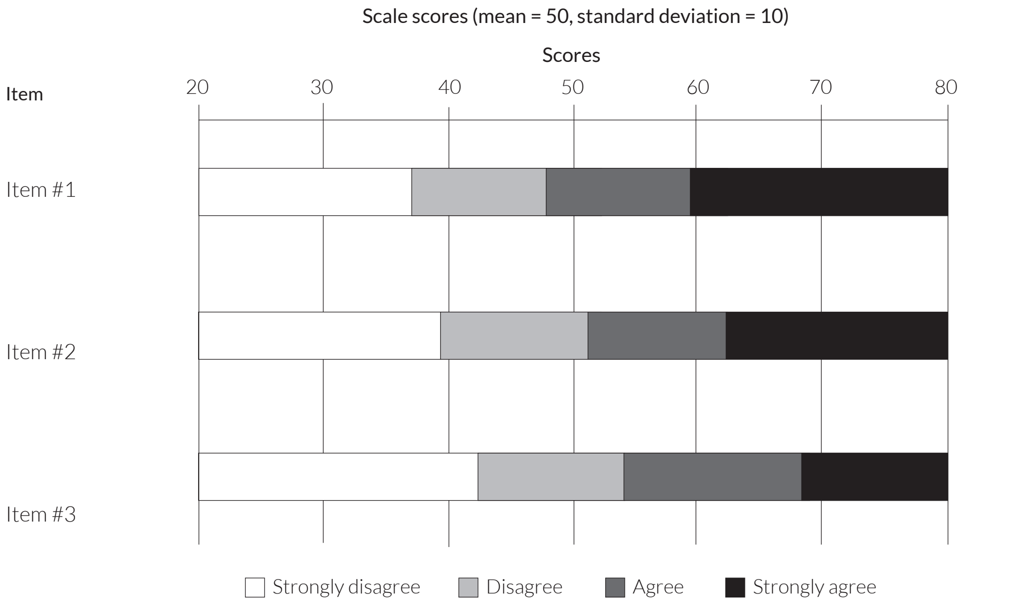


11 This approach was also used in the IEA ICCS 2009 survey (see Schulz & Friedman, 2011) and the ICILS 2013 survey (see Schulz & Friedman, 2015).

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.[12] 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 2016 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 11.2). For each item, it is thus possible to determine Thurstonian thresholds, namely 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.

Figure 11.2: 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 50% probability to strongly disagree with all three items

#2: A respondent with score 40 has more than 50% probability not to strongly disagree with items 1 and 2 but to strongly disagree with item 3

#3: A respondent with score 50 has more than 50% probability to agree with item 1 and to disagree with items 2 and 3

#4: A respondent with score 60 has more than 50% probability to strongly agree with item 1 and to at least agree with items 2 and 3

#5: A respondent with score 70 has more than 50% probability to strongly agree with items 1, 2 and 3

12 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).

This information can also be summarized by calculating the average thresholds across all items in a scale. For example, it is possible to do this for the second threshold of a four-point Likert-type scale, which allows the prediction of how likely it would be for a respondent with a certain scale score to have responses in the two lower or upper categories (on average across items). The ICCS 2016 team used this approach in the case of items measuring agreement to distinguish between scale scores for respondents who were most likely to agree or disagree with the "average item" used for measuring the respective latent trait.

In the reporting tables for questionnaire scales, we depicted national average scale scores 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 darker shaded area indicated that, on average across items, 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 lighter shaded 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

Student questionnaire

National index of students' socioeconomic background

The multivariate analyses presented in the international report (Schulz et al., 2018) include a composite index reflecting students' socioeconomic background. The *national index of students' socioeconomic background* (S_NISB) was derived from the following three indices: highest occupational status of parents (S_HISEI), highest educational level of parents (S_HISCED), and the number of books at home (S_HOMLIT). For the S_HISCED index, we collapsed the lowest two categories to have an indicator variable with four categories: lower-secondary or below, upper-secondary, tertiary non-university, and university education. The S_HOMLIT index was reduced from five to four categories (0 to 10 books; 11 to 25 books; 26 to 100 books; more than 100 books) collapsing the two highest categories. This was done for both indices on parental education and home literacy, as prior analyses had shown approximately linear associations across these categories with civic knowledge test scores and the other indicators of socioeconomic background.

In order to impute values for students who had missing data for only one of the three indicators, we used predicted values plus 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 for each national sample separately.

After converting the resulting variables including the imputed values into z-standardized variables (with a mean of 0 and a standard deviation of 1 for each national dataset), principal component analysis of these indicator variables were conducted separately for each weighted national sample.

The final S_NISB scores consists of factor scores for the first principal component with national averages of 0 and national standard deviations of 1. We calculated the factor loadings and reliabilities for each national sample (Table 11.3).

Table 11.3: Factor loadings and reliabilities for the national index of students' socioeconomic background

Country	Factor loadings for:			Cronbach's alpha
	Highest parental occupation	Highest parental education	Books at home	
Belgium (Flemish)	0.82	0.81	0.68	0.65
Bulgaria	0.81	0.86	0.78	0.75
Chile	0.85	0.86	0.61	0.67
Chinese Taipei	0.79	0.82	0.73	0.68
Colombia	0.80	0.80	0.67	0.63
Croatia	0.83	0.85	0.70	0.71
Denmark	0.84	0.83	0.66	0.67
Dominican Republic	0.80	0.81	0.59	0.58
Estonia	0.82	0.82	0.64	0.64
Finland	0.83	0.84	0.59	0.63
Hong Kong SAR	0.80	0.83	0.70	0.67
Italy	0.83	0.83	0.68	0.67
Korea, Republic of	0.67	0.71	0.71	0.47
Latvia	0.81	0.83	0.66	0.65
Lithuania	0.81	0.81	0.71	0.67
Malta	0.81	0.78	0.65	0.60
Mexico	0.85	0.87	0.61	0.68
Netherlands	0.77	0.77	0.70	0.61
North-Rhine Westphalia (Germany)*	0.78	0.73	0.70	0.58
Norway	0.78	0.78	0.69	0.61
Peru	0.84	0.84	0.64	0.68
Russian Federation	0.81	0.81	0.62	0.61
Slovenia	0.83	0.82	0.71	0.68
Sweden	0.79	0.76	0.72	0.62
ICCS 2016 average	0.81	0.81	0.67	0.64

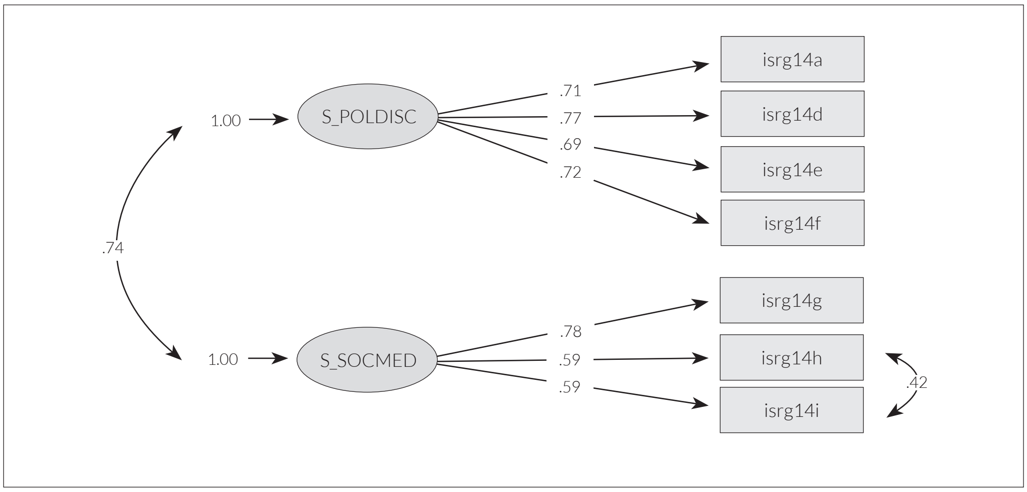
Notes:
*Benchmarking participant.
The ICCS 2016 average based on data from countries meeting IEA sample participation requirements.

Students' participation in out-of-school activities

Question 14 of the student questionnaire required students to indicate their level of involvement in a series of activities outside of school. Students reported their frequency of undertaking the following activities using the response categories "Never or hardly ever," "Monthly (at least once a month)," "Weekly (at least once a week)," and "Daily or almost daily".

Two scales were derived from this question. Four of the items were used to derive a scale reflecting *students' discussion of political and social issues outside of school* (S_POLDISC) and another three items were used to derive a scale of *students' engagement with social media* (S_SOCMED). Higher scale scores on both scales reflect more frequent engagement in each type of these activities.

Figure 11.3: Confirmatory factor analysis of items measuring students' participation in out-of-school activities



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.086	0.092	0.094	0.084
CFI	0.96	0.97	0.95	0.93
TLI	0.94	0.94	0.94	0.95

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' participation in out-of-school activities (Figure 11.3). After including a correlation between residuals of two of the items measuring S_SOCMED (items *h* and *i*, which reflected highly similar content), the model fit was marginally satisfactory and the two latent factors (S_POLDISC and S_SOCMED) were highly correlated ($r = 0.74$). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit changed only marginally which indicates a relatively high degree of invariance for this model.

The average reliabilities (Cronbach's alpha) across national samples were 0.74 for S_POLDISC (ranging from 0.65 to 0.81) and 0.63 for S_SOCMED (ranging from 0.47 to 0.78) (see Table 11.4). These scales were based on selected item parameters (Table 11.5). The scale scores for S_POLDISC were linked to the scale established in the ICCS 2009 so that they were comparable across the two cycles.

Table 11.4: Reliabilities for scales measuring students' participation in out-of-school activities

Country	Scale reliability (Cronbach's alpha)	
	S_POLDISC	S_SOCMED
Belgium (Flemish)	0.75	0.62
Bulgaria	0.70	0.67
Chile	0.74	0.78
Chinese Taipei	0.77	0.64
Colombia	0.65	0.63
Croatia	0.74	0.47
Denmark	0.79	0.57
Dominican Republic	0.67	0.66
Estonia	0.76	0.60
Finland	0.80	0.62
Hong Kong SAR	0.81	0.78
Italy	0.66	0.59
Korea, Republic of	0.78	0.70
Latvia	0.79	0.76
Lithuania	0.76	0.60
Malta	0.69	0.63
Mexico	0.68	0.70
Netherlands	0.75	0.64
North-Rhine Westphalia (Germany)*	0.75	0.64
Norway	0.79	0.63
Peru	0.67	0.65
Russian Federation	0.75	0.63
Slovenia	0.73	0.49
Sweden	0.81	0.67
ICCS 2016 average	0.74	0.63

Note:
*Benchmarking participant.

Table 11.5: Item parameters for scales measuring students' participation in out-of-school activities

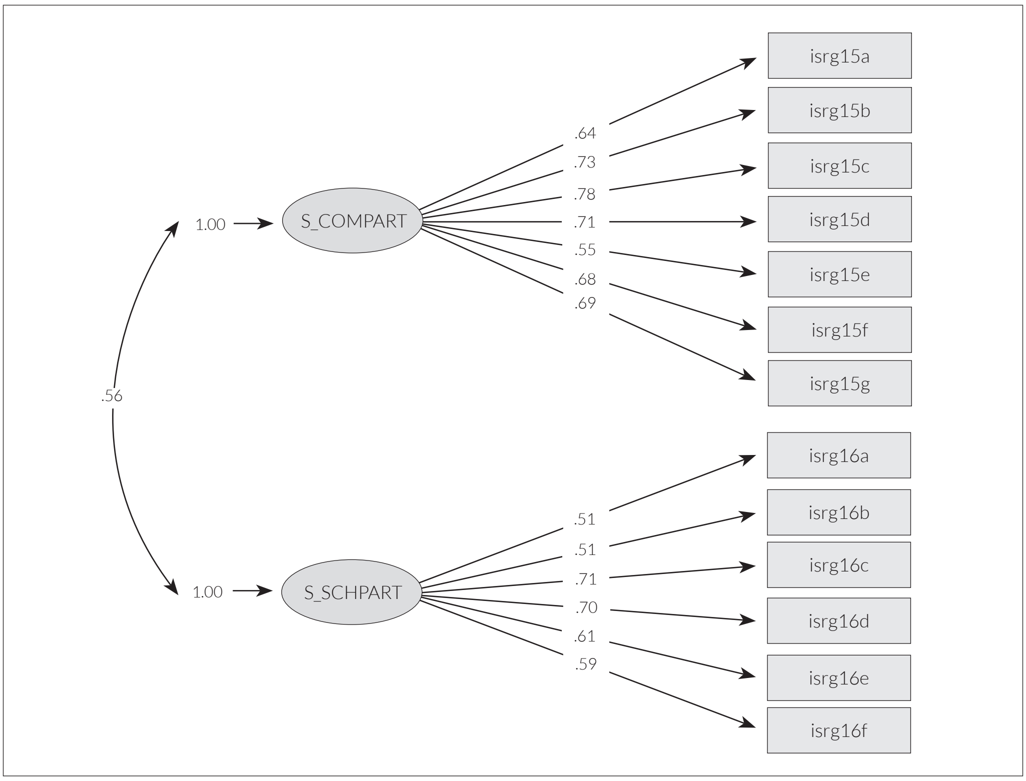
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_POLDISC	*How often are you involved in each of the following activities?*				
IS3G14A	Talking with your parent(s) about political or social issues	0.16	–0.93	–0.09	1.02
IS3G14D	Talking with friends about political or social issues	0.64	–0.83	–0.06	0.89
IS3G14E	Talking with your parent(s) about what is happening in other countries	–0.78	–1.55	–0.08	1.63
IS3G14F	Talking with friends about what is happening in other countries	–0.02	–1.38	0.00	1.38
S_SOCMED	*How often are you involved in each of the following activities?*				
IS3G14G	Using the internet to find information about political or social issues	–0.86	–0.81	–0.06	0.87
IS3G14H	Posting a comment or image regarding a political or social issue on the internet or social media	0.47	–0.01	–0.21	0.22
IS3G14I	Sharing or commenting on another person's online post regarding a political or social issue	0.38	–0.19	–0.14	0.33

Students' civic participation in the community and at school

Question 15 asked students to state whether they had participated in ten different organizations, clubs, or groups in the wider community either "within the last 12 months," "more than year ago," or "never." We used seven of these items to derive a scale reflecting students' civic participation in the wider community (S_COMPART) where positive values on this scale reflect higher levels of civic participation.

Question 16 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 allowed us to derive a scale reflecting students' civic participation at school (S_SCHPART), with the positive values reflecting higher levels of civic participation.

Figure 11.4: Confirmatory factor analysis of items measuring students' civic participation in the community and at school



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.047	0.043	0.073	0.078
CFI	0.95	0.95	0.84	0.79
TLI	0.93	0.94	0.83	0.81

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.4) confirmed that the model fit was satisfactory and the two latent dimensions (S_COMPART and S_SCHPART) were highly correlated (r = 0.56). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for the more constrained models, which suggests a certain degree of measurement variance for this model across countries. However, there was still an acceptable fit for the most constrained model.

The scale reliability (Cronbach's alpha) for S_COMPART was 0.70 on average with coefficients ranging from 0.63 to 0.80 (see Table 11.6). For S_SCHPART the average reliability was 0.67 with coefficient ranging from 0.54 to 0.83. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.7). Given modifications to the two questions since ICCS 2009, the ICCS 2016 scale scores were not equated and are therefore not comparable to those from the previous cycle.

Table 11.6: Reliabilities for scales measuring students' civic participation in the community and at school

Country	Scale reliability (Cronbach's alpha)	
	S_COMPART	S_SCHPART
Belgium (Flemish)	0.66	0.70
Bulgaria	0.75	0.72
Chile	0.80	0.73
Chinese Taipei	0.69	0.72
Colombia	0.71	0.63
Croatia	0.63	0.55
Denmark	0.67	0.65
Dominican Republic	0.74	0.68
Estonia	0.67	0.69
Finland	0.66	0.66
Hong Kong SAR	0.79	0.75
Italy	0.67	0.54
Korea, Republic of	0.75	0.83
Latvia	0.66	0.71
Lithuania	0.74	0.72
Malta	0.75	0.69
Mexico	0.78	0.71
Netherlands	0.63	0.64
North-Rhine Westphalia (Germany)*	0.68	0.65
Norway	0.67	0.69
Peru	0.74	0.65
Russian Federation	0.75	0.71
Slovenia	0.70	0.69
Sweden	0.72	0.68
ICCS 2016 average	0.70	0.67
Minimum value	0.63	0.54
Maximum value	0.80	0.83

Note:
*Benchmarking participant.

Table 11.7: Item parameters for scale measuring students' civic participation in the community and at school

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_COMPART	*Have you ever been involved in activities of any of the following organizations, clubs or groups?*				
IS3G15A	A youth organization affiliated with a political party or union	0.68	0.87	–0.87	n/a
IS3G15B	An environmental action group or organization	–0.05	–0.28	0.28	n/a
IS3G15C	A Human Rights organization	0.47	0.40	–0.40	n/a
IS3G15D	A voluntary group doing something to help the community	–0.47	–0.20	0.20	n/a
IS3G15E	An organization collecting money for a social cause	–0.62	–0.16	0.16	n/a
IS3G15F	A group of young people campaigning for an issue	–0.02	0.20	–0.20	n/a
IS3G15G	An animal rights or animal welfare group	0.02	0.08	–0.08	n/a
S_SCHPART	*At school, have you ever done any of the following activities?*				
IS3G16A	Active participation in an organized debate	–0.23	0.18	–0.18	n/a
IS3G16B	Voting for <class representative> or <school parliament>	–0.98	0.02	–0.02	n/a
IS3G16C	Taking part in decision-making about how the school is run	0.37	0.13	–0.13	n/a
IS3G16D	Taking part in discussions at a <student assembly>	0.40	0.21	–0.21	n/a
IS3G16E	Becoming a candidate for <class representative> or <school parliament>	0.31	0.12	–0.12	n/a
IS3G16F	Participating in an activity to make the school more <environmentally friendly> (e.g. through water-saving or recycling)	0.13	–0.08	0.08	n/a

Note:
n/a = not applicable.

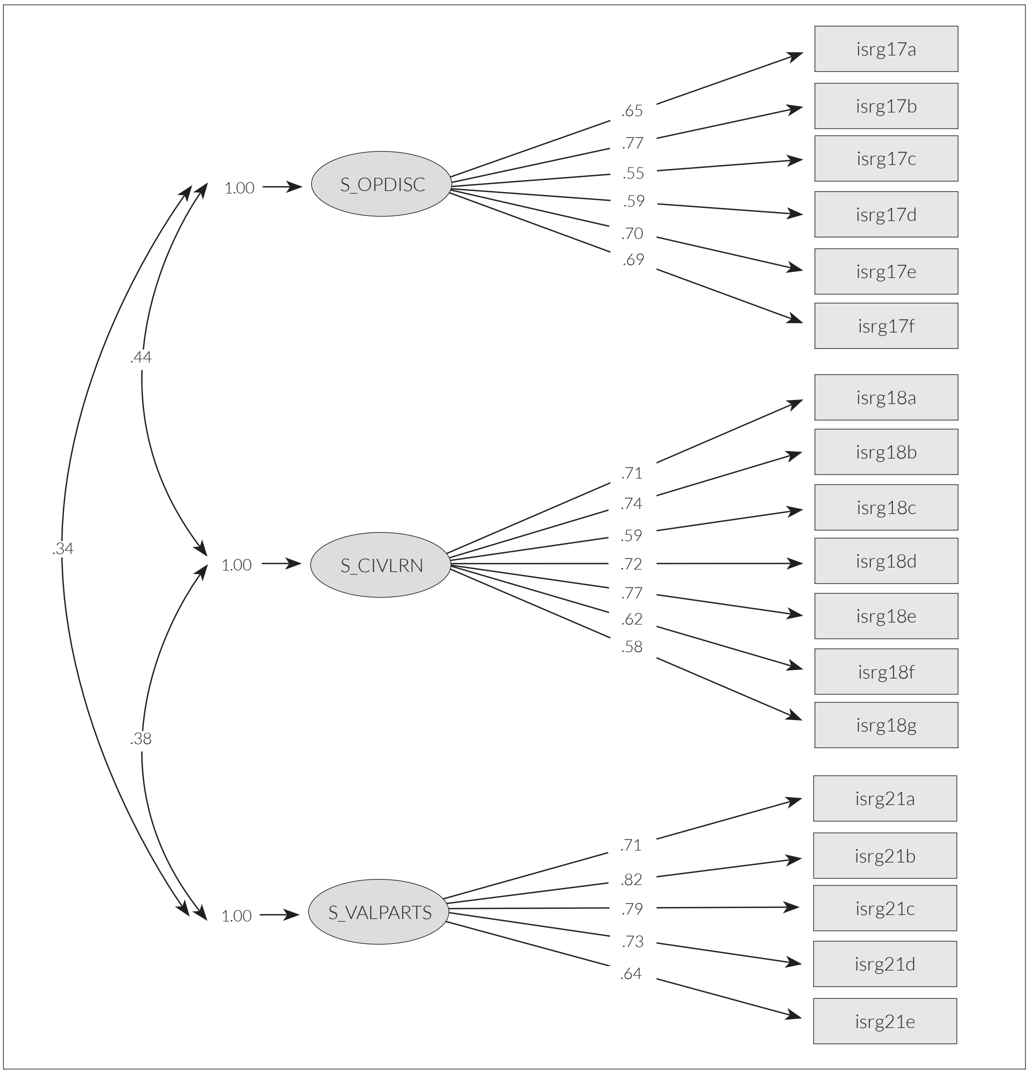
Students' perceptions of civic learning and participation at school

Question 17, which was unchanged from ICCS 2009, asked students how frequently ("never," "rarely," "sometimes," "often") situations happened during regular lessons regarding the discussion of political and social issues indicating the openness of the classroom climate for these activities. We used six of the question items to derive the scale reflecting *students' perceptions of openness in classroom discussions* (S_OPDISC). The higher values on the scale reflect increased perceptions of classroom discussions of political and social issues.

Question 18, which was new to ICCS 2016, inquired about the extent ("to a large extent," "to a moderate extent," "to a small extent," "not at all") to which students thought they had learned about seven different topics. The seven items were used to derive a scale reflecting *students' reports on civic learning at school* (S_CIVLRN) with higher values indicating higher levels of civic learning 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. While four items were identical to those administered in ICCS 2009, one item was new to ICCS 2016. We included all five question items in the scale *students' perceptions of the value of participation at school* (S_VALPARTS). Higher scores on this scale correspond to a higher extent of agreement with statements about the value of participation at school.

Figure 11.5: Confirmatory factor analysis of items measuring students' perceptions of civic learning and participation at school



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.042	0.045	0.050	0.069
CFI	0.96	0.97	0.95	0.89
TLI	0.96	0.96	0.95	0.91

The results from confirmatory factor analyses assume a three-dimensional model for the (scaled) items from these three questions (Figure 11.5). The model fit was satisfactory and the three latent dimensions were moderately correlated with r = 0.44 between S_OPDISC and S_CIVLRN, r = 0.34 between S_OPDISC and S_VALPARTS, and r = 0.38 between S_CIVLRN and S_VALPARTS. When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for the more constrained models but model fit was still in an acceptable range for the most constrained (scalar) model.

The reliabilities (Cronbach's alpha) were satisfactory for the three scales (Table 11.8): for S_OPDISC we observed an average reliability of 0.77 (ranging from 0.66 to 0.90), for S_CIVLRN of 0.80 (ranging from 0.71 to 0.89), and for S_VALPARTS of 0.78 (ranging from 0.70 to 0.91).

Selected item parameters were used to scale the items corresponding to these three scales (Table 11.9). Given that question 17 was identical and for question 19 only one item had been exchanged, the scale scores for S_OPDISC and S_VALPARTS were equated and are therefore comparable with those from the previous cycle.

Table 11.8: Reliabilities for scales measuring students' perceptions of civic learning and participation at school

Country	Scale reliability (Cronbach's alpha)		
	S_OPDISC	S_CIVLRN	S_VALPARTS
Belgium (Flemish)	0.74	0.77	0.70
Bulgaria	0.76	0.80	0.77
Chile	0.86	0.88	0.90
Chinese Taipei	0.84	0.87	0.87
Colombia	0.73	0.75	0.73
Croatia	0.78	0.82	0.76
Denmark	0.78	0.75	0.77
Dominican Republic	0.75	0.71	0.71
Estonia	0.75	0.80	0.82
Finland	0.77	0.84	0.85
Hong Kong SAR	0.88	0.89	0.89
Italy	0.70	0.76	0.70
Korea, Republic of	0.90	0.89	0.91
Latvia	0.73	0.80	0.77
Lithuania	0.80	0.82	0.79
Malta	0.73	0.80	0.76
Mexico	0.77	0.81	0.79
Netherlands	0.76	0.82	0.78
North-Rhine Westphalia (Germany)*	0.76	0.77	0.72
Norway	0.83	0.83	0.79
Peru	0.66	0.74	0.71
Russian Federation	0.81	0.83	0.81
Slovenia	0.78	0.78	0.78
Sweden	0.81	0.82	0.82
ICCS 2016 average	0.77	0.80	0.78
Minimum value	0.66	0.71	0.70
Maximum value	0.90	0.89	0.91

Note:
*Benchmarking participant.

Table 11.9: Item parameters for scale measuring students' perceptions of civic learning and participation at school

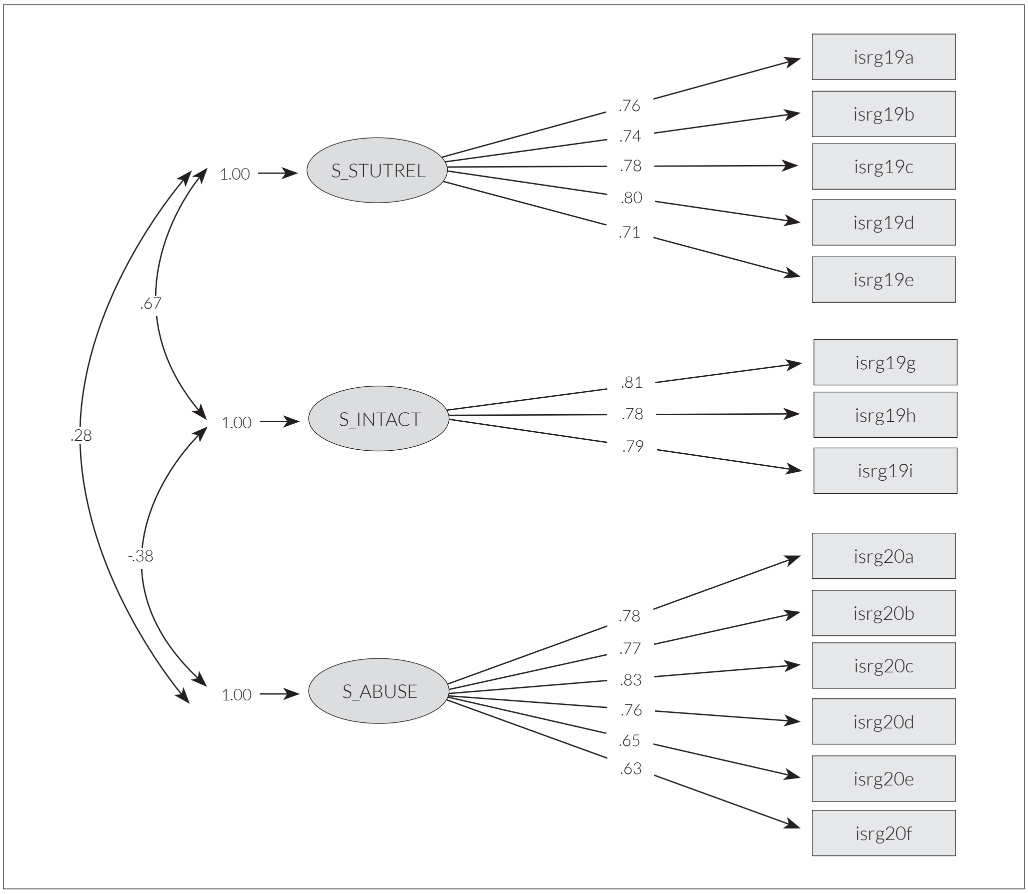
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_OPDISC	*When discussing political or social issues during regular lessons, how often do the following things happen?*				
IS3G17A	Teachers encourage students to make up their own minds	–0.18	–0.84	–0.44	1.28
IS3G17B	Teachers encourage students to express their opinions	–0.73	–0.64	–0.32	0.96
IS3G17C	Students bring up current political events for discussion in class	0.89	–1.50	–0.04	1.54
IS3G17D	Students express opinions in class even when their opinions are different from most of the other students	–0.26	–1.21	–0.20	1.41
IS3G17E	Teachers encourage students to discuss the issues with people having different opinions	0.39	–1.03	–0.22	1.26
IS3G17F	Teachers present several sides of the issues when explaining them in class	–0.11	–1.00	–0.28	1.28
S_CIVLRN	*At school, to what extent have you learned about the following topics?*				
IS3G18A	How citizens can vote in local or national elections	0.01	–1.37	–0.18	1.56
IS3G18B	How laws are introduced and changed in <country of test>	0.14	–1.50	–0.09	1.59
IS3G18C	How to protect the environment (e.g. through energy-saving or recycling)	–0.97	–1.34	0.03	1.31
IS3G18D	How to contribute to solving problems in the <local community>	0.25	–1.58	0.03	1.55
IS3G18E	How citizen rights are protected in <country of test>	0.02	–1.31	–0.01	1.32
IS3G18F	Political issues and events in other countries	0.37	–1.65	0.00	1.65
IS3G18G	How the economy works	0.18	–1.39	0.00	1.38
S_VALPARTS	*How much do you agree or disagree with the following statements about student participation at school?*				
IS3G21A	Student participation in how schools are run can make schools better	–0.06	–1.86	–0.96	2.82
IS3G21B	Lots of positive changes can happen in schools when students work together	–0.44	–1.73	–1.07	2.80
IS3G21C	Organizing groups of students to express their opinions could help solve problems in schools	–0.03	–2.19	–0.80	2.99
IS3G21D	Students can have more influence on what happens in schools if they act together rather than alone	–0.06	–1.85	–0.89	2.74
IS3G21E	Voting in student elections can make a difference to what happens at schools	0.59	–2.01	–0.61	2.63

Students' perceptions of school climate and interactions

Question 19 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." We used five of the ten items from ICCS 2009 to derive the scale students' perceptions of student–teacher relations at school (S_STUTREL), while another three new ICCS 2016 items were used to measure students' perceptions of student interaction at school (S_INTACT). For both scales, higher values reflected more positive perceptions of interactions at school.

Question 20 asked students to indicate how often they themselves had experienced situations of physical and/or verbal abuse during the last three months ("not at all," "once," "two to four times," "five times or more"). We used the six question items to form the scale *students' experiences of physical and verbal abuse at school* (S_ABUSE). Higher values on this scale reflect more frequent experiences with physical and verbal aggression at school.

Figure 11.6: Confirmatory factor analysis of items measuring students' perceptions of school climate and interactions



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.052	0.052	0.053	0.057
CFI	0.97	0.97	0.97	0.95
TLI	0.96	0.97	0.96	0.96

The results from confirmatory factor analyses assume a three-dimensional model for the (scaled) items from these three questions (Figure 11.6). The model fit was satisfactory and there was a strong positive correlation between S_STUTREL and S_INTACT with r = 0.67. S_ABUSE has moderate negative correlations with S_STUTREL (r = –0.28) and S_INTACT (r = –0.38). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was only slightly less satisfactory for the most constrained model suggesting a relatively high degree of measurement invariance.

The reliabilities (Cronbach's alpha) were satisfactory for the three scales (Table 11.10): for S_STUTREL we observed an average reliability of 0.81 (ranging from 0.72 to 0.90), for S_INTACT of 0.76 (ranging from 0.71 to 0.85), and for S_ABUSE of 0.75 (ranging from 0.68 to 0.80).

Selected item parameters were used to scale the items corresponding to these three scales (Table 11.11). Given that the items measuring S_STUTREL (both stem and item wording) were identical with those used in ICCS 2009, the scale was equated so that its scores are comparable with those from the previous cycle.

Table 11.10: Reliabilities for scales measuring students' perceptions of school climate and interactions

Country	Scale reliability (Cronbach's alpha)		
	S_STUTREL	S_INTACT	S_ABUSE
Belgium (Flemish)	0.78	0.74	0.73
Bulgaria	0.78	0.75	0.78
Chile	0.86	0.82	0.79
Chinese Taipei	0.90	0.83	0.72
Colombia	0.78	0.75	0.71
Croatia	0.82	0.71	0.77
Denmark	0.84	0.79	0.74
Dominican Republic	0.74	0.72	0.75
Estonia	0.80	0.75	0.74
Finland	0.83	0.77	0.76
Hong Kong SAR	0.89	0.85	0.78
Italy	0.79	0.72	0.70
Korea, Republic of	0.88	0.85	0.68
Latvia	0.78	0.73	0.76
Lithuania	0.82	0.72	0.77
Malta	0.83	0.75	0.79
Mexico	0.79	0.75	0.76
Netherlands	0.78	0.75	0.71
North-Rhine Westphalia (Germany)*	0.80	0.74	0.74
Norway	0.87	0.82	0.80
Peru	0.72	0.73	0.73
Russian Federation	0.78	0.76	0.76
Slovenia	0.79	0.72	0.76
Sweden	0.85	0.81	0.76
ICCS 2016 average	0.81	0.76	0.75
Minimum value	0.72	0.71	0.68
Maximum value	0.90	0.85	0.80

Note:
*Benchmarking participant.

Table 11.11: Item parameters for scales measuring students' perceptions of school climate and interactions

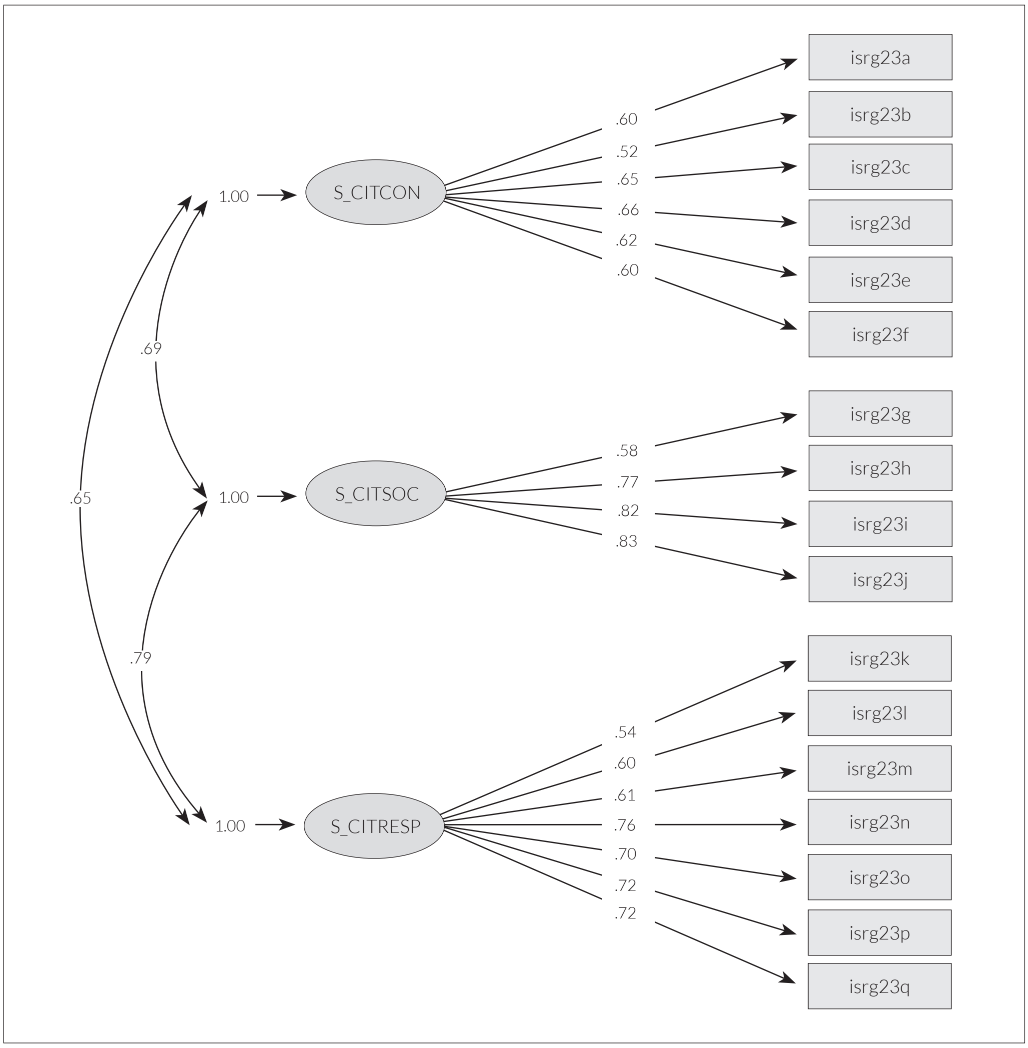
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_STUTREL	*How much do you agree or disagree with the following statements about teachers and students at your school?*				
IS3G19A	Most of my teachers treat me fairly	–0.23	–1.90	–0.78	2.69
IS3G19B	Students get along well with most teachers	0.48	–2.81	–0.31	3.13
IS3G19C	Most teachers are interested in students' well-being	–0.03	–2.18	–0.58	2.76
IS3G19D	Most of my teachers listen to what I have to say	0.08	–2.31	–0.48	2.79
IS3G19E	If I need extra help. I receive it from my teachers	–0.30	–1.96	–0.79	2.75
S_INTACT	*How much do you agree or disagree with the following statements about teachers and students at your school?*				
IS3G19G	Most students at my school treat each other with respect	0.42	–2.62	–0.38	3.00
IS3G19H	Most students at my school get along well with each other	–0.14	–2.72	–0.65	3.37
IS3G19I	My school is a place where students feel safe	–0.28	–2.44	–0.77	3.21
S_ABUSE	*During the last three months. how often did you experience the following situations at your school?*				
IS3G20A	A student called you by an offensive nickname	–1.16	–0.64	0.21	0.43
IS3G20B	A student said things about you to make others laugh	–1.09	–0.81	0.04	0.77
IS3G20C	A student threatened to hurt you	0.30	0.00	–0.18	0.18
IS3G20D	You were physically attacked by another student	0.50	–0.05	–0.05	0.11
IS3G20E	A student broke something belonging to you on purpose	0.53	–0.52	0.16	0.36
IS3G20F	A student posted offensive pictures or text about you on the Internet	0.92	0.24	–0.17	–0.06

Students' perceptions of good citizenship behaviors

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 three scales that we derived from this question, and which are included in the student database, are:

- Students' perceptions of the importance of conventional citizenship (S_CITCON) based on six ICCS 2009 items;
- Students' perceptions of the importance of social-movement-related citizenship (S_CITSOC) based on four ICCS 2009 items;
- Students' perceptions of the importance of personal responsibility for citizenship (S_CITRESP) based on two ICCS 2009 and five newly developed items.

Figure 11.7: Confirmatory factor analysis of items measuring students' perceptions of the importance of citizenship behaviors



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.069	0.077	0.071	0.081
CFI	0.91	0.91	0.91	0.85
TLI	0.89	0.89	0.90	0.88

The results from confirmatory factor analyses assume a three-dimensional model for the (scaled) items from these three questions (Figure 11.7). The model fit was satisfactory and the three latent dimensions were strongly correlated with r = 0.69 between S_CITCON and S_CITSOC, r = 0.65 between S_CITCON and S_CITRESP, and r = 0.79 between S_CITSOC and S_CITRESP. When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was notably less satisfactory for the most constrained model indicating a certain lack of measurement invariance. This finding suggests that comparisons of scale scores across participating countries should be interpreted with some caution.

The reliabilities (Cronbach's alpha) were mostly satisfactory for the three scales (Table 11.12): for S_CITCON we observed an average reliability of 0.71 (ranging from 0.60 to 0.82), for S_CITSOC of 0.74 (ranging from 0.60 to 0.88), and for S_CITRESP of 0.78 (ranging from 0.71 to 0.88).

Selected item parameters were used to scale the items corresponding to these three scales (Table 11.13). Given that all items measuring S_CITCON and S_CITSOC (both stem and item wording) were identical with those used in ICCS 2009, both scales were equated so that its scores are comparable with those from the previous cycle.

Table 11.12: Reliabilities for scales measuring students' perceptions of the importance of citizenship behaviors

Country	Scale reliability (Cronbach's alpha)		
	S_CITCON	S_CITSOC	S_CITRESP
Belgium (Flemish)	0.66	0.74	0.74
Bulgaria	0.71	0.76	0.81
Chile	0.82	0.85	0.88
Chinese Taipei	0.78	0.81	0.84
Colombia	0.65	0.65	0.72
Croatia	0.72	0.70	0.82
Denmark	0.66	0.81	0.71
Dominican Republic	0.60	0.60	0.82
Estonia	0.71	0.74	0.71
Finland	0.74	0.77	0.80
Hong Kong SAR	0.80	0.84	0.87
Italy	0.65	0.71	0.72
Korea, Republic of	0.81	0.88	0.88
Latvia	0.69	0.71	0.73
Lithuania	0.72	0.75	0.77
Malta	0.72	0.76	0.82
Mexico	0.75	0.75	0.82
Netherlands	0.69	0.72	0.76
North-Rhine Westphalia (Germany)*	0.70	0.73	0.74
Norway	0.74	0.79	0.79
Peru	0.64	0.66	0.74
Russian Federation	0.78	0.77	0.76
Slovenia	0.70	0.72	0.78
Sweden	0.73	0.81	0.77
ICCS 2016 average	0.71	0.74	0.78
Minimum value	0.60	0.60	0.71
Maximum value	0.82	0.88	0.88

Note:
*Benchmarking participant.

Table 11.13: Item parameters for scales measuring students' perceptions of the importance of citizenship behaviors

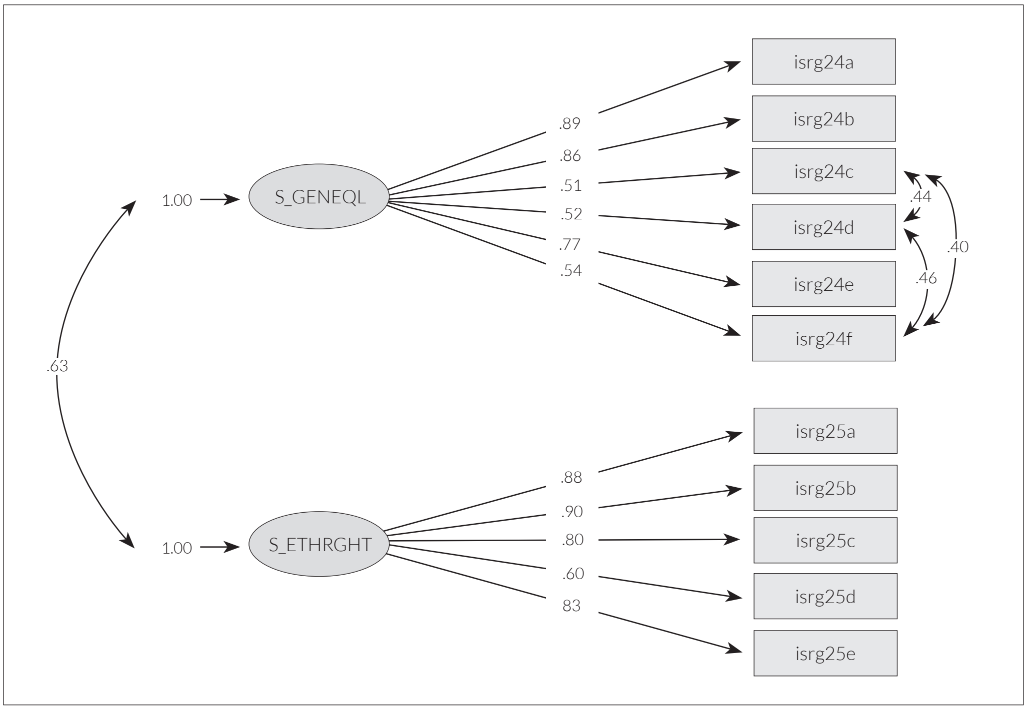
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_CITCON	*How important are the following behaviors for being a good adult citizen?*				
IS3G23A	Voting in every national election	–0.58	–1.49	–0.05	1.54
IS3G23B	Joining a political party	1.06	–2.13	0.70	1.43
IS3G23C	Learning about the country's history	–0.45	–1.40	–0.06	1.45
IS3G23D	Following political issues in the newspaper, on the radio, on TV or on the Internet	–0.24	–1.66	–0.19	1.85
IS3G23E	Showing respect for government representatives	–0.51	–1.29	–0.47	1.75
IS3G23F	Engaging in political discussions	0.73	–2.13	0.40	1.73
S_CITSOC	*How important are the following behaviors for being a good adult citizen?*				
IS3G23G	Participating in peaceful protests against laws believed to be unjust	0.89	–2.10	0.09	2.02
IS3G23H	Participating in activities to benefit people in the <local community>	–0.13	–2.14	–0.29	2.43
IS3G23I	Taking part in activities promoting human rights	–0.31	–2.04	–0.26	2.30
IS3G23J	Taking part in activities to protect the environment	–0.44	–1.84	–0.30	2.14
S_CITRESP	*How important are the following behaviors for being a good adult citizen?*				
IS3G23K	Working hard	0.33	–1.47	–0.38	1.85
IS3G23L	Always obeying the law	–0.13	–0.95	–0.49	1.45
IS3G23M	Ensuring the economic welfare of their families	–0.34	–0.98	–0.66	1.64
IS3G23N	Making personal efforts to protect natural resources (e.g., through saving water or recycling waste)	0.00	–1.37	–0.45	1.81
IS3G23O	Respecting the rights of others to have their own opinions	–0.42	–0.88	–0.73	1.61
IS3G23P	Supporting people who are worse off than you	0.01	–1.13	–0.70	1.83
IS3G23Q	Engaging in activities to help people in less developed countries	0.54	–1.67	–0.33	2.00

Students' endorsement of equal rights and opportunities

Question 24 presented a series of seven ICCS 2009 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. We used the first six question items to form the scale *students' attitudes toward gender equality* (S_GENEQL) where higher values reflect stronger agreement with the notion of gender equality or stronger disagreement with negative views of gender equality.

Question 25 contained a series of statements from ICCS 2009 regarding the rights and responsibilities of all 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. We used all five question items to construct the *students' attitudes toward equal rights for all ethnic/racial groups* (S_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.

Figure 11.8: Confirmatory factor analysis of items measuring students' endorsement of equal rights and opportunities



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.048	0.053	0.049	0.064
CFI	0.99	0.99	0.99	0.97
TLI	0.98	0.98	0.99	0.98

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.8). The model fit was poor for the original model, but was satisfactory after controlling for the common residual variance between negatively worded statements on gender equality. The two latent dimensions were highly correlated (r = 0.63). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for the most constrained models, which suggests a certain degree of measurement variance for this model across countries. However, model fit for the most constrained (scalar) model was still within an acceptable range.

The scale reliability (Cronbach's alpha) for S_GENEQL was mostly satisfactory, being 0.77 on average with coefficients ranging from 0.58 to 0.86 (see Table 11.14). For S_ETHRGHT the average reliability was 0.82 with coefficients ranging from 0.66 to 0.93. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.15). Given that the two questions were identical to those administered in ICCS 2009, both scales were equated so that scale scores are comparable to those from the previous cycle.

Table 11.14: Reliabilities for scales measuring students' endorsement of equal rights and opportunities

Country	Scale reliability (Cronbach's alpha)	
	S_GENEQL	S_ETHRGHT
Belgium (Flemish)	0.77	0.79
Bulgaria	0.74	0.83
Chile	0.77	0.89
Chinese Taipei	0.80	0.83
Colombia	0.73	0.76
Croatia	0.80	0.83
Denmark	0.86	0.85
Dominican Republic	0.60	0.66
Estonia	0.81	0.81
Finland	0.86	0.89
Hong Kong SAR	0.81	0.93
Italy	0.79	0.83
Korea, Republic of	0.73	0.89
Latvia	0.78	0.77
Lithuania	0.78	0.82
Malta	0.79	0.79
Mexico	0.58	0.80
Netherlands	0.80	0.83
North-Rhine Westphalia (Germany)*	0.83	0.85
Norway	0.83	0.90
Peru	0.73	0.73
Russian Federation	0.73	0.86
Slovenia	0.83	0.82
Sweden	0.84	0.90
ICCS 2016 average	0.77	0.82
Minimum value	0.58	0.66
Maximum value	0.86	0.93

Note:
*Benchmarking participant.

Table 11.15: Item parameters for scales measuring students' endorsement of equal rights and opportunities

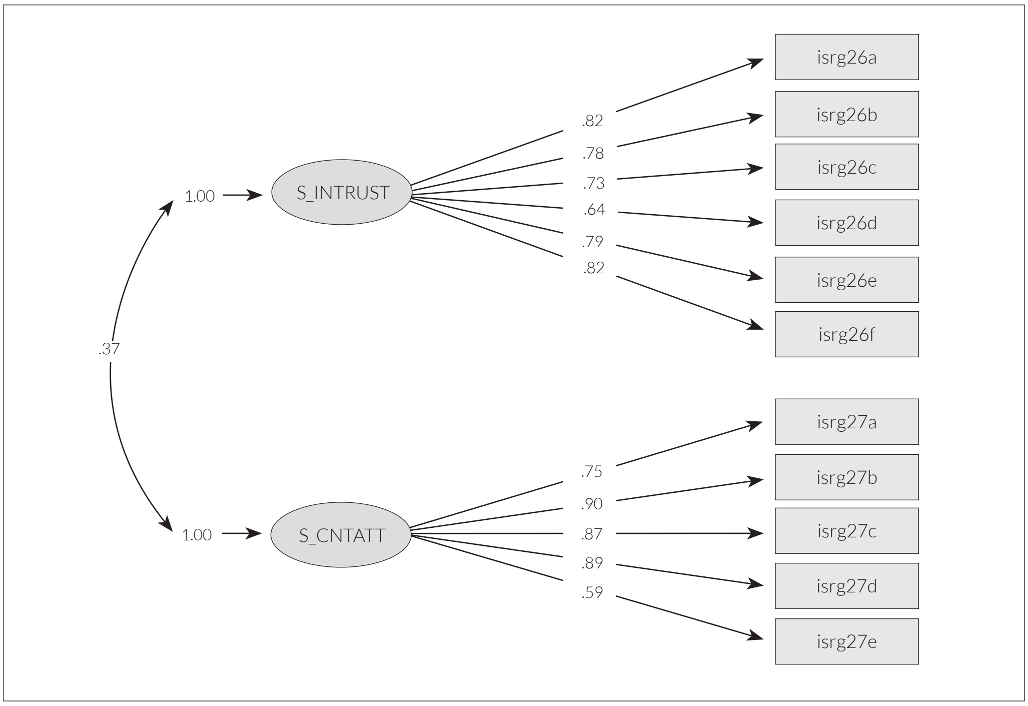
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_GENEQL	*How much do you agree or disagree with the following statements?*				
IS3G24A	Men and women should have equal opportunities to take part in government	–0.89	–0.49	–0.82	1.30
IS3G24B	Men and women should have the same rights in every way	–0.70	–1.12	–0.18	1.30
IS3G24C	Women should stay out of politics	0.41	–0.59	–0.61	1.20
IS3G24D	When there are not many jobs available, men should have more right to a job than women	0.74	–0.69	–0.41	1.10
IS3G24E	Men and women should get equal pay when they are doing the same jobs	–0.41	–0.69	–0.33	1.01
IS3G24F	Men are better qualified to be political leaders than women	0.85	–0.92	–0.42	1.34
S_ETHRGHT	*How much do you agree or disagree with the following statements?*				
IS3G25A	All <ethnic/racial groups> should have an equal chance to get a good education in <country of test>	–0.53	–1.27	–1.26	2.53
IS3G25B	All <ethnic/racial groups> should have an equal chance to get good jobs in <country of test>	–0.34	–1.82	–0.88	2.69
IS3G25C	Schools should teach students to respect <members of all ethnic/racial groups>	–0.25	–1.79	–0.66	2.45
IS3G25D	<Members of all ethnic/racial groups> should be encouraged to run in elections for political office	1.26	–2.65	–0.36	3.01
IS3G25E	<Members of all ethnic/racial groups> should have the same rights and responsibilities	–0.14	–1.38	–0.94	2.33

Students' attitudes toward civic institutions and their country of residence

Question 26 required students to indicate their level of trust ("completely," "quite a lot," "a little," "not at all") in 15 different institutions or groups. We used the level of trust reported for six of the items to derive the scale *students' trust in civic institutions* (S_INTRUST) where higher values on this scale reflect greater trust in civic institutions.

The items in Question 27 were five statements about the country of residence, which, in contrast to ICCS 2009, did not include statements about their satisfaction with specific aspects of the country. Students were asked to indicate their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with those statements. We used all question items to form a scale reflecting *students' attitudes toward their country* (S_CNTATT). The higher scores on the scale came from students who held the more favorable attitudes toward their country.

Figure 11.9: Confirmatory factor analysis of items measuring students' attitudes toward civic institutions and their country of residence



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.073	0.077	0.078	0.104
CFI	0.97	0.98	0.97	0.93
TLI	0.97	0.97	0.97	0.95

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.9). The model fit based on data from the pooled ICCS 2016 dataset with equally weighted countries was satisfactory and the two latent dimensions were moderately correlated (r = 0.37). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was no longer satisfactory for the most constrained models, which suggests a lack of measurement invariance across countries. Consequently, comparisons across countries for scores from these scales should be interpreted with some caution.

The scale reliability (Cronbach's alpha) for S_INTRUST was satisfactory, being 0.85 on average with coefficients ranging from 0.82 to 0.91 (see Table 11.16). For S_CNTATT the average reliability was 0.82 with coefficients ranging from 0.72 to 0.89. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.17). Given that items measuring S_INTRUST were identical to those administered in ICCS 2009, this scale was equated so that scale scores are comparable to those from the previous cycle. However, in the case of S_CNTATT, we refrained from equating given considerable differences when comparing IRT parameters from both cycles, and because this question in 2016 was presented without items measuring students' satisfaction with specific aspects of the country of residence (such as "The political system in <country of test> works well").

Table 11.16: Reliabilities for scales measuring students' attitudes toward civic institutions and their country of residence

Country	Scale reliability (Cronbach's alpha)	
	S_INTRUST	S_CNTATT
Belgium (Flemish)	0.83	0.76
Bulgaria	0.87	0.79
Chile	0.89	0.89
Chinese Taipei	0.88	0.88
Colombia	0.85	0.81
Croatia	0.84	0.81
Denmark	0.83	0.80
Dominican Republic	0.83	0.72
Estonia	0.82	0.88
Finland	0.87	0.85
Hong Kong SAR	0.87	0.87
Italy	0.82	0.76
Korea, Republic of	0.91	0.86
Latvia	0.85	0.88
Lithuania	0.83	0.84
Malta	0.82	0.84
Mexico	0.87	0.83
Netherlands	0.86	0.78
North-Rhine Westphalia (Germany)*	0.86	0.77
Norway	0.87	0.83
Peru	0.85	0.78
Russian Federation	0.85	0.87
Slovenia	0.86	0.83
Sweden	0.87	0.82
ICCS 2016 average	0.85	0.82
Minimum value	0.82	0.72
Maximum value	0.91	0.89

Note:
*Benchmarking participant.

Table 11.17: Item parameters for scales measuring students' attitudes toward civic institutions and their country of residence

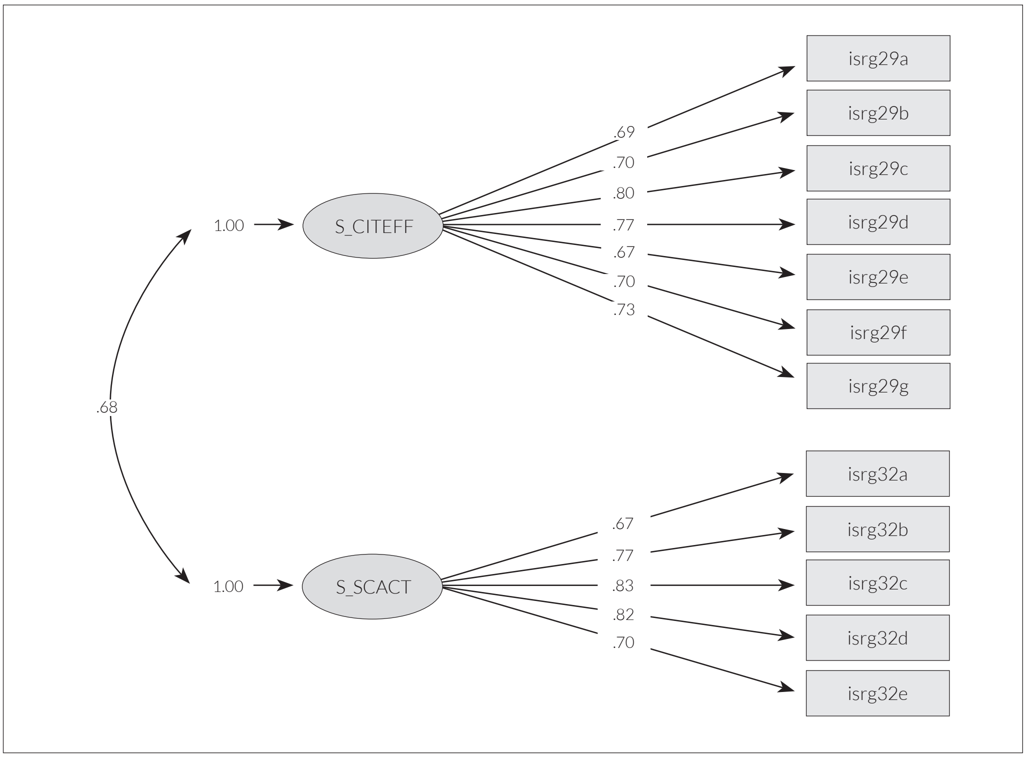
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_INTRUST	*How much do you trust each of the following groups, institutions or sources of information?*				
IS3G26A	The <national government> of <country of test>	–0.08	–2.46	–0.17	2.63
IS3G26B	The <local government> of your town or city	–0.21	–2.80	–0.27	3.07
IS3G26C	Courts of justice	–0.37	–2.49	–0.18	2.67
IS3G26D	The police	–0.53	–1.99	–0.20	2.19
IS3G26E	Political parties	0.91	–2.79	0.02	2.77
IS3G26F	<National Parliament>	0.28	–2.42	–0.20	2.62
S_CNTATT	*How much do you agree or disagree with the following statements about <country of test>?*				
IS3G27A	The <flag of country of test> is important to me	0.10	–1.51	–0.51	2.01
IS3G27B	I have great respect for <country of test>	–0.54	–1.66	–0.78	2.45
IS3G27C	In <country of test> we should be proud of what we have achieved	–0.39	–1.65	–0.60	2.25
IS3G27D	I am proud to live in <country of test>	–0.16	–1.49	–0.46	1.95
IS3G27E	Generally speaking, <country of test> is a better country to live in than most other countries	0.99	–2.00	–0.10	2.10

Students' dispositions toward civic engagement

Question 29 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* (S_CITEFF), and the higher values on this scale denote higher levels of confidence with respect to this form of self-efficacy.

Question 32 asked students to indicate how likely ("very likely," "likely," "not very likely," "not at all likely") they considered participating in the future in five different activities at school if they were given a chance. We used all question items to form a scale reflecting students' willingness to participate in school activities (S_SCACT). Higher scores on the scale came from students who perceived future participation in these activities as more likely.

Figure 11.10: Confirmatory factor analysis of items measuring students' dispositions toward civic engagement



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.087	0.096	0.090	0.095
CFI	0.94	0.94	0.94	0.91
TLI	0.93	0.93	0.94	0.93

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.10). The model fit based on data from the pooled ICCS 2016 dataset with equally weighted countries was marginally satisfactory and the two latent dimensions were strongly correlated ($r = 0.68$). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for more constrained models, suggesting a lack of measurement invariance across countries.

The scale reliability (Cronbach's alpha) for S_CITEFF was satisfactory, being 0.84 on average with coefficients ranging from 0.76 to 0.92 (see Table 11.18). For S_SCACT the average reliability was 0.81 with coefficients ranging from 0.74 to 0.87. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.19). Given that the question measuring S_CITEFF was identical (in stem and item wording) to those administered in ICCS 2009, this scale was equated so that scale scores are comparable to those from the previous cycle.

Table 11.18: Reliabilities for scales measuring students' dispositions toward civic engagement

Country	Scale reliability (Cronbach's alpha)	
	S_CITEFF	S_SCACT
Belgium (Flemish)	0.81	0.82
Bulgaria	0.83	0.83
Chile	0.90	0.86
Chinese Taipei	0.88	0.78
Colombia	0.83	0.79
Croatia	0.82	0.78
Denmark	0.83	0.77
Dominican Republic	0.76	0.74
Estonia	0.83	0.82
Finland	0.86	0.85
Hong Kong SAR	0.92	0.87
Italy	0.80	0.78
Korea, Republic of	0.91	0.84
Latvia	0.82	0.84
Lithuania	0.84	0.82
Malta	0.85	0.82
Mexico	0.85	0.82
Netherlands	0.84	0.84
North-Rhine Westphalia (Germany)*	0.83	0.82
Norway	0.87	0.79
Peru	0.78	0.75
Russian Federation	0.84	0.83
Slovenia	0.84	0.81
Sweden	0.87	0.83
ICCS 2016 average	0.84	0.81
Minimum value	0.76	0.74
Maximum value	0.92	0.87

Note:
*Benchmarking participant.

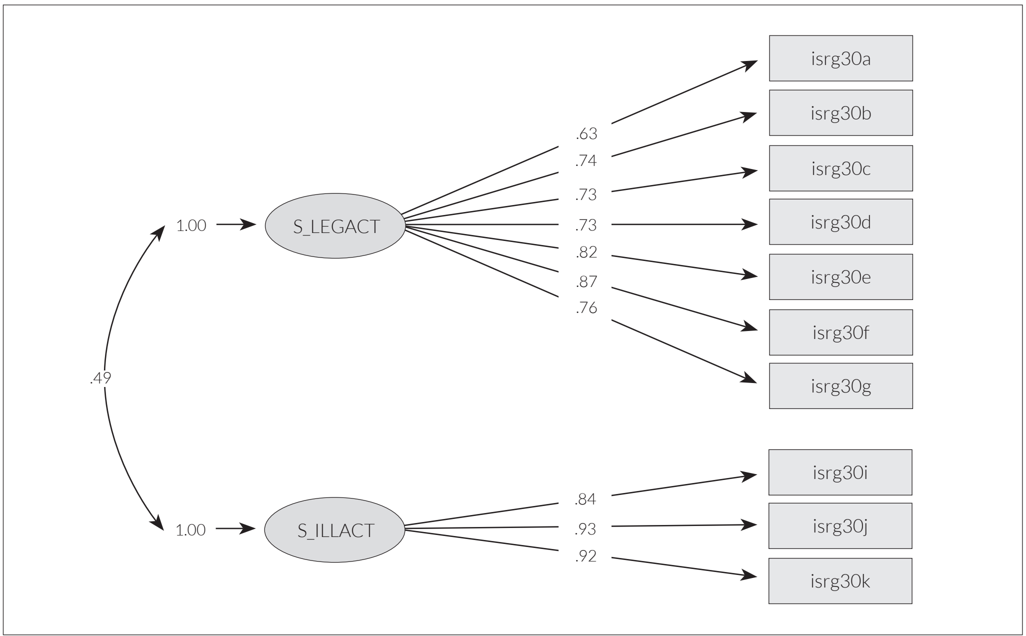
Table 11.19: Item parameters for scales measuring students' dispositions toward civic engagement

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_CITEFF	*How well do you think you would do the following activities?*				
IS3G29A	Discuss a newspaper article about a conflict between countries	−0.11	−2.44	0.02	2.42
IS3G29B	Argue your point of view about a controversial political or social issue	−0.31	−2.27	0.00	2.27
IS3G29C	Stand as a candidate in a <school election>	0.10	−1.96	0.09	1.87
IS3G29D	Organize a group of students in order to achieve changes at school	−0.13	−1.89	−0.02	1.91
IS3G29E	Follow a television debate about a controversial issue	0.14	−2.19	0.08	2.11
IS3G29F	Write a letter or email to a newspaper giving your view on a current issue	0.14	−1.91	0.03	1.88
IS3G29G	Speak in front of your class about a social or political issue	0.17	−1.65	−0.02	1.67
S_SCACT	*If you were given the chance, how likely is it that you would participate in each activity?*				
IS3G32A	Vote in a school election of <class representatives> or <school parliament>	−1.06	−1.42	−0.04	1.45
IS3G32B	Join a group of students campaigning for an issue you agree with	−0.28	−2.04	0.10	1.94
IS3G32C	Become a candidate for <class representative> or <school parliament>	0.42	−1.71	0.45	1.25
IS3G32D	Take part in discussions in a <student assembly>	0.22	−1.90	0.19	1.71
IS3G32E	Participate in writing articles for a school newspaper or website	0.70	−1.74	0.32	1.42

Students' expectations to engage in activities to express their opinion

Question 30 contained a list of 11 possible ways that citizens can use to express their opinions ("I would certainly do this," "I would probably do this," "I would probably not do this," and "I would certainly not do this"). We used the first six of these items, which are related to legal activities, to construct the scale *students' expected participation in future legal activities* (S_LEGACT). Further, we used the three items from Question 31 to form the scale *students' expected participation in future illegal activities* (S_ILLACT). The higher values on both of these scales reflect a greater likelihood of participation in the respective activities.

Figure 11.11: Confirmatory factor analysis of items measuring students' expectations to engage in activities to express their opinion



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.086	0.100	0.097	0.095
CFI	0.98	0.97	0.97	0.95
TLI	0.97	0.96	0.96	0.96

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.11). The model fit based on data from the pooled ICCS 2016 dataset with equally weighted countries was marginally satisfactory and the two latent dimensions were moderately correlated (r = 0.49). When reviewing measurement invariance using across multiple-group models with different constraints, the results showed that the model was measuring in similar ways across countries albeit with an only marginally satisfactory fit.

The scale reliability (Cronbach's alpha) for S_LEGACT was satisfactory, being 0.85 on average with coefficients ranging from 0.80 to 0.91 (see Table 11.20). For S_ILLACT the average reliability was 0.87 with coefficients ranging from 0.82 to 0.95. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.21). Given that the question had been modified from ICCS 2009 (now referring to activities to express opinions instead of protest), these two scales were not equated, so it is not possible to compare scale scores with the two scales of similar content from the previous cycle.

Table 11.20: Reliabilities for scales measuring students' expectations to engage in activities to express their opinion

Country	Scale reliability (Cronbach's alpha)	
	S_LEGACT	S_ILLACT
Belgium (Flemish)	0.84	0.87
Bulgaria	0.84	0.82
Chile	0.91	0.89
Chinese Taipei	0.89	0.92
Colombia	0.85	0.85
Croatia	0.83	0.88
Denmark	0.82	0.88
Dominican Republic	0.85	0.83
Estonia	0.82	0.85
Finland	0.87	0.90
Hong Kong SAR	0.91	0.95
Italy	0.82	0.83
Korea, Republic of	0.91	0.90
Latvia	0.84	0.85
Lithuania	0.83	0.88
Malta	0.86	0.88
Mexico	0.86	0.88
Netherlands	0.87	0.87
North-Rhine Westphalia (Germany)*	0.80	0.87
Norway	0.89	0.87
Peru	0.81	0.82
Russian Federation	0.85	0.88
Slovenia	0.83	0.87
Sweden	0.87	0.90
ICCS 2016 average	0.85	0.87
Minimum value	0.80	0.82
Maximum value	0.91	0.95

Note:
*Benchmarking participant.

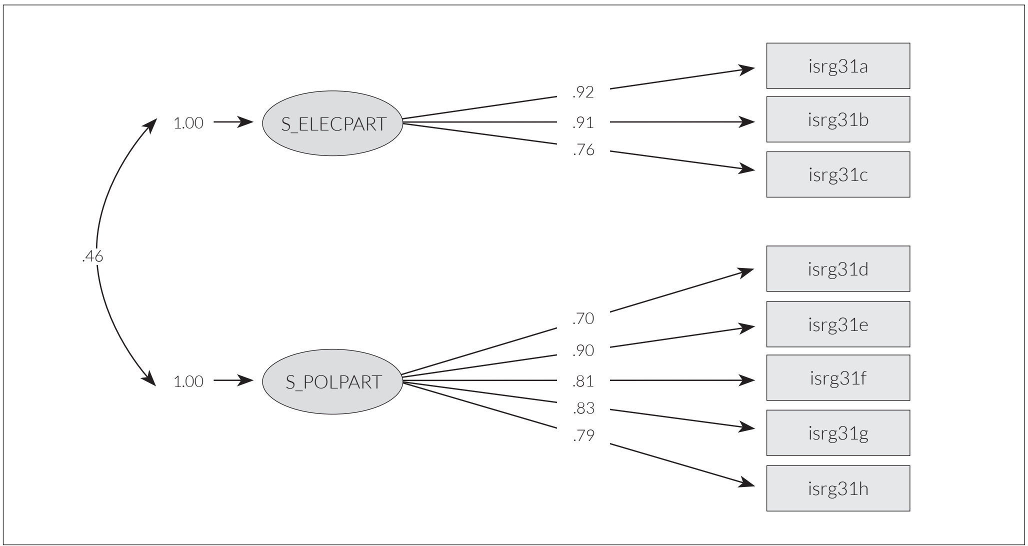
Table 11.21: Item parameters for scales measuring students' expectations to engage in activities to express their opinion

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_LEGACT	*Would you take part in any of the following activities to express your opinion in the future?*				
IS3G30A	Talk to others about your views on political or social issues	–0.71	–2.30	–0.06	2.36
IS3G30B	Contact an <elected representative>	0.31	–2.58	0.32	2.27
IS3G30C	Take part in a peaceful march or rally	–0.18	–2.27	0.14	2.12
IS3G30D	Collect signatures for a petition	–0.13	–2.30	0.18	2.13
IS3G30E	Contribute to an online discussion forum about social or political issues	0.14	–2.42	0.23	2.18
IS3G30F	Organize an online group to take a stance on a controversial political or social issue	0.43	–2.39	0.41	1.98
IS3G30G	Participate in an online campaign	0.14	–2.34	0.15	2.19
S_ILLACT	*Would you take part in any of the following activities to express your opinion in the future?*				
IS3G30I	Spray-paint protest slogans on walls	–0.28	–2.80	0.50	2.30
IS3G30J	Stage a protest by blocking traffic	0.05	–2.76	0.55	2.22
IS3G30K	Occupy public buildings as a sign of protest	0.22	–2.61	0.56	2.05

Students' expectations of political participation

Question 31 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"). We used the first three items for the question to construct the scale *students' expected electoral participation* (S_ELECPART) and five items to construct the S_POLPART scale (*students' expected active political participation*). Higher scores for these two scales indicated greater expectancy of participation in the respective activities.

Figure 11.12: Confirmatory factor analysis of items measuring students' expected political participation



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0,103	0.118	0.118	0.112
CFI	0,98	0.98	0.97	0.96
TLI	0,97	0.97	0.97	0.97

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.12). The model fit based on data from the pooled ICCS 2016 dataset with equally weighted countries was only marginally satisfactory and the two latent dimensions were moderately correlated ($r = 0.46$). When reviewing measurement invariance using across multiple-group models with different constraints, the results showed that the model was measuring in similar ways across countries, albeit with non-satisfactory fit.

The scale reliability (Cronbach's alpha) for S_ELECPART was satisfactory, being 0.83 on average with coefficients ranging from 0.76 to 0.94 (see Table 11.22). For S_POLPART the average reliability was 0.85 with coefficients ranging from 0.72 to 0.91. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.23). Given that question 31 was identical with a question included in ICCS 2009 and for S_POLPART only one item had been added, the scale scores for both S_ELECPART and S_POLPART were equated and are therefore comparable with those from the previous cycle.

Table 11.22: Reliabilities for scales measuring students' expected political participation

Country	Scale reliability (Cronbach's alpha)	
	S_ELECPART	S_POLPART
Belgium (Flemish)	0.77	0.86
Bulgaria	0.84	0.88
Chile	0.91	0.91
Chinese Taipei	0.85	0.84
Colombia	0.83	0.86
Croatia	0.79	0.85
Denmark	0.79	0.72
Dominican Republic	0.76	0.84
Estonia	0.83	0.83
Finland	0.84	0.82
Hong Kong SAR	0.94	0.89
Italy	0.82	0.81
Korea, Republic of	0.86	0.90
Latvia	0.86	0.86
Lithuania	0.84	0.86
Malta	0.79	0.87
Mexico	0.85	0.89
Netherlands	0.83	0.84
North-Rhine Westphalia (Germany)*	0.84	0.79
Norway	0.89	0.84
Peru	0.76	0.84
Russian Federation	0.85	0.89
Slovenia	0.87	0.82
Sweden	0.83	0.81
ICCS 2016 average	0.83	0.85
Minimum value	0.76	0.72
Maximum value	0.94	0.91

Note:
*Benchmarking participant.

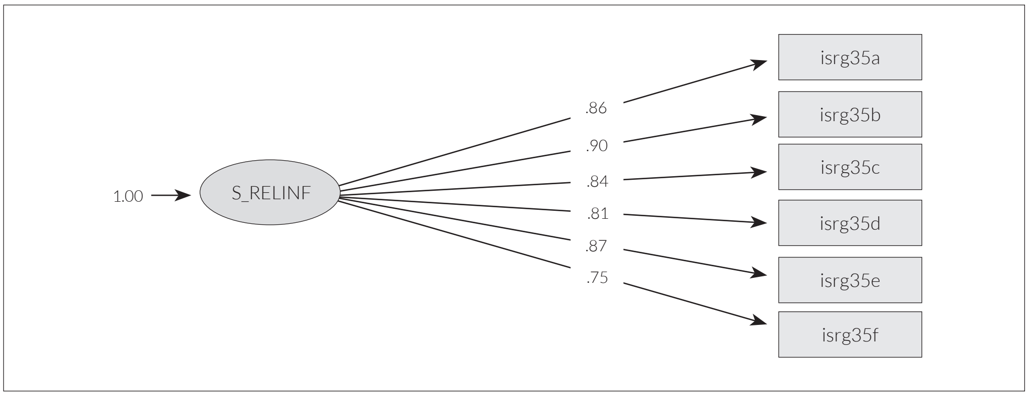
Table 11.23: Item parameters for scales measuring students' expected political participation

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_ELECPART	*When you are an adult, what do you think you will do?*				
IS3G31A	Vote in <local elections>	–0.11	–1.98	–0.62	2.60
IS3G31B	Vote in <national elections>	–0.12	–2.06	–0.46	2.53
IS3G31C	Get information about candidates before voting in an election	0.23	–2.27	–0.28	2.55
S_POLPART	*When you are an adult, what do you think you will do?*				
IS3G31D	Help a candidate or party during an election campaign	–0.69	–2.91	0.42	2.49
IS3G31E	Join a political party	0.30	–2.48	0.60	1.88
IS3G31F	Join a trade union	0.05	–2.53	0.35	2.18
IS3G31G	Stand as a candidate in <local elections>	0.41	–2.27	0.61	1.66
IS3G31H	Join an organization for a political or social cause	–0.07	–2.62	0.37	2.25

Students' endorsement of religious influence in society

Question 35 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"). We used five ICCS 2009 items and one new ICCS 2016 item to form a scale reflecting *students' endorsement of the influence of religion on society* (S_RELINF). Higher scores on this scale indicated stronger agreement with the notion that religion should play an important role in shaping society.

Figure 11.13: Confirmatory factor analysis of items measuring students' endorsement of religious influence in society



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.090	0.102	0.093	0.096
CFI	0.99	0.99	0.99	0.98
TLI	0.99	0.98	0.99	0.99

The results from confirmatory factor analyses assume a one-dimensional model for these items (Figure 11.13). The model fit based on data from the pooled ICCS 2016 dataset with equally weighted countries was marginally satisfactory. When reviewing measurement invariance using across multiple-group models with different constraints, the results showed that the model was measuring in similar ways across countries, albeit with only marginally satisfactory fit.

The scale reliability (Cronbach's alpha) for S_RELINF was satisfactory, being 0.87 on average with coefficients ranging from 0.73 to 0.93 (see Table 11.24). Selected item parameters were used to scale the items (Table 11.25). Given that five of the six items used for scaling were identical with those included in ICCS 2009, the scale was equated and scale scores are therefore comparable with those from the previous cycle.

Table 11.24: Reliabilities for scale measuring students' endorsement of religious influence in society

Country	Scale reliability (Cronbach's alpha)
	S_RELINF
Belgium (Flemish)	0.89
Bulgaria	0.87
Chile	0.92
Chinese Taipei	0.86
Colombia	0.82
Croatia	0.87
Denmark	0.88
Dominican Republic	0.73
Estonia	0.89
Finland	n/a
Hong Kong SAR	0.91
Italy	n/a
Korea, Republic of	0.93
Latvia	0.90
Lithuania	0.88
Malta	0.85
Mexico	n/a
Netherlands	0.91
North-Rhine Westphalia (Germany)*	n/a
Norway	0.90
Peru	0.80
Russian Federation	n/a
Slovenia	0.90
Sweden	0.92
ICCS 2016 average	0.87
Minimum value	0.73
Maximum value	0.93

Notes:
*Benchmarking participant.
n/a = not applicable.

Table 11.25: Item parameters for scale measuring students' endorsement of religious influence in society

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
S_RELINF	*How much do you agree or disagree with the following statements about religion?*				
IS3G35A	Religion is more important to me than what is happening in national politics	–0.34	–2.39	0.37	2.02
IS3G35B	Religion helps me to decide what is right and what is wrong	–0.42	–2.17	0.01	2.17
IS3G35C	Religious leaders should have more power in society.	0.51	–2.68	0.44	2.25
IS3G35D	Religion should influence people's behaviour towards others	–0.46	–2.13	–0.25	2.38
IS3G35E	Rules of life based on religion are more important than civil laws	0.33	–2.54	0.36	2.18
IS3G35G	Religious people are better citizens	0.39	–2.24	0.33	1.91

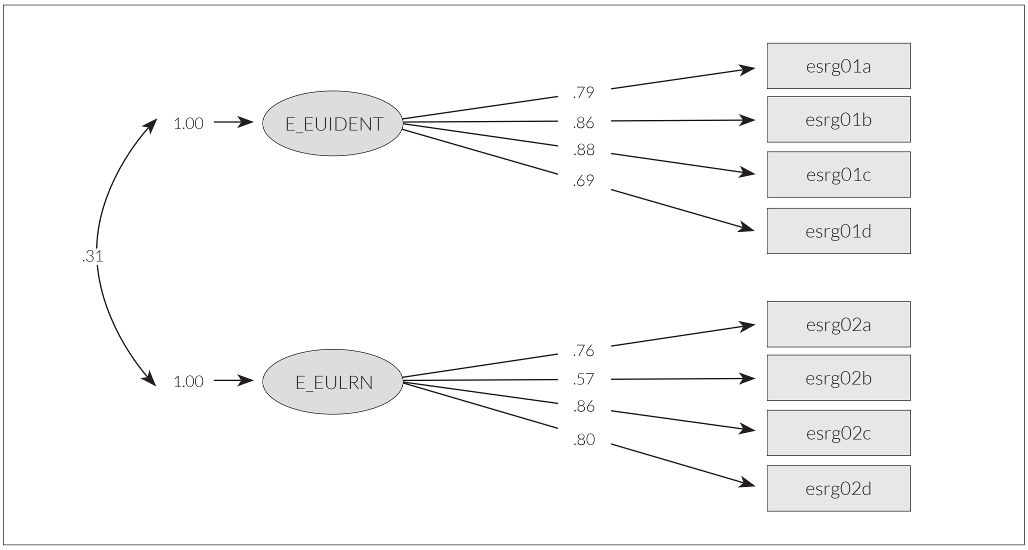
European student questionnaire

Students' perceptions of European identify and learning about Europe at school

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"). We used four of the question items to construct the scale *students' sense of European identity* (E_EUIDENT). The higher values on this scale reflect a stronger sense of European identity.

Question 2 asked students to indicate to what extent they had had opportunities to learn about Europe at school ("to a large extent," "to some extent," "to a small extent," "not at all"). We used the four items associated with the question to derive the scale *students' reports on opportunities for learning about Europe at school* (E_EULRN), where higher scores on this scale correspond to greater amounts of opportunity.

Figure 11.14: Confirmatory factor analysis of items measuring students' perceptions of European identity and learning about Europe at school



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.033	0.038	0.040	0.061
CFI	1.00	1.00	0.99	0.97
TLI	0.99	0.99	0.99	0.98

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.14). There was close model fit (i.e. $p < 0.05$) based on data from the pooled ICCS 2016 dataset with equally weighted countries and the two latent dimensions were moderately correlated ($r = 0.31$). When reviewing measurement invariance using across multiple-group models with different constraints, the results showed that the model fit for more constrained models was not as close but still within a satisfactory range.

The scale reliability (Cronbach's alpha) for E_EUIDENT was quite high, being 0.80 on average with coefficients ranging from 0.74 to 0.84 (see Table 11.26). For E_EULRN the average reliability was 0.77 with coefficients ranging from 0.72 to 0.81. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.27). Given that the four items measuring E_EUIDENT were identical to four out of the five items used to measure this construct in ICCS 2009, the scale was equated so that it is possible to compare scale scores with those from the previous cycle.

Table 11.26: Reliabilities for scales measuring students' perceptions of European identity and learning about Europe at school

Country	Scale reliability (Cronbach's alpha)	
	E_EUIDENT	E_EULRN
Belgium (Flemish)	0.79	0.76
Bulgaria	0.84	0.79
Croatia	0.83	0.80
Denmark	0.76	0.72
Estonia	0.83	0.77
Finland	0.82	0.79
Italy	0.77	0.72
Latvia	0.80	0.75
Lithuania	0.77	0.74
Malta	0.82	0.81
Netherlands	0.80	0.75
North-Rhine Westphalia (Germany)*	0.74	0.77
Norway	0.83	0.79
Slovenia	0.80	0.76
Sweden	0.80	0.81
ICCS 2016 average	0.80	0.77
Minimum value	0.74	0.72
Maximum value	0.84	0.81

Note:
*Benchmarking participant.

Table 11.27: Item parameters for scales measuring students' perceptions of European identity and learning about Europe at school

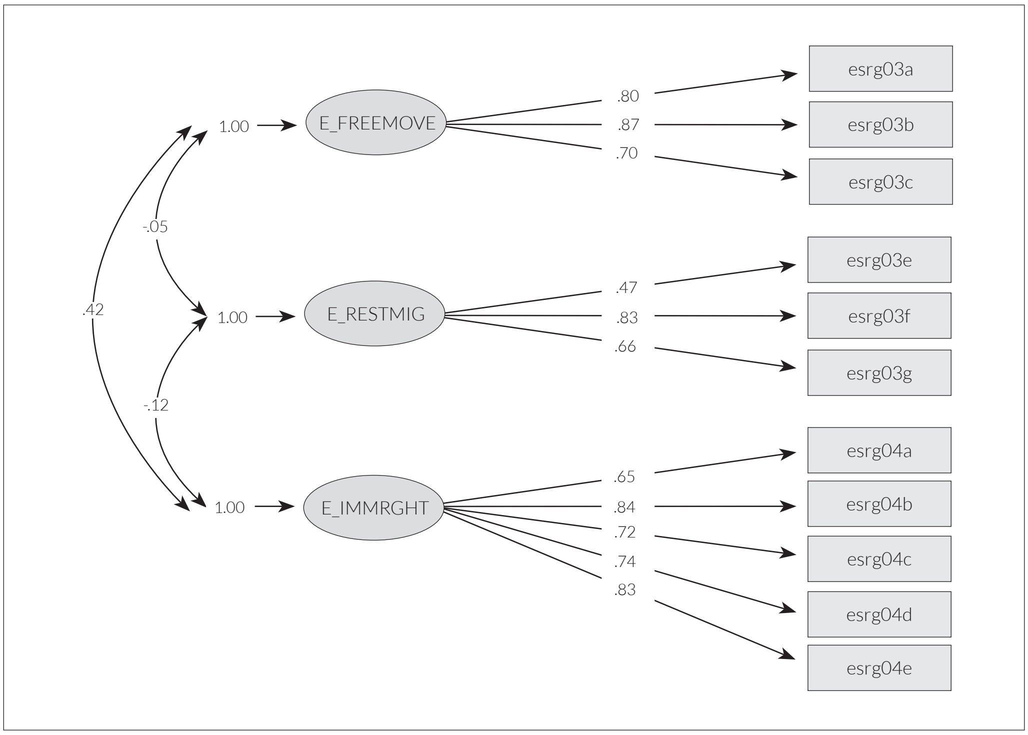
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
E_EUIDENT	*How much do you agree or disagree with the following statements?*				
ES3G01A	I see myself as European	–0.67	–1.59	–1.03	2.62
ES3G01B	I am proud to live in Europe	–0.52	–2.26	–0.97	3.23
ES3G01C	I feel part of Europe	0.22	–2.68	–0.56	3.24
ES3G01D	I see myself first as a citizen of Europe and then as a citizen of the world	0.97	–2.81	–0.32	3.14
E_EULRN	*At school, to what extent have you had the opportunity to learn about the following issues or topics?*				
ES3G02A	Political and economic systems of other European countries	0.33	–2.57	–0.42	2.99
ES3G02B	The history of Europe	–0.97	–2.00	–0.28	2.28
ES3G02C	Political and social issues in other European countries	0.35	–2.75	–0.24	2.99
ES3G02D	Political and economic integration between European) countries (for example the European Union	0.29	–2.41	–0.30	2.71

Students' attitudes toward migration in Europe

Question 3 of the European student questionnaire asked students to indicate their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with a range of statements about possibilities for European citizens to move and work in other European countries. Three were positive statements endorsing the freedom of individuals to live and work in their choice of European countries, while three other statements were concerned with restricting freedom of movement. We used each of these item sets to construct two scales reflecting *students' endorsement of freedom of migration within Europe* (E_FREEMOVE) and of *students' endorsement of restricting migration in Europe* (E_RESTMIG).[13] Higher scores on both scales indicate higher levels of endorsement of the respective statements.

Question 4 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 of these statements. We used five of the ICCS 2009 items to construct the scale *students' endorsement of equal rights for immigrants* (E_IMMRGHT).[14] Students with higher scores on this scale reflected higher levels of endorsement of equal rights for immigrants.

Figure 11.15: Confirmatory factor analysis of items measuring students' attitudes toward migration in Europe



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.068	0.081	0.080	0.085
CFI	0.96	0.95	0.95	0.92
TLI	0.95	0.94	0.94	0.93

13 In the European ICCS 2016 report (Losito, Agrusti, Damiani, & Schulz, 2018), these two scales were called *students' attitudes toward freedom of movement and students' attitudes toward restricting migration in Europe.*

14 In the European ICCS 2016 report (Losito et al., 2018), this scale was called *students' attitudes toward equal rights for immigrants.*

The results from confirmatory factor analyses assume a three-dimensional model for the (scaled) items from these three questions (Figure 11.15). The model fit was satisfactory. Correlations between the three latent dimensions show a moderate correlation of r = 0.42 between E_FREEMOVE and E_IMMRGHT, only weak negative correlations were observed between E_FREEMOVE and E_RESTMIG (r = –0.05), and E_IMMRGHT and E_RESTMIG (r = –0.12). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for the more constrained model indicating a lack of measurement invariance although it was still within an acceptable range.

The reliabilities (Cronbach's alpha) were satisfactory for the two scales E_FREEMOVE (with an average of 0.74, ranging from 0.67 to 0.79) and E_IMMRGHT (with an average of 0.81 ranging from 0.77 to 0.87), while E_RESTMIG had only marginally satisfactory reliability with 0.64 (ranging from 0.47 to 0.73) (Table 11.28). Given the conceptual importance of this construct, we decided to retain this (short) scale; but the lack of internal scale consistency should be borne in mind, in particular in countries where it was quite low, such as Latvia and Norway.

Selected item parameters were used to scale the items corresponding to these three scales (Table 11.29). Given that all items measuring E_IMMRGHT (both stem and item wording) were identical with those used in ICCS 2009, this scale was equated so that its scores are comparable with those from the previous cycle.

Table 11.28: Reliabilities for scales measuring students' attitudes toward migration in Europe

Country	Scale reliability (Cronbach's alpha)		
	E_FREEMOVE	E_RESTMIG	E_IMMRGHT
Belgium (Flemish)	0.76	0.65	0.78
Bulgaria	0.74	0.60	0.82
Croatia	0.78	0.60	0.82
Denmark	0.73	0.63	0.80
Estonia	0.76	0.66	0.77
Finland	0.77	0.73	0.85
Italy	0.67	0.60	0.80
Latvia	0.67	0.59	0.78
Lithuania	0.71	0.68	0.77
Malta	0.75	0.63	0.79
Netherlands	0.79	0.68	0.77
North-Rhine Westphalia (Germany)*	0.76	0.69	0.84
Norway	0.78	0.47	0.84
Slovenia	0.71	0.70	0.80
Sweden	0.75	0.63	0.87
ICCS 2016 average	0.74	0.64	0.81
Minimum value	0.67	0.47	0.77
Maximum value	0.79	0.73	0.87

Note:
*Benchmarking participant.

Table 11.29: Item parameters for scales measuring students' attitudes toward migration in Europe

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
E_FREEMOVE	*How much do you agree or disagree with the following statements related to the possibilities for European citizens to work in other European countries?*				
ES3G03A	Allowing citizens of European countries to work anywhere in Europe is good for the European economy	–0.13	–2.81	–1.15	3.95
ES3G03B	Citizens of European countries should be allowed to work anywhere in Europe	–0.15	–2.99	–0.71	3.70
ES3G03C	Allowing citizens of European countries to work anywhere in Europe helps to reduce unemployment	0.28	–3.01	–0.61	3.62
E_RESTMIG	*How much do you agree or disagree with the following statements related to the possibilities for European citizens to work in other European countries?*				
ES3G03D	Citizens of European countries should be allowed to work in another European country only if their skills are needed there	–0.74	–2.28	0.09	2.19
ES3G03E	Citizens of European countries who wish to work in another country should be allowed to take only the jobs that no one in the other country wants to do	0.33	–1.92	0.36	1.56
ES3G03F	Only a limited number of people should be allowed to move for work from one European country to another	0.41	–1.84	0.15	1.69
E_IMMRGHT	*How much do you agree or disagree with the following statements about <immigrants>?*				
ES3G04A	<Immigrants> should have the opportunity to continue speaking their own language	0.70	–2.24	–0.42	2.66
ES3G04B	<Immigrant> children should have the same opportunities for education that other children in the country have	–0.93	–1.38	–1.17	2.55
ES3G04C	<Immigrants> who live in a country for several years should have the opportunity to vote in elections	0.24	–2.21	–0.33	2.53
ES3G04D	<Immigrants> should have the opportunity to continue their own customs and lifestyle	0.49	–2.05	–0.53	2.58
ES3G04E	<Immigrants> should have the same rights that everyone else in the country has	–0.50	–1.49	–0.74	2.22

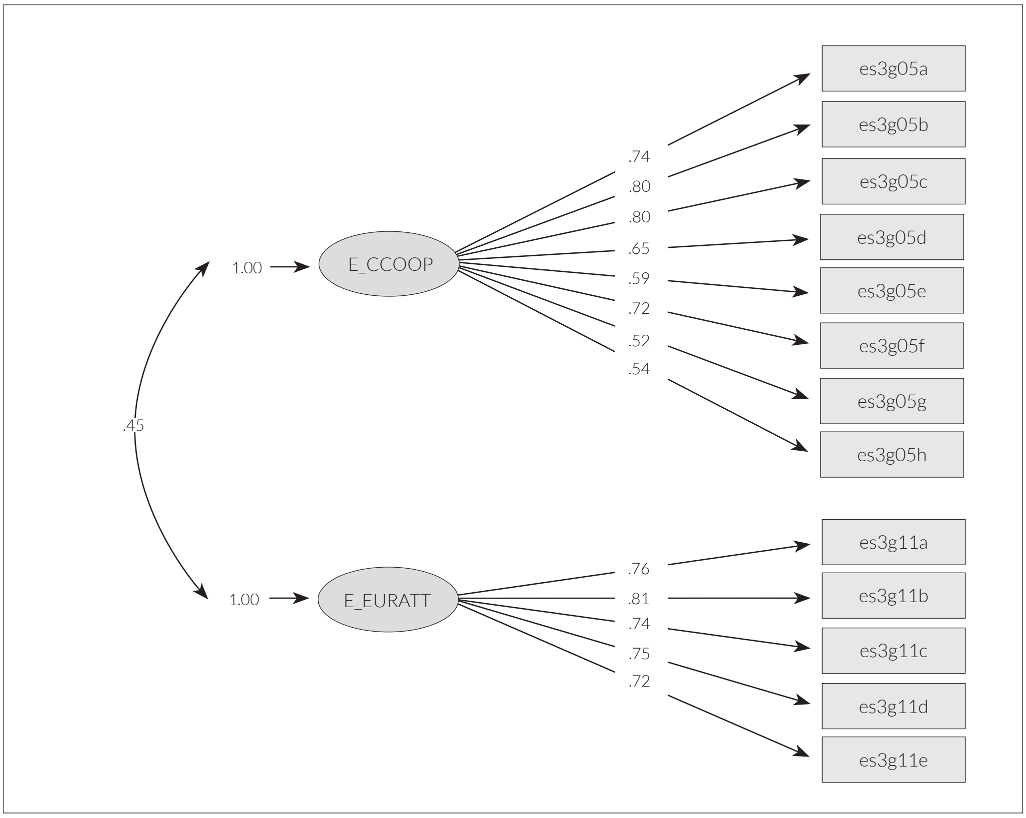
Students' attitudes toward European integration

Question 5 of the European regional questionnaire asked students to indicate their level of agreement with eight positively worded statements about cooperation between European countries ("strongly agree," "agree," "disagree," "strongly disagree"). We used these eight question items to construct the scale *students' endorsement of European cooperation* (E_CCOOP).[15] Higher values on this scale reflected higher levels of endorsement of cooperation between European countries.

Question 11 of the European regional questionnaire asked students to indicate their level of agreement with five positively worded statements about the European Union ("strongly agree," "agree," "disagree," "strongly disagree"). We used these question items to construct the scale *students' positive attitudes toward the European Union* (E_EURATT). The higher values on this scale reflected more positive views of the European Union.

15 In the European ICCS 2016 report (Losito et al., 2018), this scale was called *students' attitudes toward cooperation among European countries.*

Figure 11.16: Confirmatory factor analysis of items measuring students' attitudes toward European integration



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.051	0.058	0.052	0.060
CFI	0.97	0.97	0.97	0.95
TLI	0.96	0.96	0.97	0.96

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.16). The model fit was satisfactory based on data from the pooled ICCS 2016 dataset with equally weighted countries and the two latent dimensions were moderately correlated (r = 0.45). When reviewing measurement invariance using across multiple-group models with different constraints, the results showed that the model fit for more constrained models was not as close, although still within a satisfactory range.

The scale reliability (Cronbach's alpha) for both scales was satisfactory (Table 11.30): E_CCOOP had an average reliability of 0.79 (ranging from 0.72 to 0.83) and E_EURATT had average reliability of 0.80 with coefficients in participating countries ranging from 0.70 to 0.87. Selected item parameters were used to scale the items corresponding to these two scales, which were new to ICCS 2016 (Table 11.31).

Table 11.30: Reliabilities for scales measuring students' attitudes toward European integration

Country	Scale reliability (Cronbach's alpha)	
	E_CCOOP	E_EURATT
Belgium (Flemish)	0.77	0.77
Bulgaria	0.83	0.81
Croatia	0.83	0.87
Denmark	0.74	0.70
Estonia	0.81	0.81
Finland	0.79	0.83
Italy	0.72	0.77
Latvia	0.78	0.78
Lithuania	0.82	0.82
Malta	0.81	0.80
Netherlands	0.79	0.81
North-Rhine Westphalia (Germany)*	0.81	0.79
Norway	0.78	0.84
Slovenia	0.79	0.79
Sweden	0.81	0.82
ICCS 2016 average	0.79	0.80
Minimum value	0.72	0.70
Maximum value	0.83	0.87

Note:
*Benchmarking participant.

Table 11.31: Item parameters for scales measuring students' attitudes toward European integration

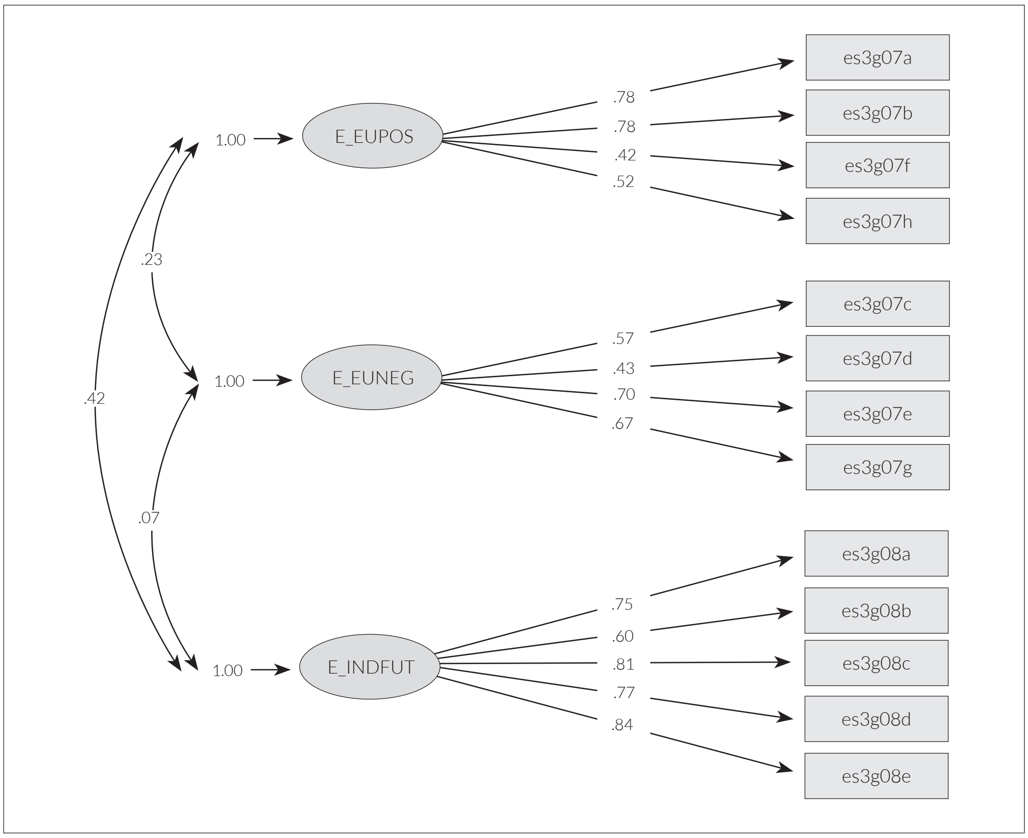
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
E_CCOOP	*How much do you agree or disagree with the following statements?*				
ES3G05A	European countries should cooperate to protect the environment (e.g. through programmes to limit pollution, programmes to combat climate change)	–0.64	–0.75	–1.49	2.24
ES3G05B	European countries should cooperate to guarantee high levels of employment	–0.24	–1.70	–1.11	2.81
ES3G05C	European countries should cooperate to strengthen their economies	–0.25	–1.65	–0.96	2.61
ES3G05D	European countries should recognize all educational qualifications achieved in any other European country	0.26	–1.88	–0.66	2.54
ES3G05E	European countries should have a European army for peace keeping missions	0.48	–2.03	–0.42	2.45
ES3G05F	European countries should cooperate to prevent and combat terrorism	–0.39	–0.72	–0.90	1.61
ES3G05G	European countries should cooperate to combat illegal entry from non-European countries	0.37	–1.64	–0.55	2.19
ES3G05H	European countries should cooperate to provide shelter to people escaping persecution in their countries for reasons of race, religion, or political opinions	0.41	–1.14	–0.95	2.09
E_EURATT	*How much do you agree or disagree with the following statements?*				
ES3G11A	<EU> guarantees respect for human rights all over Europe	–0.38	–3.01	–0.82	3.83
ES3G11B	<EU> makes Europe a safe place to live	–0.07	–3.14	–0.72	3.86
ES3G11C	<EU> takes care of the environment	0.28	–3.52	–0.37	3.89
ES3G11D	<EU> is good for the economy of individual countries	0.33	–3.15	–0.84	3.99
ES3G11E	<EU> is good because countries share a common set of rules and laws	–0.16	–2.57	–1.06	3.63

Students' expectations for Europe's and their own individual future

Question 7 of the European student questionnaire asked students how likely they viewed eight different scenarios for Europe in the future ("very likely," "likely," "unlikely," "very unlikely"). Four were positive scenarios, while the other four suggested rather negative developments. We used each of these item sets to construct two scales reflecting *students' positive expectations for European future* (E_EUPOS) and *students' negative expectations for European future* (E_EUNEG). For the E_EUPOS scale, higher scores reflected higher levels of perceived likelihood for the respective scenarios, while for the E_EUNEG scale, higher scores reflected lower levels.

Question 8 presented students with a list of five statements about their life in the future. Students were asked to indicate how likely they thought their individual future would resemble these statements ("very likely," "likely," "unlikely," "very unlikely"). We used these five items to construct the scale students' positive expectations for their own individual future (E_INDFUT). Students with higher scores on this scale reflected greater perceived likelihood for the respective scenarios.

Figure 11.17: Confirmatory factor analysis of items measuring students' expectations for Europe's and their own individual future



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.052	0.081	0.080	0.085
CFI	0.96	0.95	0.95	0.92
TLI	0.95	0.94	0.94	0.93

The results from confirmatory factor analyses showed a satisfactory model fit for a three-dimensional model (Figure 11.17). The results also showed a moderate positive correlation of $r = 0.42$ between E_EUPOS and E_INDFUT, and a weak positive correlation between E_EUPOS and E_EUNEG ($r = 0.23$); the (positive) correlation between E_INDFUT and E_EUNEG was negligible ($r = 0.07$). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for the more constrained model indicating a lack of measurement invariance, although the fit was still within a marginally satisfactory range.

The scale reliability (Cronbach's alpha) of two of the scales was only marginally satisfactory (Table 11.32); E_EUPOS had an average of 0.64, ranging from 0.58 to 0.70, and E_EUNEG had an average of 0.62, ranging from 0.50 to 0.71. Meanwhile, the E_INDFUT scale had a highly satisfactory reliability of 0.80 (ranging from 0.76 to 0.83). Selected item parameters were used to scale the items corresponding to these three scales, which were new to ICCS 2016 (Table 11.33).

Table 11.32: Reliabilities for scales measuring students' expectations for Europe's and their own individual future

Country	Scale reliability (Cronbach's alpha)		
	E_EUPOS	E_EUNEG	E_INDFUT
Belgium (Flemish)	0.64	0.57	0.78
Bulgaria	0.69	0.68	0.82
Croatia	0.68	0.71	0.83
Denmark	0.58	0.50	0.76
Estonia	0.66	0.63	0.81
Finland	0.60	0.62	0.83
Italy	0.64	0.58	0.76
Latvia	0.66	0.63	0.79
Lithuania	0.67	0.62	0.80
Malta	0.63	0.59	0.77
Netherlands	0.59	0.56	0.80
North-Rhine Westphalia (Germany)*	0.70	0.64	0.77
Norway	0.59	0.64	0.80
Slovenia	0.67	0.70	0.79
Sweden	0.63	0.63	0.83
ICCS 2016 average	0.64	0.62	0.80
Minimum value	0.58	0.50	0.76
Maximum value	0.70	0.71	0.83

Note:
*Benchmarking participant.

Table 11.33: Item parameters for scales measuring students' expectations for Europe's and their own individual future

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
E_EUPOS	*What is Europe likely to look like in 10 years?*				
ES3G07A	There will be stronger cooperation among European countries	–0.71	–1.92	–0.65	2.57
ES3G07B	There will be greater peace across Europe	0.09	–2.31	0.08	2.22
ES3G07F	There will be less air and water pollution in Europe	0.86	–2.14	0.09	2.05
ES3G07H	Democracy will be strengthened across Europe.	–0.24	–2.25	–0.46	2.71
E_EUNEG	*What is Europe likely to look like in 10 years?*				
ES3G07C	Terrorism will be more of a threat all across Europe	0.39	–1.95	–0.10	2.05
ES3G07D	Europe will be more influenced by non-European powers like China, India and the United States	0.26	–2.22	0.08	2.14
ES3G07E	The economy will be weaker in all European countries	–0.43	–2.28	–0.42	2.70
ES3G07G	There will be a rise in poverty and unemployment in Europe	–0.21	–2.25	–0.18	2.43
E_INDFUT	*How likely do you think it is that your future will look like this?*				
ES3G08A	I will find a steady job	–0.55	–2.15	–0.90	3.05
ES3G08B	My financial situation will be better than that of my parents	0.81	–3.67	–0.13	3.81
ES3G08C	I will find a job I like	–0.19	–2.47	–0.49	2.96
ES3G08D	I will have the opportunity to travel abroad for leisure	0.27	–2.17	–0.61	2.78
ES3G08E	I will earn enough money to start a family	–0.34	–1.45	–1.37	2.82

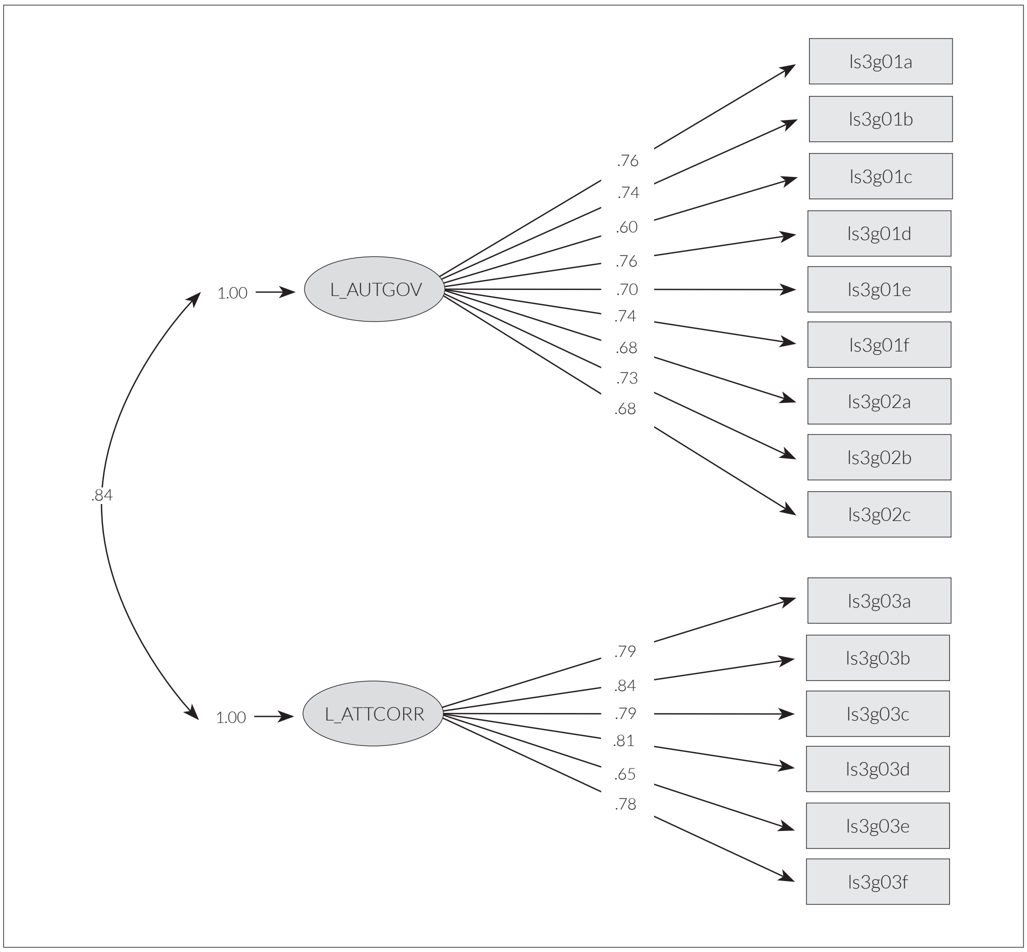
Latin American student questionnaire

Students' attitudes toward authoritarian and corrupt government practices

Questions 1 and 2 of the Latin American student questionnaire required students to rate their level of agreement with eleven different statements about government and its leaders, or its power ("strongly agree," "agree," "disagree," or "strongly disagree"). We used all the items from Question 1 and the first three items from Question 2 to derive the scale *students' endorsement of authoritarian government practices* (L_AUTGOV). Higher scale scores correspond to greater acceptance of governments engaging in authoritarian practices.

In Question 3, students were asked to rate their level of agreement ("strongly agree," "agree," "disagree," "strongly disagree") with six (positive) statements about the permissiveness of corruption in public service and the government. We used these items to derive the scale *students' endorsement of corrupt practices in government* (L_ATTCOR). The higher scores on this scale denote a greater degree of acceptance of corrupt government practices.

Figure 11.18: Confirmatory factor analysis of items measuring students' attitudes toward authoritarian and corrupt government practices



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.049	0.052	0.051	0.050
CFI	0.98	0.98	0.98	0.98
TLI	0.98	0.98	0.98	0.98

The results from confirmatory factor analyses showed a close model fit (i.e. $p < 0.05$) for a two-dimensional model for the combined set of items (Figure 11.18). The two latent dimensions were strongly correlated ($r = 0.84$). When reviewing measurement invariance usin remained equally good with more constrained models.

Scale reliabilities (Cronbach's alpha) for both scales (Table 11.34) were high with an average reliability of 0.85 for L_AUTGOV (ranging 0.82 to 0.89) and 0.85 for L_ATTCOR (ranging from 0.83 to 0.89). Selected item parameters were used to scale the items corresponding to these two scales (Table 11.35). Given that both constructs were measured with exactly the same items as in ICCS 2009, the scales were equated so that it was possible to compare scale scores with those from the previous cycle.

Table 11.34: Reliabilities for scales measuring students' attitudes toward authoritarian and corrupt government practices

Country	Scale reliability (Cronbach's alpha)	
	L_AUTGOV	L_ATTCOR
Chile	0.88	0.88
Colombia	0.85	0.84
Dominican Republic	0.82	0.83
Mexico	0.89	0.89
Peru	0.82	0.83
ICCS 2016 average	0.85	0.85
Minimum value	0.82	0.83
Maximum value	0.89	0.89

Table 11.35: Item parameters for scales measuring students' attitudes toward authoritarian and corrupt government practices

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
L_AUTGOV	*How much do you agree or disagree with the following statements about the government and its leaders?*				
LS3G01A	It is better for government leaders to make decisions without consulting anybody	0.49	–1.25	0.82	0.43
LS3G01B	People in government must enforce their authority even if it means violating the rights of some citizens	0.34	–1.36	0.25	1.11
LS3G01C	People in government lose part of their authority when they admit their mistakes	–0.44	–1.62	–0.01	1.64
LS3G01D	People whose opinions are different than those of the government must be considered its enemies	0.54	–1.56	0.78	0.78
LS3G01E	The most important opinion of a country should be that of the president	–0.52	–1.14	0.08	1.05
LS3G01F	It is fair that the government does not comply with the law when it thinks it is not necessary	0.33	–1.23	0.39	0.85
LS3G02A	Concentration of power in one person guarantees order	–0.85	–1.73	0.06	1.67
LS3G02B	The government should close communication media that are critical	0.30	–1.69	0.40	1.29
LS3G02C	If the president does not agree with <Congress>, he/she should <dissolve> it	–0.19	–1.80	0.33	1.47
L_ATTCORR	*How much do you agree or disagree with the following statements about the public service and government?*				
LS3G03A	It is acceptable for a civil servant to accept bribes if his salary is too low	0.36	–1.69	0.81	0.88
LS3G03B	It is acceptable for a civil servant to use the resources of the institution in which he/she works for personal benefit	0.17	–1.77	0.18	1.59
LS3G03C	Good candidates grant personal benefits to voters in return for their votes	–0.08	–1.72	0.20	1.52
LS3G03D	Paying an additional amount to a civil servant in order to obtain a personal benefit is acceptable	0.13	–1.86	0.33	1.53
LS3G03E	It is acceptable that a civil servant helps his/her friends. by giving them employment in his/her office	–0.59	–1.78	–0.13	1.91
LS3G03F	Since public resources belong to everyone, it is acceptable that those who can keep part of them	0.01	–1.89	0.35	1.53

Students' attitudes toward violence and disobedience to the law

Question 4 of the Latin American student questionnaire asked students to rate their level of agreement (ranging from "strongly agree" to "strongly disagree") with ten statements relating to peace, conflict and the use of violence. We used six items of this question to derive the scale *students' endorsement of the use of violence* (L_ATTVIOL); the higher scale scores indicate more positive attitudes toward the use of violence.

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. We used nine of the ten items to derive the scale students' endorsement of disobeying the law (L_DISLAW). Higher values on this scale reflected greater agreement with the notion that it is acceptable to disobey a law under certain circumstances.

Figure 11.19: Confirmatory factor analysis of items measuring students' attitudes toward violence and disobedience to the law

Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.061	0.062	0.063	0.059
CFI	0.97	0.97	0.97	0.97
TLI	0.97	0.97	0.97	0.97

The results from a confirmatory factor analysis (Figure 11.19) showed satisfactory model fit for a two-dimensional model for the combined set of items after introducing correlations for two pairs of items which reflected more similar contents with regard to justifications for disobedience to the law (items A and B, and H and I). The two latent dimensions were strongly correlated (r = 0.56). When reviewing measurement invariance using across multiple-group models with different constraints, the results showed that model fit remained equally satisfactory with more constrained models suggesting a high degree of measurement invariance.

Scale reliabilities (Cronbach's alpha) for both scales (Table 11.36) were high with an average reliability of 0.86 for L_ATTVIOL (ranging 0.84 to 0.87) and 0.85 for L_DISLAW (ranging from 0.81 to 0.88). Selected item parameters were used to scale the items corresponding to these two scales (Table 11.37). Given that both constructs were measured with exactly the same items as in ICCS 2009, the two scales were equated to enable comparison of scale scores across the two cycles.

Table 11.36: Reliabilities for scales measuring students' attitudes toward violence and disobedience to the law

Country	Scale reliability (Cronbach's alpha)	
	L_ATTVIOL	L_DISLAW
Chile	0.87	0.88
Colombia	0.87	0.85
Dominican Republic	0.85	0.86
Mexico	0.87	0.87
Peru	0.84	0.81
ICCS 2016 average	0.86	0.85
Minimum value	0.84	0.81
Maximum value	0.87	0.88

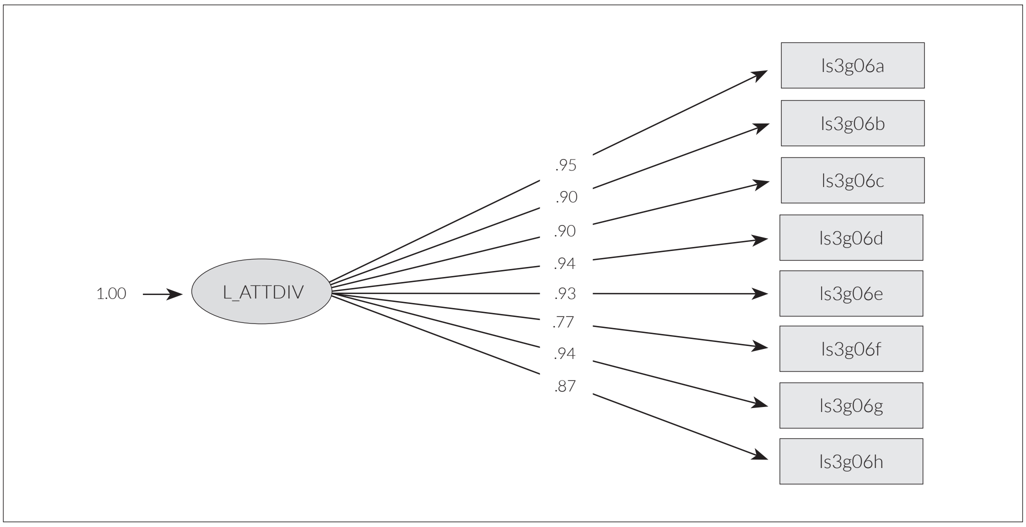
Table 11.37: Item parameters for scales measuring students' attitudes toward violence and disobedience to the law

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
L_ATTVIOL	*How much do you agree or disagree with the following statements?*				
LS3G04E	He who does me harm will have to pay for it	–0.97	–2.31	0.57	1.74
LS3G04F	Watching fights between classmates is fun	0.31	–2.21	0.95	1.26
LS3G04G	If you can't succeed by doing good things, <try> the bad ones	–0.22	–2.10	0.66	1.44
LS3G04H	You have to fight so people do not think you are a coward	0.29	–2.13	0.98	1.15
LS3G04I	Revenge is sweet	0.11	–2.00	0.74	1.26
LS3G04J	Aggression serves to achieve what one wants	0.48	–1.97	1.05	0.92
L_DISLAW	*How much do you agree or disagree with the following statements about situations where the law is disobeyed?*				
LS3G05A	when it is the only alternative left for achieving important objectives	–0.54	–1.79	0.05	1.74
LS3G05B	when it is the only way one has to help one's family	–0.80	–1.76	–0.25	2.01
LS3G05C	when others who disobeyed it were not punished	0.30	–1.97	0.57	1.40
LS3G05D	when others do it	0.58	–1.86	0.59	1.27
LS3G05E	when one distrusts the enacting body	0.19	–2.01	0.38	1.63
LS3G05F	when one is sure nobody will realize	0.51	–1.87	0.62	1.25
LS3G05H	when nobody gets hurt	–0.34	–1.53	–0.05	1.58
LS3G05I	when it is not done with bad intentions	–0.32	–1.65	–0.06	1.71
LS3G05J	when one can obtain economic benefits	0.42	–1.62	0.52	1.09

Students' acceptance of neighborhood diversity

Question 6 asked students whether they would be bothered ("yes," "no") by 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. We used the eight items associated with the question to construct the scale *students' acceptance of neighborhood diversity* (L_ATTDIV); higher scale scores correspond to higher levels of acceptance of neighborhood diversity.

Figure 11.20: Confirmatory factor analysis of items measuring students' acceptance of neighborhood diversity



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.045	0.041	n/a	0.037
CFI	1.00	1.00	n/a	1.00
TLI	1.00	1.00	n/a	1.00

Note:
n/a = not applicable

The model fit for the one-dimensional model (Figure 11.20) based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory and remained satisfactory fit with the most constrained (scalar) model.[16]

The scale reliability (Cronbach's alpha) for L_ATTDIV was very high, being 0.90 on average with coefficients ranging from 0.82 to 0.93 across participating countries (see Table 11.38). Selected item parameters were used to scale the items in this new scale for ICCS 2016 (Table 11.39). Even though most items were identical to a question used in ICCS 2009, changes to the stem and question format mean this scale cannot be used for comparisons across cycles.

16 Given the binary nature of the items it was not possible to estimate a metric model for these data.

Table 11.38: Reliabilities for scale measuring students' acceptance of neighborhood diversity

Country	Scale reliability (Cronbach's alpha)
	L_ATTDIV
Chile	0.92
Colombia	0.92
Dominican Republic	0.90
Mexico	0.93
Peru	0.82
ICCS 2016 average	0.90
Minimum value	0.82
Maximum value	0.93

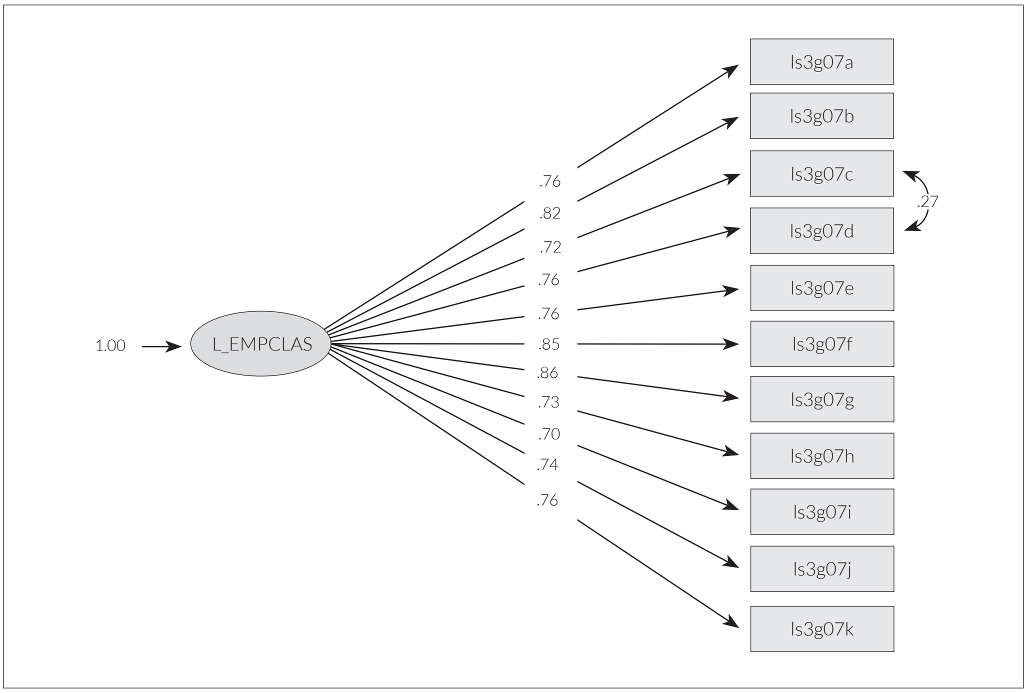
Table 11.39: Item parameters for scale measuring students' acceptance of neighborhood diversity

Scale or item	Question/item wording	Item parameters
		Delta
L_ATTDIV	*Would it bother you having neighbours belonging to the following groups?*	
LS3G06A	Persons with different skin colour than yours	–0.23
LS3G06B	Persons of a different social class than yours	0.03
LS3G06C	Persons of a different religion than yours	0.23
LS3G06D	Persons who come from another region of the country	–0.23
LS3G06E	Persons with physical disabilities	–0.30
LS3G06F	Persons with mental disorders	0.66
LS3G06G	Persons from a different country	–0.13
LS3G06H	Persons of indigenous origin	–0.03

Students' feelings of empathy toward classmates

Question 7 presented students with 11 items describing 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"). We used all 11 items to derive the scale *students' feelings of empathy toward classmates* (L_EMPCLAS). Students who scored highly on this scale were students who expressed a greater degree of empathy toward their classmates.

Figure 11.21: Confirmatory factor analysis of items measuring students' feelings of empathy toward classmates



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.051	0.053	0.047	0.052
CFI	0.98	0.98	0.98	0.98
TLI	0.98	0.98	0.98	0.98

The model fit for the one-dimensional model (Figure 11.21) based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory and remained satisfactory fit for more constrained models. Given the highly similar content of two items (c and d) we included a correlation of their respective residual terms.

The scale reliability (Cronbach's alpha) for L_EMPCLAS was very high, being 0.89 on average with coefficients ranging from 0.86 to 0.91 across participating countries (see Table 11.40). Selected item parameters were used to scale the items in this new scale for ICCS 2016 (Table 11.41). Even though there are items where the question is similar to one from the ICCS 2009 Latin American student questionnaire, it is not possible to make comparisons due to changes to the stem and category wording.

Table 11.40: Reliabilities for scale measuring students' feelings of empathy toward classmates

Country	Scale reliability (Cronbach's alpha)
	L_EMPCLAS
Chile	0.89
Colombia	0.86
Dominican Republic	0.91
Mexico	0.91
Peru	0.86
ICCS 2016 average	0.89
Minimum value	0.86
Maximum value	0.91

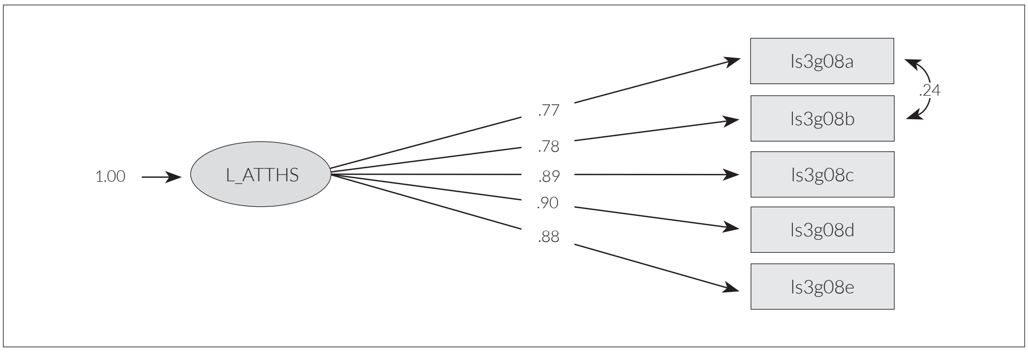
Table 11.41: Item parameters for scale measuring students' feelings of empathy toward classmates

Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
L_EMPCLAS	*How do you feel when you witness the following situations at your school?*			
LS3G07A	A classmate falls and gets hurt	0.36	–0.43	0.43
LS3G07B	A classmate gets beaten up	–0.41	–1.16	1.16
LS3G07C	A classmate gets unfairly reprimanded	0.23	–1.21	1.21
LS3G07D	A classmate gets unfairly punished	–0.05	–1.19	1.19
LS3G07E	A classmate gets something stolen from him/her	–0.46	–1.29	1.29
LS3G07F	A classmate gets ridiculed	0.18	–1.13	1.13
LS3G07G	A classmate gets insulted	–0.09	–1.39	1.39
LS3G07H	A classmate looks very sad	–0.56	–1.34	1.34
LS3G07I	A classmate gets bad grades	0.46	–2.09	2.09
LS3G07J	A classmate has nobody to play with	–0.06	–1.56	1.56
LS3G07K	There is a fight between classmates	0.40	–0.90	0.90

Students' attitudes toward homosexuality

Question 8 of the Latin American student questionnaire asked students to rate their level of agreement (ranging from "strongly agree" to "strongly disagree") with five statements relating to the acceptance of homosexuality. We used these question items to derive the scale *students' acceptance of homosexuality* (L_ATTHS); higher scale scores indicate higher levels of acceptance of equal rights for homosexuals.

Figure 11.22: Confirmatory factor analysis of items measuring students' acceptance of homosexuality



Model fit indices:	Pooled sample	Multiple-group models		
		Configural	Metric	Scalar
RMSEA	0.038	0.061	0.046	0.060
CFI	1.00	1.00	1.00	1.00
TLI	1.00	1.00	1.00	1.00

After including a correlation between the residual error terms between two items (a and b) with related content, the model fit for the one-dimensional model (Figure 11.22) based on data from the pooled ICCS 2016 dataset with equally weighted countries was satisfactory. When reviewing measurement invariance across models with different constraints, results showed that the most constrained (scalar) model had somewhat less satisfactory fit, but this was still within an acceptable range.

The scale reliability (Cronbach's alpha) for L_ATTHS was very high, being 0.88 on average with coefficients ranging from 0.84 to 0.94 across participating countries (see Table 11.42). Table 11.43 shows the item parameters for this new ICCS 2016 scale.

Table 11.42: Reliabilities for scale measuring students' acceptance of homosexuality

Country	Scale reliability (Cronbach's alpha)
	L_ATTHS
Chile	0.94
Colombia	0.87
Dominican Republic	0.84
Mexico	0.89
Peru	0.87
ICCS 2016 average	0.88
Minimum value	0.84
Maximum value	0.94

Table 11.43: Item parameters for scale measuring students' acceptance of homosexuality

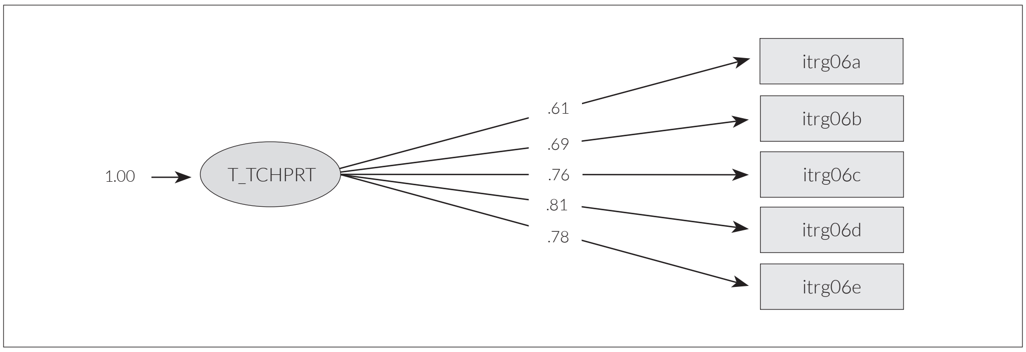
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
L_ATTHS	*How much do you agree or disagree with the following statements with regard to homosexual orientations?*				
LS3G08A	Persons of the same sex should have the right to get married	0.42	–1.32	–0.45	1.76
LS3G08B	Two persons of the same sex should have the right to adopt children	0.50	–1.48	–0.45	1.93
LS3G08C	Homosexuals should have the same rights as all other citizens	–0.64	–1.27	–0.71	1.97
LS3G08D	All schools should accept homosexuals	–0.18	–1.34	–0.64	1.98
LS3G08E	Homosexuals should have the right to hold any political or public position	–0.10	–1.41	–0.56	1.96

Teacher questionnaire

Teachers' perceptions of teacher participation at school

Question 6 of the teacher questionnaire required teachers to rate how many teachers in their school have participated in five different activities related to a teacher's cooperation with the running of their school ("All of them," "Most of them," "Some of them," "None or hardly any"). We used these items to derive the scale *teachers' perception of teacher participation at school* (T_TCHPART); higher scale scores indicate higher levels of teacher participation.

Figure 11.23: Confirmatory factor analysis of items measuring teachers' perceptions of teacher participation at school



Model fit indices:	Pooled sample
RMSEA	0.049
CFI	0.99
TLI	0.98

The model fit for the one-dimensional model (Figure 11.23) based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory.

The scale reliability (Cronbach's alpha) for T_TCHPART was generally satisfactory, being 0.77 on average with coefficients ranging from 0.63 to 0.85 across participating countries (see Table 11.44). Selected item parameters were used to scale the items in this new scale for ICCS 2016 (Table 11.45).

Table 11.44: Reliabilities for scale measuring teachers' perceptions of teacher participation at school

Country	Scale reliability (Cronbach's alpha)
	T_TCHPART
Belgium (Flemish)	0.74
Bulgaria	0.78
Chile	0.83
Chinese Taipei	0.77
Colombia	0.81
Croatia	0.79
Denmark	0.79
Dominican Republic	0.77
Estonia	0.68
Finland	0.80
Italy	0.81
Korea, Republic of	0.85
Latvia	0.72
Lithuania	0.72
Malta	0.75
Mexico	0.85
Netherlands	0.63
Norway	0.71
Peru	0.79
Russian Federation	0.76
Slovenia	0.77
Sweden	0.73
ICCS average	0.77
Minimum value	0.63
Maximum value	0.85

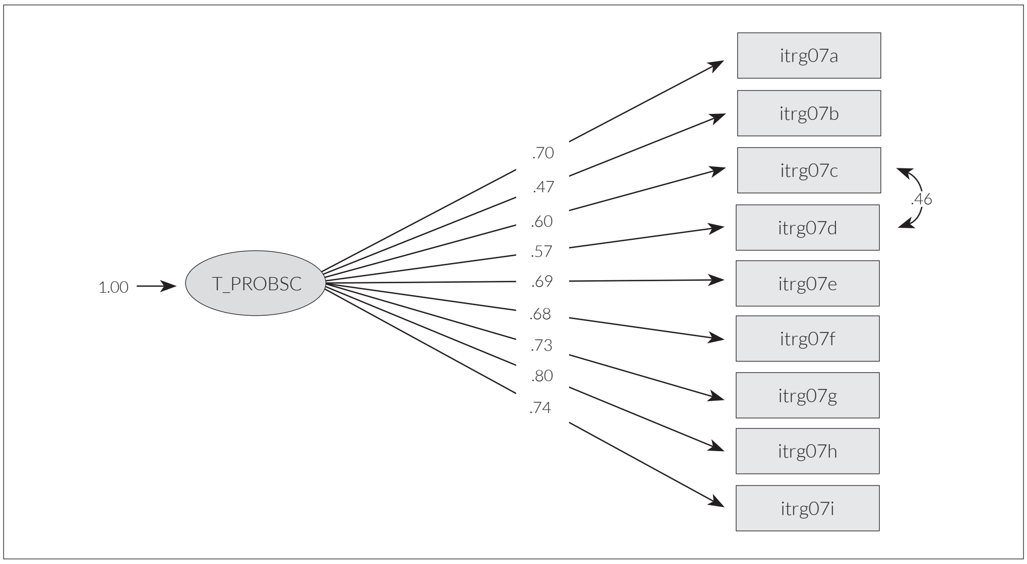
Table 11.45: Item parameters for scale measuring teachers' perceptions of teacher participation at school

Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
T_TCHPRT	*Teachers' perceptions of teacher participation at school*			
IT3G06A	Working with one another in devising teaching activities	0.05	–0.92	0.92
IT3G06B	Helping in solving conflict situations arising among students in the school	–0.49	–1.24	1.24
IT3G06C	Taking on tasks and responsibilities in addition to teaching (tutoring, school projects, etc.)	0.31	–1.04	1.04
IT3G06D	Actively taking part in school <development/ improvement activities>	–0.01	–1.08	1.08
IT3G06E	Engaging in <guidance and counselling activities>	0.14	–0.80	0.80

Teachers' perceptions of social problems at school

Question 7 required teachers to indicate the frequency with which a series of anti-social behaviors occurred among students in their school (ranging from "Never" to "Very often"). These items were used to derive the scale *teachers' perceptions of social problems at school* (T_PROBSC) for which the higher scale scores indicate higher perceived social problems.

Figure 11.24: Confirmatory factor analysis of items measuring teachers' perceptions of social problems at school



Model fit indices:	Pooled sample
RMSEA	0.076
CFI	0.95
TLI	0.92

The model fit for the one-dimensional model (Figure 11.24) based on data from the pooled ICCS 2016 dataset with equally weighted countries was acceptable, once the residual correlation of 0.46 between two items reflecting highly related problems at school (religious and ethnic intolerance) was taken into account.

The scale reliability (Cronbach's alpha) for T_PROBSC was good on average (0.76) and across countries with coefficients ranging from 0.70 to 0.83 across participating countries (see Table 11.46). Selected item parameters were used to scale the items corresponding to this new scale for ICCS 2016 (Table 11.47).

Table 11.46: Reliabilities for scale measuring teachers' perceptions of social problems at school

Country	Scale reliability (Cronbach's alpha)
	T_PROBSC
Belgium (Flemish)	0.78
Bulgaria	0.80
Chile	0.82
Chinese Taipei	0.79
Colombia	0.80
Croatia	0.79
Denmark	0.77
Dominican Republic	0.75
Estonia	0.72
Finland	0.74
Italy	0.70
Korea, Republic of	0.74
Latvia	0.73
Lithuania	0.72
Malta	0.82
Mexico	0.83
Netherlands	0.75
Norway	0.76
Peru	0.77
Russian Federation	0.71
Slovenia	0.77
Sweden	0.79
ICCS average	0.76
Minimum value	0.70
Maximum value	0.83

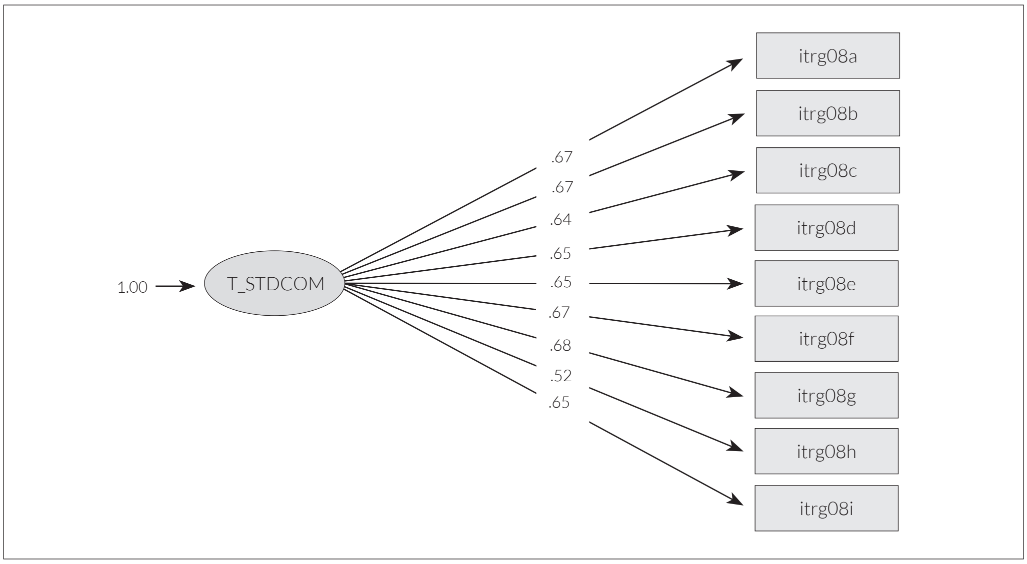
Table 11.47: Item parameters for scale measuring teachers' perceptions of social problems at school

Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
T_PROBSC	*Please indicate how frequently each of the following problems occurs among students at this school.*			
IT3G07A	Vandalism	–0.36	–1.93	1.93
IT3G07B	Truancy	–2.17	–2.36	2.36
IT3G07C	Ethnic intolerance	0.18	–1.80	1.80
IT3G07D	Religious intolerance	0.85	–1.55	1.55
IT3G07E	<Bullying>	–1.61	–2.26	2.26
IT3G07F	Violence	–0.67	–2.12	2.12
IT3G07G	Sexual harassment	1.51	–1.95	1.95
IT3G07H	Drug abuse	1.21	–1.50	1.50
IT3G07I	Alcohol abuse	1.05	–1.69	1.69

Teachers' perceptions of student activities in the community

Question 8 of the Teacher questionnaire asked teachers to indicate ("Yes" or "No") whether their students had taken part in activities carried out by the school in cooperation with external groups or organizations. The nine items in this question were used to derive the scale *teachers' perceptions of student activities in the community* (T_STDCOM); higher scale scores indicate greater levels of participation in the community.

Figure 11.25: Confirmatory factor analysis of items measuring teachers' perceptions of student activities in the community



Model fit indices:	Pooled sample
RMSEA	0.042
CFI	0.98
TLI	0.97

The model fit for the one-dimensional model (Figure 11.25) based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory.

The scale reliability (Cronbach's alpha) for T_STDCOM was good, being 0.72 on average with coefficients ranging from 0.63 to 0.81 across participating countries (see Table 11.48). Selected item parameters were used to scale the items corresponding to this new scale for ICCS 2016 (Table 11.49).

Table 11.48: Reliabilities for scale measuring teachers' perceptions of student activities in the community

Country	Scale reliability (Cronbach's alpha)
	T_STDCOM
Belgium (Flemish)	0.70
Bulgaria	0.71
Chile	0.76
Chinese Taipei	0.79
Colombia	0.71
Croatia	0.75
Denmark	0.63
Dominican Republic	0.78
Estonia	0.63
Finland	0.63
Italy	0.74
Korea, Republic of	0.78
Latvia	0.75
Lithuania	0.74
Malta	0.81
Mexico	0.69
Netherlands	0.66
Norway	0.66
Peru	0.69
Russian Federation	0.78
Slovenia	0.74
Sweden	0.71
ICCS average	0.72
Minimum value	0.63
Maximum value	0.81

Table 11.49: Item parameters for scale measuring teachers' perceptions of student activities in the community

Scale or item	Question/item wording	Item parameters
		Delta
T_STDCOM	*During the current school year, have you and your <target grade> students taken part in any of these activities?*	
IT3G08A	Activities related to environmental sustainability (e.g. <energy and water saving, recycling>)	–0.52
IT3G08B	Human rights projects	0.67
IT3G08C	Activities for underprivileged people or groups	0.52
IT3G08D	Cultural activities (e.g. theatre, music)	–1.60
IT3G08E	Multicultural and intercultural activities within the <local community> (e.g. <promotion and celebration of cultural diversity, food street market>)	0.20
IT3G08F	Campaigns to raise people's awareness, such as <campaigns to raise people's awareness about social issues, campaigns to raise people's awareness of environmental issues>	–0.30
IT3G08G	Activities aimed at protecting the cultural heritage in the <local community>	0.47
IT3G08H	Visits to political institutions (e.g. <Parliament house, Prime Minister's/President's official residence>)	2.07
IT3G08I	Sports events	–1.51

Teachers' perceptions of student behavior, classroom climate and bullying at school

In Question 9 of the teacher questionnaire, respondents were asked to indicate how many students in the school exhibited six different behavior types (response options included "all or nearly all," "most of them" "some of them," and "none or hardly any"). These six items were used to derive a scale of *teachers' perceptions of student behavior at school* (T_STUDB); higher scale scores indicated a greater proportion of students exhibited positive behaviors.

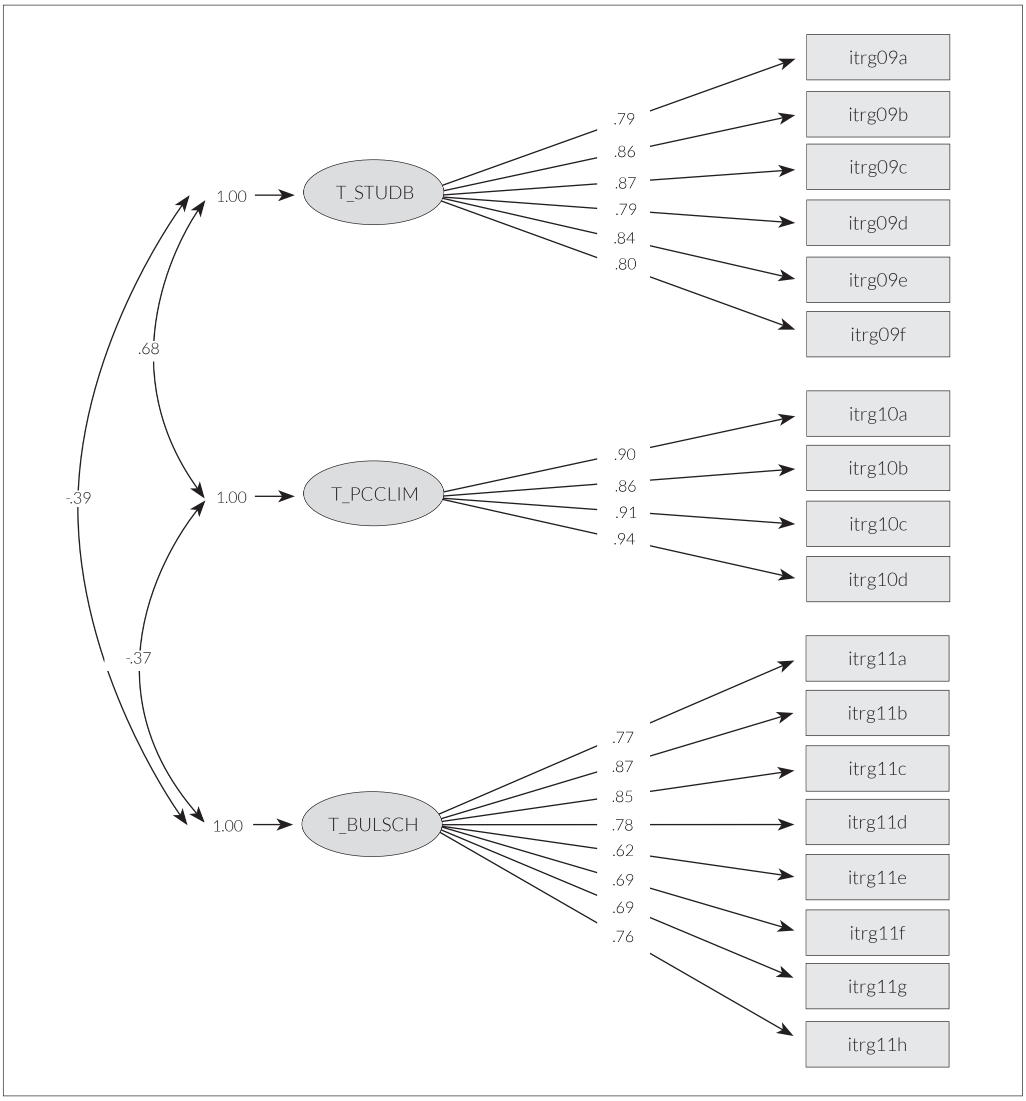
Similarly, teachers were also asked to indicate the proportion of their target grade students who exhibited positive interactions with their classmates (Question 10) using the same set of response options as the previous question. The four items in the question were used to derive the scale *teachers' perceptions of classroom climate* (T_PCCLIM), where a higher scale score corresponded to a more positive climate in the classroom.

In Question 11, teachers were asked to respond to how often eight situations involving bullying occurred in the school year ("never," "less than once a month," "1 to 5 times a month," and "more than 5 times a month"). All eight items were used to derive the scale for *teachers' perceptions of bullying at school* (T_BULSCH). A higher scale score indicated that occurrences of bullying in school were more frequent.

The results from the confirmatory factor analyses assume a three-dimensional model for these items (Figure 11.26). The model fit based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory. While there was a strong positive correlation between T_STUDB and T_PCCLIM, T_BULSCH had moderate negative correlations with the other two latent factors.

The reliabilities (Cronbach's alpha) for each of the three scales were good (see Table 11.50). T_STUDB had an average reliability of 0.86 (with national coefficients ranging from 0.78 to 0.91), T_PCCLIM had an average reliability of 0.89 (with national coefficients ranging from 0.80 to 0.91) and T_BULSCH had an average reliability of 0.83 (with national coefficients ranging from 0.77 to 0.87). Selected item parameters were used to scale the items in these three new scales for ICCS 2016 (Table 11.51).

Figure 11.26: Confirmatory factor analysis of items measuring teachers' perceptions of student behavior, classroom climate and bullying at school



Model fit indices:	Pooled sample
RMSEA	0.035
CFI	0.95
TLI	0.92

Table 11.50: Reliabilities for scales measuring teachers' perceptions of student behavior, classroom climate and bullying at school

Country	Scale reliability (Cronbach's alpha)		
	T_STUDB	T_PCCLIM	T_BULSCH
Belgium (Flemish)	0.83	0.91	0.80
Bulgaria	0.90	0.88	0.86
Chile	0.91	0.90	0.87
Chinese Taipei	0.90	0.91	0.85
Colombia	0.86	0.87	0.87
Croatia	0.86	0.89	0.85
Denmark	0.85	0.90	0.79
Dominican Republic	0.85	0.80	0.84
Estonia	0.87	0.91	0.84
Finland	0.82	0.87	0.77
Italy	0.87	0.89	0.81
Korea, Republic of	0.90	0.91	0.86
Latvia	0.85	0.90	0.87
Lithuania	0.83	0.89	0.84
Malta	0.90	0.91	0.81
Mexico	0.87	0.84	0.83
Netherlands	0.78	0.87	0.78
Norway	0.86	0.89	0.82
Peru	0.86	0.84	0.86
Russian Federation	0.88	0.91	0.82
Slovenia	0.87	0.88	0.83
Sweden	0.86	0.90	0.82
ICCS 2016 average	0.86	0.89	0.83
Minimum value	0.78	0.80	0.77
Maximum value	0.91	0.91	0.87

Table 11.51: Item parameters for scales measuring teachers' perceptions of student behavior, classroom climate and bullying at school

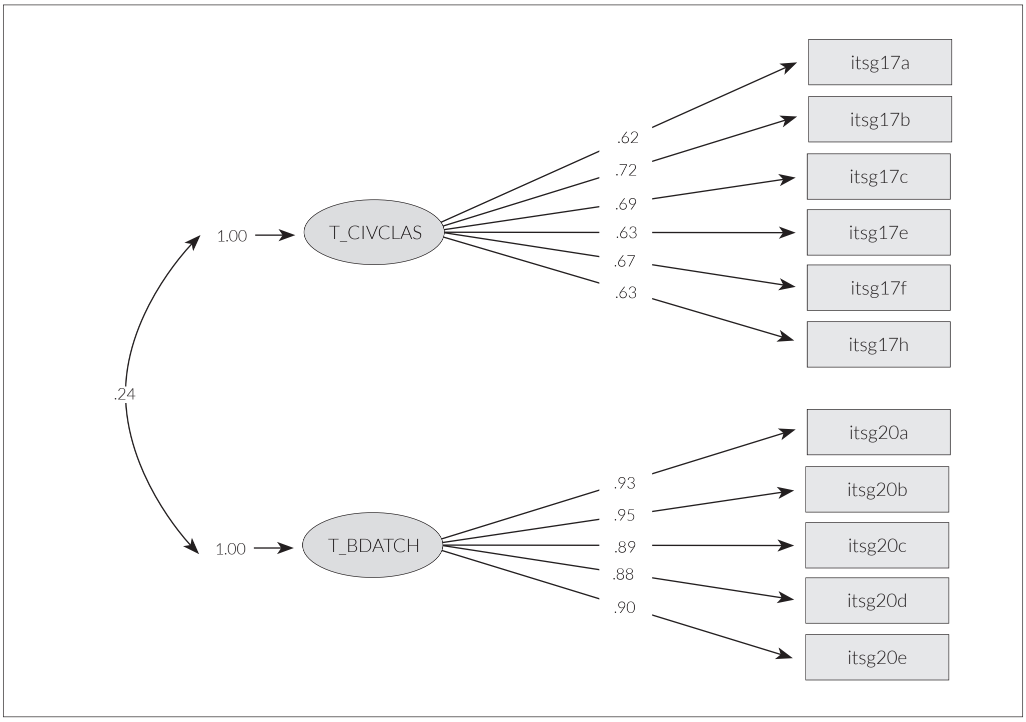
Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
T_STUDB	*In your opinion, how many students in this school …*			
IT3G09A	are well behaved on entering and leaving the school premises?	–0.35	–3.13	3.13
IT3G09B	have a positive attitude towards their own school?	0.22	–3.18	3.18
IT3G09C	have a good relationship with the school teachers and staff?	–0.92	–3.42	3.42
IT3G09D	show care for school facilities and equipment?	0.77	–3.03	3.03
IT3G09E	are well behaved during breaks?	0.15	–3.39	3.39
IT3G09F	show they feel part of the school community?	0.12	–2.68	2.68
T_PCCLIM	*In your opinion, how many of your <target grade> students …*			
IT3G10A	get on well with their classmates?	–0.59	–4.52	4.52
IT3G10B	are well integrated in the class?	–0.13	–4.18	4.18
IT3G10C	respect their classmates even if they are different?	0.80	–3.66	3.66
IT3G10D	have a good relationship with other students?	–0.09	–4.36	4.36
T_BULSCH	*How often have any of the following situations happened during the current school year?*			
IT3G11A	A student informed you about aggressive or destructive behaviours by other students	–1.43	–1.90	1.90
IT3G11B	A student informed you that s/he was <bullied> by another student	–0.93	–1.82	1.82
IT3G11C	A teacher informed you that a student was <bullied> by other students	–0.70	–1.86	1.86
IT3G11D	A teacher informed you that a student helped another student who was being <bullied>	0.21	–1.71	1.71
IT3G11E	A student informed you that s/he was <bullied> by a teacher	1.76	–1.21	1.21
IT3G11F	A parent informed you that his/her son/daughter was <bullied> by other students	0.66	–1.97	1.97
IT3G11G	A teacher informed you that s/he was <bullied> by students	0.95	–1.15	1.15
IT3G11H	You witnessed students' <bullying> behaviors	–0.50	–1.43	1.43

Teachers' reports on civic-related activities in class and professional development activities for teaching methods

The teacher questionnaire included a set of questions to be administered only to teachers who taught civic and citizenship education related subjects at the target grade. In Question 17, this group of teachers were asked how often activities relating to civic and citizenship education took place in their target grade lessons ("Never," "Sometimes," "Rarely," and "Often"). Six out of the eight items in this question were used to derive a scale for *teachers' reports on civic-related activities in class* (T_CIVCLAS). Higher scale scores corresponded to a greater frequency of this type of activity occurring in their target grade lessons.

In Question 20, the same group of teachers were asked whether they had attended any teacher training courses addressing different teaching methods and approaches, and, if so, whether this was pre-service and/or in-service training. There were five items in the question, and all were used to derive a scale of *teachers' PD (professional development) activities for teaching methods* (T_PDATCH). Higher T_PDATCH scores corresponded to greater attendance to a range of different PD activities in this area.

Figure 11.27: Confirmatory factor analysis of items measuring teachers' reports on civic-related activities in class and PD activities for teaching methods



Model fit indices:	Pooled sample
RMSEA	0.035
CFI	0.98
TLI	0.98

The model fit for the two-dimensional model (Figure 11.27) based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory, and there was a moderate positive correlation of 0.24 between these two factors.

The scale reliability (Cronbach's alpha) for T_CIVCLAS was good, being 0.72 on average with coefficients ranging from 0.57 to 0.83 across participating countries (see Table 11.52). The scale reliability for T_PDATCH was also good; 0.83 on average, ranging from 0.74 to 0.92. Selected item parameters were used to scale the items corresponding to these two scales for ICCS 2016 (Table 11.53).

Table 11.52: Reliabilities for scales measuring teachers' reports of civic-related activities in class and PD activities for teaching methods

Country	Scale reliability (Cronbach's alpha)	
	T_CIVCLAS	T_PDATCH
Belgium (Flemish)	0.69	0.76
Bulgaria	0.83	0.88
Chile	0.79	0.89
Chinese Taipei	0.79	0.77
Colombia	0.78	0.83
Croatia	0.79	0.91
Denmark	0.57	0.79
Dominican Republic	0.81	0.85
Estonia	0.74	0.78
Finland	0.70	0.84
Italy	0.71	0.81
Korea, Republic of	0.72	0.83
Latvia	0.74	0.74
Lithuania	0.74	0.81
Malta	0.68	0.91
Mexico	0.67	0.85
Netherlands	0.69	0.75
Norway	0.60	0.87
Peru	0.65	0.86
Russian Federation	0.77	0.92
Slovenia	0.75	0.77
Sweden	0.68	0.86
ICCS 2016 average	0.72	0.83
Minimum value	0.57	0.74
Maximum value	0.83	0.92

Table 11.53: Item parameters for scales measuring teachers' reports of civic-related activities in class and PD activities for teaching methods

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
T_CIVCLAS	*How often do the following activities take place during your <target grade> lessons related to <civic and citizenship education>?*				
IT3G17A	Students work on projects that involve gathering information outside school (e.g. interviews in the neighborhood, small scale surveys)	1.01	–2.26	0.86	1.40
IT3G17B	Students work in small groups on different topics/issues	–0.63	–2.71	0.44	2.27
IT3G17C	Students participate in role plays	0.58	–2.24	0.55	1.69
IT3G17E	Students discuss current issues	–1.59	–2.69	0.27	2.42
IT3G17F	Students research and/or analyze information gathered from multiple Internet sources (e.g. wikis, online newspapers)	–0.25	–2.49	0.53	1.96
IT3G17H	Students propose topics/issues for the following lessons	0.89	–2.57	0.87	1.70
T_PDATCH	*Have you attended any teacher training courses addressing the following teaching methods and approaches?*				
IT3G20A	Pair and group work	–0.91			
IT3G20B	Classroom discussion	–0.10			
IT3G20C	Role play	0.52			
IT3G20D	Research work	0.62			
IT3G20E	Problem solving	–0.13			

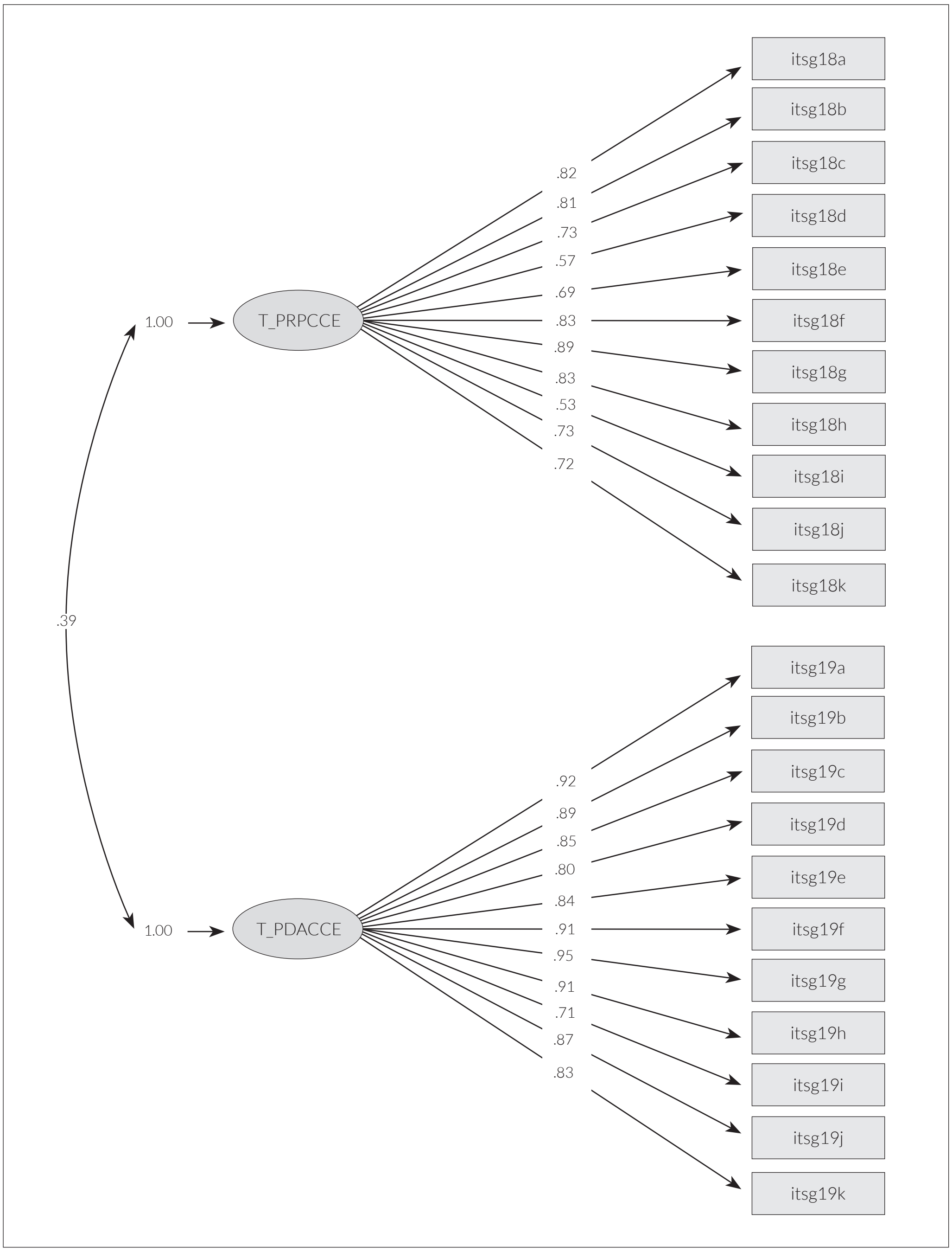
Teachers' reports on their confidence to teach civic and citizenship education topics and participation in related professional development activities

The subset of teachers who taught a civic and citizenship education (CCE) related subject at the target grade were asked in Question 18 of the teacher questionnaire how prepared they felt about teaching topics related to civic and citizenship education, one of which was optional for European countries. The 11 international items were used to derive the scale *teachers' preparedness for teaching CCE topics* (T_PRPCCE). Higher scale scores correspond to a greater sense of preparedness.

Question 19 was a new question developed for ICCS 2016 that required teachers to indicate whether they had attended training courses for the same topics listed in the previous question, and, if so, whether this was pre-service and/or in-service training. The 11 international items were used to derive the scale *teachers' PD activities for CCE topics* (T_PDACCE).

The model fit for the two-dimensional model (Figure 11.28) based on data from the pooled ICCS 2016 dataset with equally weighted countries was highly satisfactory. The two latent factors has a positive correlation of 0.39.

Figure 11.28: Confirmatory factor analysis of items measuring teachers' reports of their confidence to teach CCE topics and participation in CCE PD activities



Model fit indices:	Pooled sample
RMSEA	0.045
CFI	0.96
TLI	0.96

The scale reliabilities (Cronbach's alpha) for both T_PRPCCE and T_PDACCE were both high on average across participating countries (see Table 11.54). The average reliability for T_PRPCCE was 0.88 (ranging from 0.84 to 0.92) and for T_PDACCE it was 0.90 (ranging from 0.84 to 0.94). The item parameters for both scales are presented in Table 11.55.

Table 11.54: Reliabilities for scales measuring teachers' reports of their confidence to teach CCE topics and participation in CCE PD activities

Country	Scale reliability (Cronbach's alpha)	
	T_PRPCCE	T_PDACCE
Belgium (Flemish)	0.85	0.88
Bulgaria	0.90	0.92
Chile	0.92	0.93
Chinese Taipei	0.89	0.89
Colombia	0.90	0.89
Croatia	0.90	0.93
Denmark	0.86	0.90
Dominican Republic	0.87	0.93
Estonia	0.89	0.92
Finland	0.86	0.87
Italy	0.87	0.89
Korea, Republic of	0.92	0.93
Latvia	0.84	0.84
Lithuania	0.86	0.86
Malta	0.84	0.89
Mexico	0.88	0.89
Netherlands	0.85	0.86
Norway	0.88	0.94
Peru	0.91	0.89
Russian Federation	0.89	0.91
Slovenia	0.87	0.90
Sweden	0.88	0.93
ICCS 2016 average	0.88	0.90

Table 11.55: Item parameters for scales measuring teachers' reports of their confidence to teach CCE topics and participation in CCE PD activities

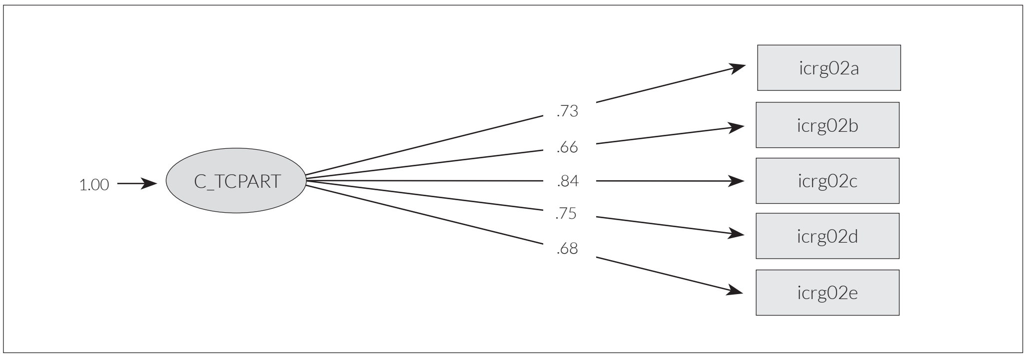
Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
T_PRPCCE	*How well prepared do you feel to teach the following topics and skills?*			
IT3G18A	Human rights	–0.18	–1.78	1.78
IT3G18B	Voting and elections	0.05	–1.33	1.33
IT3G18C	The global community and international organizations	1.05	–1.41	1.41
IT3G18E	Emigration and immigration	0.40	–1.29	1.29
IT3G18F	Equal opportunities for men and women	–0.59	–1.73	1.73
IT3G18H	The constitution and political systems	0.47	–1.13	1.13
IT3G18I social media)	Responsible Internet use (e.g. privacy, source reliability,	0.21	–1.38	1.38
IT3G18J	Critical and independent thinking	–0.52	–1.63	1.63
IT3G18K	Conflict resolution	–0.26	–1.66	1.66
T_PDACCE	*Have you attended any teacher training courses addressing the following topics and skills?*			
IT3G19A	Human rights	–0.25		
IT3G19B	Voting and elections	0.75		
IT3G19C	The global community and international organizations	0.89		
IT3G19D	The environment and environmental sustainability	–0.32		
IT3G19E	Emigration and immigration	0.82		
IT3G19F	Equal opportunities for men and women	0.12		
IT3G19H	The constitution and political systems	0.46		
IT3G19I	Responsible Internet use (e.g. privacy, source reliability, social media)	–0.59		
IT3G19J	Critical and independent thinking	-0.56		
IT3G19K	Conflict resolution	-0.92		

School questionnaire

Principals' perceptions of teacher participation in school governance

Question 2 of the school questionnaire required principals to indicate the proportion of teachers who participated in different aspects of the running of the school ("All of them," "Most of them," "Some of them," "None or hardly any"). We used these items to derive the scale *principals' perceptions of teacher participation in school governance* (C_TCPART); higher scale scores indicated higher levels of teacher participation.

Figure 11.29: Confirmatory factor analysis of items measuring principals' perceptions of teacher participation in school governance



Model fit indices:	Pooled sample
RMSEA	0.069
CFI	0.98
TLI	0.96

The model fit for the one-dimensional model (Figure 11.29) based on data from the pooled ICCS 2016 dataset with equally weighted countries was acceptable. The scale reliability (Cronbach's alpha) for C_TCPART was good, being 0.72 on average with coefficients ranging from 0.48 to 0.86 across participating countries (see Table 11.56). Selected item parameters were used to scale the items corresponding to this new scale for ICCS 2016 (Table 11.57).

Table 11.56: Reliabilities for scale measuring principals' perceptions of teacher participation in school governance

Country	Scale reliability (Cronbach's alpha)
	C_TCPART
Belgium (Flemish)	0.61
Bulgaria	n/a
Chile	0.84
Chinese Taipei	0.80
Colombia	0.86
Croatia	0.68
Denmark	0.60
Dominican Republic	0.82
Estonia	0.77
Finland	0.64
Hong Kong SAR	0.69
Italy	0.73
Korea, Republic of	0.82
Latvia	0.76
Lithuania	0.63
Malta	0.69
Mexico	0.84
Netherlands	0.48
North-Rhine Westphalia (Germany)*	0.48
Norway	0.66
Peru	0.85
Russian Federation	0.77
Slovenia	0.78
Sweden	0.74
ICCS 2016 average	0.72
Minimum value	0.48
Maximum value	0.86

Note:
*Benchmarking participant.
n/a = not applicable.

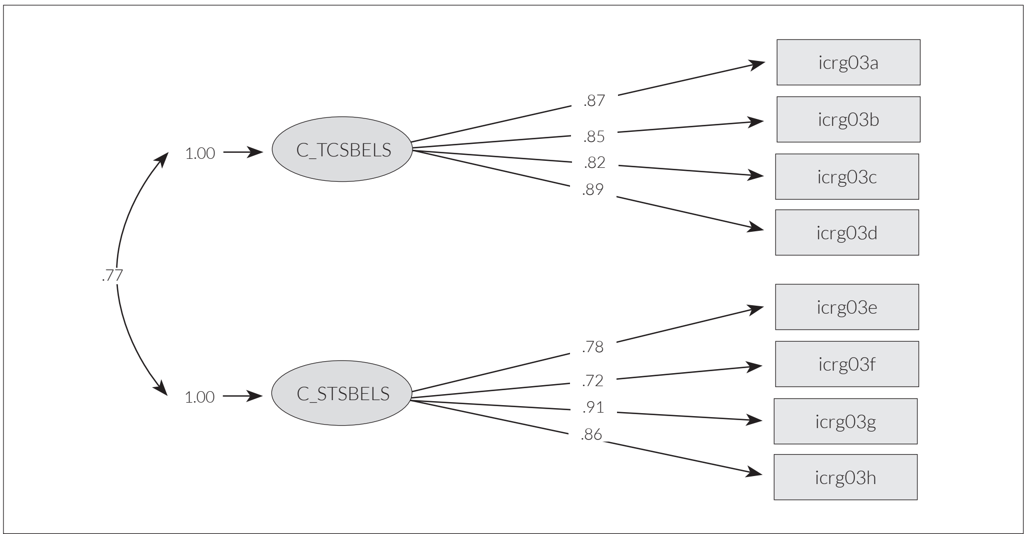
Table 11.57: Item parameters for scale measuring principals' perceptions of teacher participation in school governance

Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
C_TCPART	*In your opinion, how many teachers participate as follows at this school?*			
IC3G02A	Making useful suggestions for improving school governance	0.84	–1.27	1.27
IC3G02B	Supporting good discipline throughout the school	–1.21	–1.68	1.68
IC3G02C	Actively taking part in school <development/improvement activities>	–0.31	–1.67	1.67
IC3G02D	Encouraging students' active participation in school life	–0.82	–1.66	1.66
IC3G02E	Being willing to be members of the <school council, school governing board> as teacher representatives	1.50	–0.53	0.53

Principals' perceptions of teachers' and students' sense of belonging at school

In Question 3, principals were asked to rate the extent to which a series of statements relating students' and teachers' sense of belonging described the situation in schools (response options included "To a large extent," "To a moderate extent," "To a small extent," and "Not at all"). Four items in this question related to students, and were used to derive a scale *principals' perceptions of students' sense of belonging at school* (C_STSBELS). The remaining four items in the question related to teachers, and were used to derive a scale *principals' perceptions of teachers' sense of belonging at school* (C_TCSBELS).

Figure 11.30: Confirmatory factor analysis of items measuring principals' perceptions of students' and teachers' sense of belonging at school



Model fit indices:	Pooled sample
RMSEA	0.038
CFI	0.99
TLI	0.99

The results from confirmatory factor analyses showed a close model fit (i.e., $p < 0.05$) for a two-dimensional model for the combined set of items (Figure 11.30). The two latent dimensions were strongly correlated ($r = 0.77$). The scale reliabilities (Cronbach's alpha) for both scale were satisfactory: the average reliability for C_TCSBELS was 0.79 (with national coefficients ranging from 0.63 to 0.88 across participating countries), and for C_STSBELS it was 0.79 (with coefficients ranging from 0.65 to 0.86) (see Table 11.58). Selected item parameters were used to scale the items corresponding to this new scale for ICCS 2016 (Table 11.59).

Table 11.58: Reliabilities for scales measuring principals' perceptions of students' and teachers' sense of belonging at school

Country	Scale reliability (Cronbach's alpha)	
	C_TCSBELS	C_STSBELS
Belgium (Flemish)	0.82	0.79
Bulgaria	0.75	0.84
Chile	0.88	0.82
Chinese Taipei	0.87	0.84
Colombia	0.80	0.81
Croatia	0.85	0.86
Denmark	0.82	0.82
Dominican Republic	0.71	0.72
Estonia	0.63	0.78
Finland	0.67	0.72
Hong Kong SAR	0.85	0.83
Italy	0.80	0.81
Korea, Republic of	0.88	0.80
Latvia	0.71	0.74
Lithuania	0.76	0.77
Malta	0.85	0.74
Mexico	0.82	0.80
Netherlands	0.79	0.76
North-Rhine Westphalia (Germany)*	0.77	0.81
Norway	0.72	0.65
Peru	0.84	0.75
Russian Federation	0.74	0.79
Slovenia	0.83	0.78
Sweden	0.88	0.82
ICCS 2016 average	0.79	0.79
Minimum value	0.63	0.65
Maximum value	0.88	0.86

Note:
*Benchmarking participant.

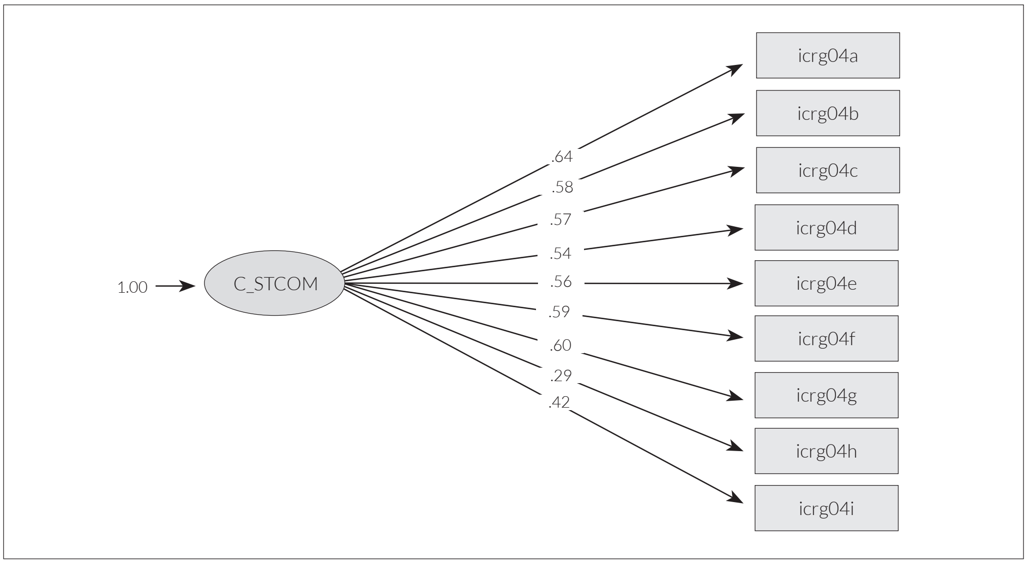
Table 11.59: Item parameters for scales measuring principals' perceptions of students' and teachers' sense of belonging at school

Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
C_TCSBELS	*In your opinion, to what extent do the following statements describe the current situation at this school?*			
IC3G03A	Teachers have a positive attitude towards the school	–0.95	–2.95	2.95
IC3G03B	Teachers feel part of the school community	–0.11	–2.47	2.47
IC3G03C	Teachers work with enthusiasm	0.86	–3.16	3.16
IC3G03D	Teachers take pride in this school	0.20	–2.58	2.58
C_STSBELS	*In your opinion, to what extent do the following statements describe the current situation at this school?*			
IC3G03E	Students enjoy being in school	–0.75	–2.93	2.93
IC3G03F	Students are actively involved in school work	0.76	–2.64	2.64
IC3G03G	Students take pride in this school	0.06	–2.66	2.66
IC3G03H	Students feel part of the school community	–0.07	–2.33	2.33

Principals' perceptions of student opportunities to participate in community activities

Question 4 of the principal questionnaire asked principals indicate how many target grade students had the opportunity to take part in different activities in the community ("all or nearly all," "most of them" "some of them," and "none or hardly any"). We used all nine of the items of this question to derive the scale *principals' perceptions of student opportunities to participate in community activities* (C_STDCOM); higher scale scores indicate higher numbers of students had such opportunities.

Figure 11.31: Confirmatory factor analysis of items measuring principals' perceptions of student opportunities to participate in community activities



Model fit indices:	Pooled sample
RMSEA	0.054
CFI	0.93
TLI	0.90

The results from a confirmatory factor analysis (Figure 11.31) showed satisfactory model fit for a one-dimensional model using all items from this question. The scale had satisfactory reliabilities (Cronbach's alpha) in almost all countries, with an average reliability of 0.74 and national coefficients ranging from 0.56 to 0.88 (Table 11.60). Selected item parameters were used to scale the items corresponding to these two scales (Table 11.61).

Table 11.60: Reliabilities for scale measuring principals' perceptions of student opportunities to participate in community activities

Country	Scale reliability (Cronbach's alpha)
	C_STDCOM
Belgium (Flemish)	0.67
Bulgaria	0.75
Chile	0.83
Chinese Taipei	0.88
Colombia	0.86
Croatia	0.80
Denmark	0.67
Dominican Republic	0.83
Estonia	0.69
Finland	0.65
Hong Kong SAR	0.77
Italy	0.74
Korea, Republic of	0.86
Latvia	0.67
Lithuania	0.74
Malta	0.67
Mexico	0.76
Netherlands	0.63
North-Rhine Westphalia (Germany)*	0.70
Norway	0.56
Peru	0.78
Russian Federation	0.79
Slovenia	0.71
Sweden	0.70
ICCS 2016 average	0.74
Minimum value	0.56
Maximum value	0.88

Note:
*Benchmarking participant.

Table 11.61: Item parameters for scale measuring principals' perceptions of student opportunities to participate in community activities

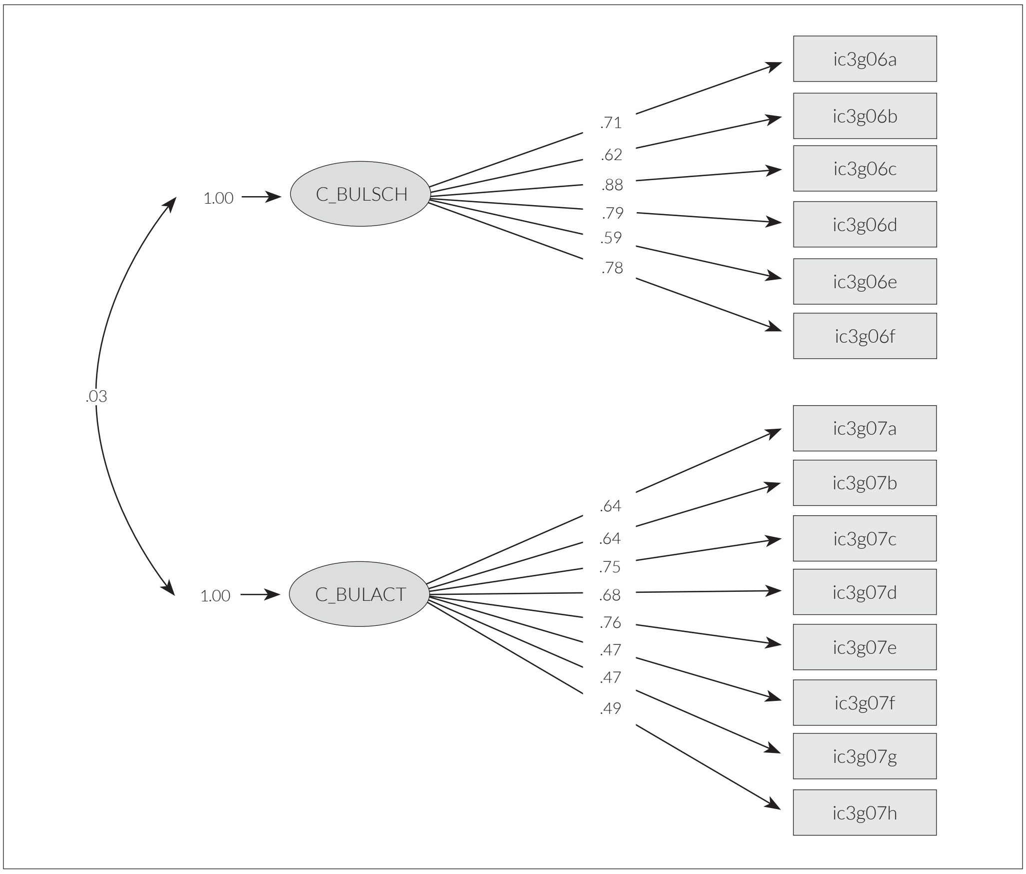
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
C_STDCOM	*During the current school year, how many <target grade> students in this school have had the opportunity to take part in any of these activities?*				
IC3G04A	Activities related to environmental sustainability (e.g. <energy and water saving, recycling>)	–0.23	–0.71	0.10	0.61
IC3G04B	Human rights projects	0.35	–0.65	0.13	0.52
IC3G04C	Activities for underprivileged people or groups	0.32	–0.75	0.23	0.52
IC3G04D	Cultural activities (e.g. theatre, music)	–0.85	–0.82	0.19	0.63
IC3G04E	Multicultural and intercultural activities within the <local community> (e.g. <promotion and celebration of cultural diversity, food street market>)	0.22	–0.38	0.04	0.34
IC3G04F	Campaigns to raise people's awareness, such as <campaigns to raise people's awareness about social issues, campaigns to raise people's awareness of environmental issues>	–0.03	–0.42	–0.02	0.44
IC3G04G	Activities aimed at protecting the cultural heritage within the <local community>	0.41	–0.43	–0.11	0.54
IC3G04H	Visits to political institutions (e.g. <Parliament house, Prime Minister's/President's official residence>)	1.04	–0.33	0.20	0.13
IC3G04I	Sports events	–1.23	–0.59	–0.02	0.60

Principals' perceptions of bullying and reports on activities against bullying at school

The school questionnaire contained a new question about the frequency of bullying in the school (Question 6). Principals were asked to indicate either "Never," "Less than once a month," "1 to 5 times a month" and "More than five times a month" for six different situations involving some form of bullying. We used all six items to derive a scale reflecting *principals' perceptions of bullying at school* (C_BULSCH), where higher values on this scale reflected more frequent occurrences of bullying.

Question 7 asked principals if a series of eight activities against bullying were undertaken at the school ("Yes" or "No"). These items enabled us to derive the scale *principals' reports on activities against bullying at school* (C_BULACT); higher values indicated more anti-bullying activities were undertaken.

Figure 11.32: Confirmatory factor analysis of items measuring principals' perceptions of bullying and reports on activities against bullying at school



Model fit indices:	Pooled sample
RMSEA	0.037
CFI	0.96
TLI	0.95

The results from confirmatory factor analyses assume a two-dimensional model for these items (Figure 11.32). The model fit was highly satisfactory and the results showed that the two latent dimensions were not correlated. The scale reliability (Cronbach's alpha) for C_BULSCH was 0.75 on average with coefficients ranging from 0.54 to 0.84 (see Table 11.62). For C_BULACT the average reliability was only marginally satisfactory, being 0.65 on average with national coefficients ranging from 0.45 to 0.87. Selected item parameters were used to scale the items corresponding to these two scales (Table 11.63).

Table 11.62: Reliabilities for scales measuring principals' perceptions of bullying and reports on activities against bullying at school

Country	Scale reliability (Cronbach's alpha)	
	C_BULSCH	C_BULACT
Belgium (Flemish)	0.69	0.54
Bulgaria	0.74	0.62
Chile	0.81	0.72
Chinese Taipei	0.81	0.72
Colombia	0.82	0.78
Croatia	0.73	0.75
Denmark	0.77	0.50
Dominican Republic	0.80	0.87
Estonia	0.65	0.63
Finland	0.72	0.50
Hong Kong SAR	0.77	0.70
Italy	0.54	0.67
Korea, Republic of	0.75	0.62
Latvia	0.82	0.45
Lithuania	0.81	0.65
Malta	0.74	0.68
Mexico	0.83	0.79
Netherlands	0.75	0.56
North-Rhine Westphalia (Germany)*	0.74	0.65
Norway	0.71	0.52
Peru	0.84	0.72
Russian Federation	0.80	0.64
Slovenia	0.59	0.62
Sweden	0.74	0.65
ICCS 2016 average	0.75	0.65
Minimum value	0.54	0.45
Maximum value	0.84	0.87

Note:
*Benchmarking participant.

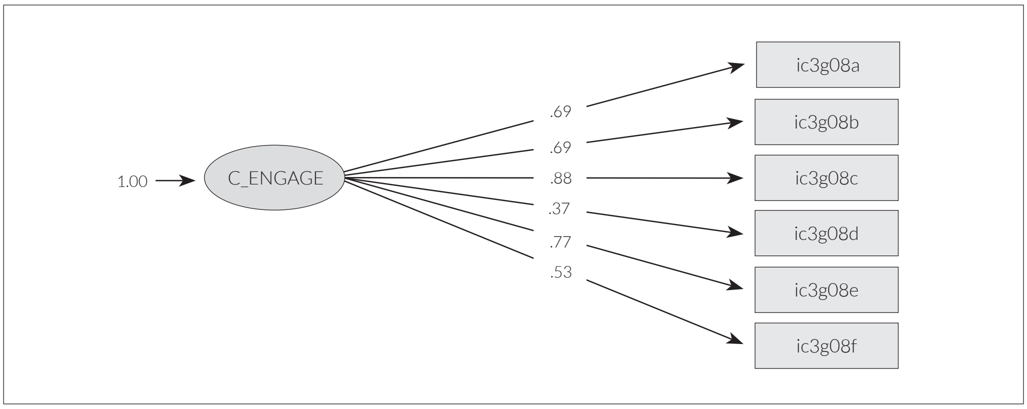
Table 11.63: Item parameters for scales measuring principals' perceptions of bullying and report on activities against bullying at school

Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
C_BULSCH	*During the current school year, how often did any of the following situations happen at this school?*			
IC3G06A	A student reported to <the principal, the headteacher, the school head> aggressive or destructive behaviours by other students	–1.34	–1.93	1.93
IC3G06B	A student reported to <the principal, the headteacher, the school head> that s/he was <bullied> by a teacher	1.33	–1.57	1.57
IC3G06C	A teacher reported to <the principal, the headteacher, the school head> that a student was <bullied> by other students	–0.99	–1.91	1.91
IC3G06D	A teacher reported to <the principal, the headteacher, the school head> that a student helped another student who was being <bullied>	0.02	–1.79	1.79
IC3G06E	A teacher reported to <the principal, the headteacher, the school head> that s/he was being <bullied> by students	1.57	–1.70	1.70
IC3G06F	A parent reported to <the principal, the headteacher, the school head> that his/her son/daughter was <bullied> by other students	–0.59	–2.37	2.37
C_BULACT	*During the current school year, are any of the following activities against <bullying> (including <cyber-bullying>) being undertaken at this school?*			
IC3G07A	Meetings aiming at informing parents about <bullying> at school	–0.53	0.00	
IC3G07B	Specific training to provide teachers with knowledge, skills and confidence to make students aware of <bullying>	0.04	0.00	
IC3G07C	Teacher training sessions on safe and responsible internet use to avoid <cyber-bullying>	0.54	0.00	
IC3G07D	Student training sessions for responsible internet use to avoid <cyber-bullying>	–0.80	0.00	
IC3G07E	Meetings aiming at raising parents' awareness on <cyber-bullying>	0.08	0.00	
IC3G07F	Development of a system to report anonymously incidents of <cyber-bullying> among students	2.22	0.00	
IC3G07G	Classroom activities aiming at raising students' awareness on <bullying>	–2.80	0.00	
IC3G07H	<Anti-bullying> conferences held by experts and/or by local authorities on <bullying> at school	1.25	0.00	

Principals' perceptions of engagement of the school community

Question 8 was newly developed for ICCS 2016, and asked principals to indicate the extent to which students, teachers and parents had the opportunity to engage with the school community (response options were "To a large extent," "To a moderate extent," "To a small extent," and "Not at all"). The six items were all used in the scale *principals' perceptions of engagement of the school community* (C_ENGAGE). Higher scores on this scale corresponded to a greater perception of engagement of the school community. A confirmatory factory analysis supported a one-dimensional model for the scale (Figure 11.33) for the pooled ICCS 2016 dataset with equally weighted countries.

Figure 11.33: Confirmatory factor analysis of items measuring principals' perceptions of engagement of the school community



Model fit indices:	Pooled sample
RMSEA	0.058
CFI	0.98
TLI	0.96

The average scale reliability (Cronbach's alpha) of C_ENGAGE was generally only marginally satisfactory, being 0.67 with coefficients ranging from 0.56 to 0.79 across countries (see Table 11.64). The parameters used for scaling the items are presented in Table 11.65.

Table 11.64: Reliabilities for scale measuring principals' perceptions of engagement of the school community

Country	Scale reliability (Cronbach's alpha)
	C_ENGAGE
Belgium (Flemish)	0.67
Bulgaria	0.61
Chile	0.76
Chinese Taipei	0.76
Colombia	0.74
Croatia	0.71
Denmark	0.65
Dominican Republic	0.68
Estonia	0.65
Finland	0.62
Hong Kong SAR	0.72
Italy	0.60
Korea, Republic of	0.66
Latvia	0.59
Lithuania	0.72
Malta	0.70
Mexico	0.65
Netherlands	0.69
North-Rhine Westphalia (Germany)*	0.56
Norway	0.73
Peru	0.79
Russian Federation	0.68
Slovenia	0.60
Sweden	0.58
ICCS 2016 average	0.67
Minimum value	0.56
Maximum value	0.79

Note:
*Benchmarking participant.

Table 11.65: Item parameters for scale measuring principals' perceptions of engagement of the school community

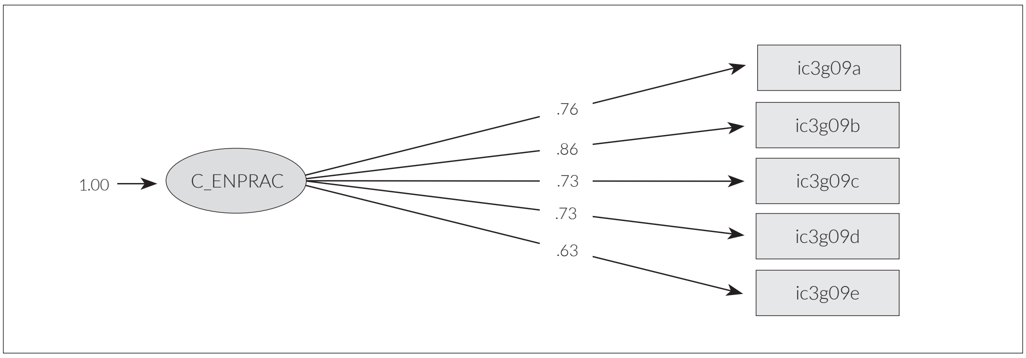
Scale or item	Question/item wording	Item parameters		
		Delta	Tau(1)	Tau(2)
C_ENGAGE	*To what extent do the following statements apply to the current situation at this school?*			
IC3G08A	Teachers are involved in decision-making processes	−1.02	−1.67	1.67
IC3G08B	Parents are involved in decision-making processes	1.84	−1.48	1.48
IC3G08C	Students' opinions are taken into account in decision-making processes	0.98	−1.70	1.70
IC3G08D	Rules and regulations are followed by teaching and non-teaching staff, students, and parents	−1.13	−1.85	1.85
IC3G08E	Students are given the opportunity to actively participate in school decisions	1.11	−1.37	1.37
IC3G08F	Parents are provided with information on the school and student performance	−1.78	−1.09	1.09

Principals' reports on environment-friendly practices at school

Question 9 of the school questionnaire asked principals to report on the extent to which five environmentally friendly practices were implemented in their schools. Respondents could respond "To a large extent," "To a moderate extent," "To a small extent," and "Not at all". All five items were used to derive the scale *principals' reports on environment-friendly practices at school* (C_ENPRAC). Higher scale scores indicate practices were more well established. Despite a marginally poor model fit, a confirmatory factor analysis did support a one-factor solution for the scale (see Figure 11.34).

The scale reliability (Cronbach's alpha) for C_ENPRAC was good, being 0.77 on average with coefficients ranging from 0.60 to 0.86 across participating countries (see Table 11.66). Selected item parameters were used to scale the items corresponding to this new scale for ICCS 2016 (Table 11.67).

Figure 11.34: Confirmatory factor analysis of items measuring principals' reports on environment-friendly practices at school



Model fit indices:	Pooled sample
RMSEA	0.120
CFI	0.96
TLI	0.92

Table 11.66: Reliabilities for scale measuring principals' reports on environment-friendly practices at school

Country	Scale reliability (Cronbach's alpha)
	C_ENPRAC
Belgium (Flemish)	0.67
Bulgaria	0.79
Chile	0.86
Chinese Taipei	0.60
Colombia	0.84
Croatia	0.80
Denmark	0.74
Dominican Republic	0.80
Estonia	0.73
Finland	0.77
Hong Kong SAR	0.83
Italy	0.80
Korea, Republic of	0.71
Latvia	0.75
Lithuania	0.75
Malta	0.86
Mexico	0.76
Netherlands	0.78
North-Rhine Westphalia (Germany)*	0.81
Norway	0.75
Peru	0.85
Russian Federation	0.73
Slovenia	0.75
Sweden	0.84
ICCS 2016 average	0.77
Minimum value	0.60
Maximum value	0.86

Note:
*Benchmarking participant.

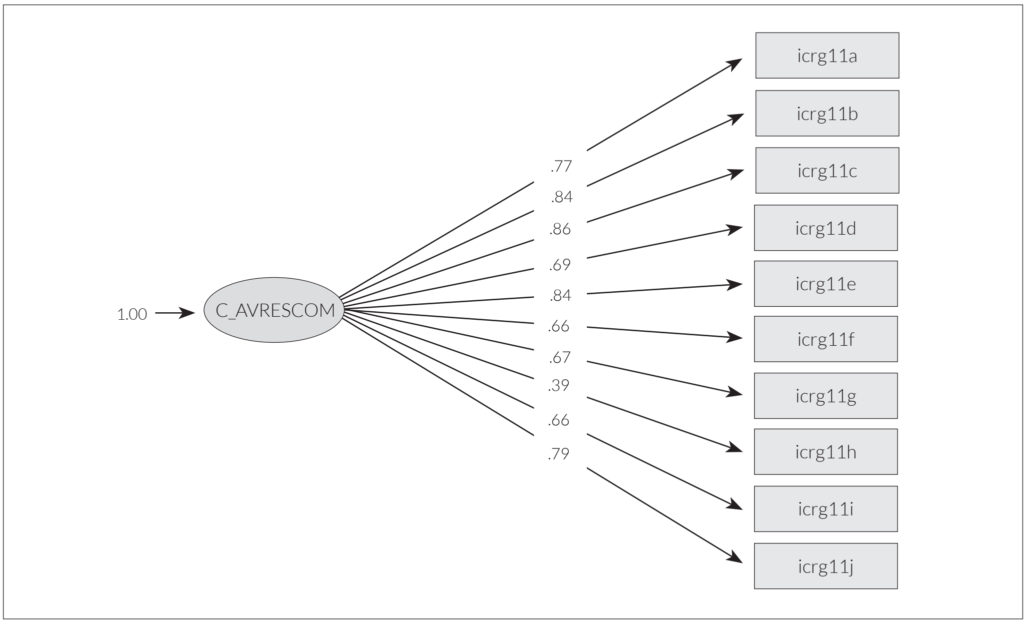
Table 11.67: Item parameters for scale measuring principals' reports on environment-friendly practices at school

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
C_ENPRAC	*To what extent are the following practices implemented at this school?*				
IC3G09A	Differential waste collection	−0.06	−0.97	−0.26	1.23
IC3G09B	Waste reduction (e.g. <encouraging waste-free lunches, limiting the use of plastic disposable products>)	0.09	−1.51	−0.03	1.54
IC3G09C	Purchasing of environmentally friendly items (e.g. <recycled paper for printing, biodegradable cutlery and dishes>)	0.57	−1.70	−0.05	1.75
IC3G09D	Energy-saving practices	−0.49	−1.66	−0.25	1.90
IC3G09E	Posters to encourage students' environmental-friendly behaviors	−0.11	−1.71	−0.10	1.81

Principals' perceptions of the availability of resources in local community

Question 11 asked principals to confirm whether there were certain resources available in the area where the school was located ("Yes" or "No"). We used all ten items to derive the scale *principals' reports on the availability of resources in local community* (C_AVRESCOM). Higher values on this scale reflected greater availability of resources in the school's local community.

Figure 11.35: Confirmatory factor analysis of items measuring principals' perceptions of the availability of resources in local community



Model fit indices:	Pooled sample
RMSEA	0.067
CFI	0.97
TLI	0.96

The results from a confirmatory factor analysis (Figure 11.35) showed satisfactory model fit for a one-dimensional model. The scale reliabilities (Cronbach's alpha) for C_AVRESCOM were generally high, being 0.75 on average with coefficients ranging from 0.52 to 0.83 (Table 11.69). Selected item parameters were used to scale the items corresponding to this scale.

Table 11.68: Reliabilities for scale measuring principals' perceptions of the availability of resources in local community

Country	Scale reliability (Cronbach's alpha)
	C_AVRESCOM
Belgium (Flemish)	0.72
Bulgaria	0.77
Chile	0.77
Chinese Taipei	0.82
Colombia	0.82
Croatia	0.83
Denmark	0.82
Dominican Republic	0.82
Estonia	0.75
Finland	0.77
Hong Kong SAR	0.63
Italy	0.70
Korea, Republic of	0.61
Latvia	0.70
Lithuania	0.63
Malta	0.75
Mexico	0.83
Netherlands	0.78
North-Rhine Westphalia (Germany)*	0.52
Norway	0.73
Peru	0.83
Russian Federation	0.81
Slovenia	0.77
Sweden	0.73
ICCS 2016 average	0.75
Minimum value	0.52
Maximum value	0.83

Note:
*Benchmarking participant.

Table 11.69: Item parameters for scale measuring principals' perceptions of the availability of resources in local community

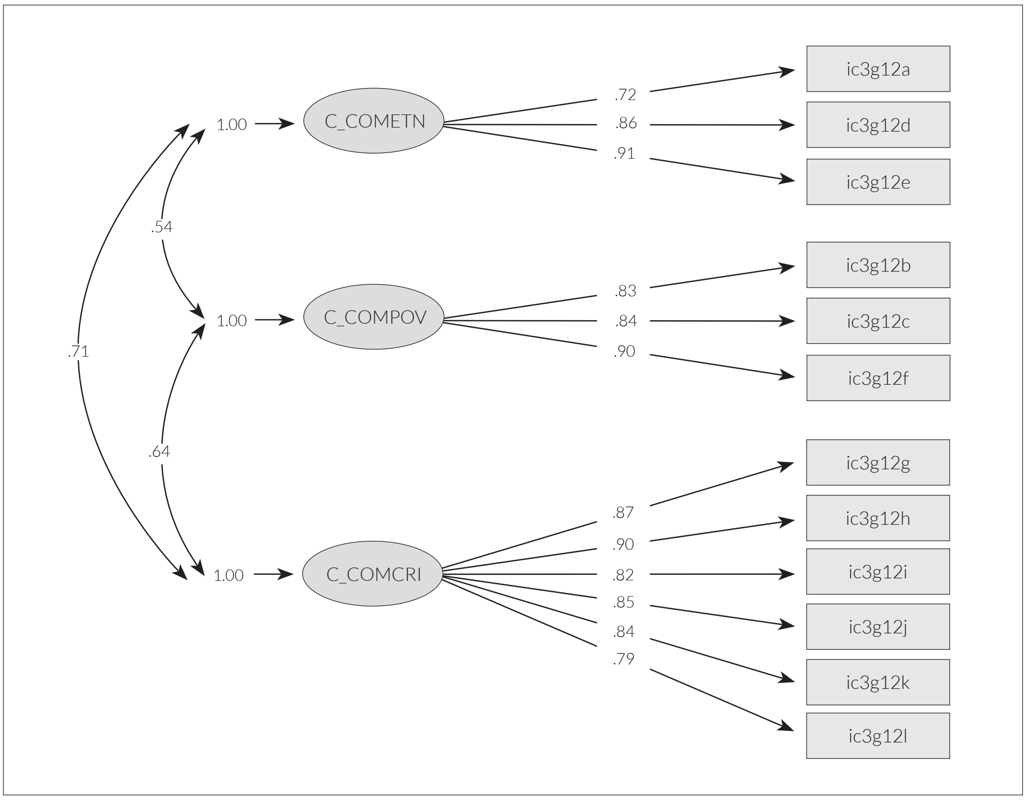
Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
C_AVRESCOM	*Are the following resources available in the immediate area where the school is located?*				
IC3G11A	Public library	–0.85	0.00	n/a	n/a
IC3G11B	Cinema	1.96	0.00	n/a	n/a
IC3G11C	Theatre or Concert Hall	1.51	0.00	n/a	n/a
IC3G11D	Language school	2.34	0.00	n/a	n/a
IC3G11E	Museum or Art Gallery	1.26	0.00	n/a	n/a
IC3G11F	Playground	–1.33	0.00	n/a	n/a
IC3G11G	Public garden or Park	–1.24	0.00	n/a	n/a
IC3G11H	Religious centre (e.g. church, mosque, synagogue)	–2.15	0.00	n/a	n/a
IC3G11I	Sports facilities (e.g. swimming pool, tennis courts, basketball court, <football> field)	–2.17	0.00	n/a	n/a
IC3G11J	Music schools	0.68	0.00	n/a	n/a

Note:
n/a = not applicable

Principals' perceptions of social tension, poverty and crime in the community

In Question 12, principals were asked to indicate the extent to which a series of 12 issues were a source of social tension ("To a large extent," "To a moderate extent," "To a small extent," and "Not at all"). The items were allocated to three scales: three were used to derive the scale *principals' perceptions of social tension in the community* (C_COMETN), three were used to form the scale *principals' perceptions of poverty in the community* (C_COMPOV), and the remaining six items were used to derive the scale *principals' perceptions of crime in the community* (C_COMCRI). Higher scores on each of these three scales corresponded to higher levels of social tension.

Figure 11.36: Confirmatory factor analysis of items measuring principals' perceptions of social tension, poverty and crime in the community



Model fit indices:	Pooled sample
RMSEA	0.061
CFI	0.98
TLI	0.97

The three scales were substantiated using a confirmatory factory analysis that revealed satisfactory fit for a three-factor model (see Figure 11.36). Each of the scales had generally satisfactory reliabilities (Cronbach's alpha) (see Table 11.70) and the average reliabilities were 0.70 for C_COMETN (with national coefficients ranging from 0.45 to 0.87), 0.82 for C_COMPOV (with national coefficients ranging between 0.61 and 0.90) and 0.86 for C_COMCRI (with national coefficients ranging between 0.71 and 0.92). Table 11.71 shows the item parameters that we used for the scaling of this item set.

Table 11.70: Reliabilities for scales measuring principals' perceptions of social tension, poverty and crime in the community

Country	Scale reliability (Cronbach's alpha)		
	C_COMETN	C_COMPOV	C_COMCRI
Belgium (Flemish)	0.87	0.90	0.90
Bulgaria	0.51	0.82	0.86
Chile	0.83	0.86	0.92
Chinese Taipei	0.74	0.74	0.90
Colombia	0.71	0.82	0.91
Croatia	0.59	0.61	0.78
Denmark	0.79	0.84	0.89
Dominican Republic	0.62	0.86	0.91
Estonia	0.53	0.71	0.71
Finland	0.80	0.81	0.88
Hong Kong SAR	0.77	0.88	0.91
Italy	0.84	0.84	0.86
Korea, Republic of	0.51	0.76	0.87
Latvia	0.45	0.86	0.77
Lithuania	0.52	0.85	0.78
Malta	0.68	0.82	0.90
Mexico	0.64	0.86	0.88
Netherlands	0.81	0.79	0.89
North-Rhine Westphalia (Germany)*	0.85	0.89	0.87
Norway	0.80	0.88	0.91
Peru	0.60	0.85	0.91
Russian Federation	0.56	0.82	0.76
Slovenia	0.81	0.75	0.78
Sweden	0.84	0.85	0.92
ICCS 2016 average	0.70	0.82	0.86
Minimum value	0.45	0.61	0.71
Maximum value	0.87	0.90	0.92

Note:
*Benchmarking participant.

Table 11.71: Item parameters for scales measuring principals' perceptions of social tension, poverty and crime in the community

Scale or item	Question/item wording	Item parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
C_COMETN	*To what extent are any of the following issues a source of social tension in the immediate area where the school is located?*				
IC3G12A	Presence of immigrants	–0.83	–1.89	0.30	1.59
IC3G12D	Religious intolerance	0.22	–2.60	0.55	2.05
IC3G12E	Ethnic conflicts	0.60	–2.35	0.64	1.71
C_COMPOV	*To what extent are any of the following issues a source of social tension in the immediate area where the school is located?*				
IC3G12B	Poor quality of housing	0.73	–2.87	0.27	2.60
IC3G12C	Unemployment	–1.04	–2.95	0.33	2.62
IC3G12F	Extensive poverty	0.31	–2.51	0.24	2.27
C_COMCRI	*To what extent are any of the following issues a source of social tension in the immediate area where the school is located?*				
IC3G12G	Organized crime	0.67	–2.20	0.72	1.48
IC3G12H	Youth gangs	0.42	–2.31	0.54	1.77
IC3G12I	Petty crime	–0.51	–3.32	0.73	2.59
IC3G12J	Sexual harassment	1.22	–2.85	0.56	2.29
IC3G12K	Drug abuse	–0.47	–2.81	0.53	2.28
IC3G12L	Alcohol abuse	–1.33	–3.06	0.40	2.66

References

Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). *ACER ConQuest: Generalised item response modelling software* [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.

Bentler, P.M., & Bonnet, D.C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606.

Bollen, K. A., & Long, S. J., Eds. (1993). *Testing structural equation models*. Newbury Park, CA: Sage Publications.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*(4), 872–882.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Fabrigar, L. R., Wegener, D.T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.

Flora, D. B., & Curren, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data, *Psychological methods*, *9*(4), 466-91

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.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement equivalence in aging research. *Experimental Aging Research*, *18*, 117–144.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

International Labour Organization. (2007). *International Standard Classification of Occupations: ISCO-2008*. Geneva, Switzerland: International Labour Office.

Kaplan, D. (2009). *Structural equation modeling, foundations and extensions*. Los Angeles, London, New Delhi & Singapore: SAGE.

Kolen, M.J., & Brennan, R.L. (2004). *Test equating, scaling, and linking* (2nd edition). New York, NY: Springer Science and Business Media, Inc.

Little, T. D. (1997). Mean and covariances structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioural Research*, *32*(1), 53–76.

Losito, B., Agrusti, G., Damiani, V., & Schulz, W. (2018). *Young people's perceptions of Europe in a time of change: IEA International Civic and Citizenship Education Study 2016 European Report*. Cham, Switzerland: Springer. Retrieved from https://www.springer.com/gp/book/9783319739595.

MacCallum, R.C., Browne, M.W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-49.

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.

Monseur, C., & Berezner, A. (2007). The computation of equating errors in international surveys in education. *Journal of Applied Measurement*, *8*(3), 323–335.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical outcomes*. Unpublished manuscript available as Mplus webnote. Retrieved from http://www.statmodel.com/bmuthen/articles/Article_075.pdf.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables. User's guide*. (Version 7). Los Angeles, CA: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edition). New York, NY: McGraw-Hill.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.

Schulz, W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study. In: *IERI Monograph Series Volume 2, Issues and Methodologies in Large-scale Assessments* (pp. 113-135). Amsterdam: IEA-ETS Research Institute.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., & Friedman, T. (2018). *Becoming citizens in a changing world. The International Civic and Citizenship Education Study 2016 International Report*. Cham, Switzerland: Springer.

Schulz, W., & Friedman, T. (2011). Scaling procedures for ICCS questionnaire items. In W. Schulz, J. Ainley, & J. Fraillon (Eds.). *ICCS 2009 Technical Report* (pp. 157–259). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., & Friedman, T. (2015). Scaling procedures for ICILS questionnaire items. In J. Fraillon, W. Schulz, T. Friedman, J. Ainley, & E. Gebhardt (Eds.) *International Computer and Literacy Information Study 2013 Technical Report* (pp. 177–220). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Tucker, L., & MacCallum, R. (1997). *Exploratory factor analysis. Unpublished manuscrip*t. Retrieved from: http://www.unc.edu/~rcm/book/factornew.htm.

UNESCO. (2011). *International Standard Classification of Education: ISCED 2011*. Retrieved from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450.

CHAPTER 12:

The reporting of ICCS 2016 results

Wolfram Schulz

Overview

In this chapter, we describe the procedures used for reporting results in the ICCS 2016 publications. We first describe the replication methodology used to estimate sampling variance and illustrate how we computed the imputation variance of the civic knowledge scores. We then describe how we conducted significance tests for differences between country and subsample means or percentages, as well as between results from ICCS 2016 and 2009 for countries that participated in both surveys.

We also include descriptions of how we estimated multilevel (hierarchical) models explaining variation in students' civic knowledge, and how we conducted our multiple regression analyses to explain variation in questionnaire scale scores reflecting students' expected electoral and active political participation.

Estimation of sampling variance

ICCS employed two-stage cluster sampling procedures to obtain the student and teacher samples. During the first stage, schools were sampled from a sampling frame with a probability proportional to their size (see Chapter 5 for further details). 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, it is not appropriate to apply the usual formulae for obtaining standard errors reflecting sampling error for population estimates.

Replication techniques provide tools with which to estimate the correct sampling variance on population estimates (Gonzalez & Foy, 2000; Wolter, 1985). For ICCS 2016, as in the previous cycle (see Schulz, 2011), we 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, we constructed sampling zones within explicit strata. When faced with occurrences of an odd number of schools within an explicit stratum or the sampling frame, we randomly divided the remaining school into two halves, thereby forming a sampling zone of two "quasi-schools."

Each of the countries participating in ICCS 2016 had up to 75 sampling zones. In countries where we had larger numbers of schools, we combined some schools into bigger "pseudo-schools" in order to keep the total number to 75. In the Russian Federation, we applied a three-stage sample design; the first stage consisted of a sample of regions. If a selected region was large enough to be selected with certainty, we paired schools; if not, we paired regions in the sampling zones. In countries, where we surveyed all schools, classrooms and students, we defined classrooms as PSUs. This method was applied to Malta and one explicit stratum in Croatia. For the teacher survey, respondents from the same school were randomly assigned to different sampling zones in these two cases.

Within each of the sampling zones, we randomly assigned one school a value of 2 and the other school a value of 0. For each of the sampling zones, we computed replicate weights. 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 multiplying student or teacher weights with the jackknife indicators once only for each sampling zone.

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. This procedure can be illustrated by a simple example featuring 24 students from six different schools (A–F) paired into three sampling zones (Table 12.1).

Table 12.1: Example of the 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

For each country sample, we computed 75 replicate weights 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, μ, 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_μ estimate is computed using the formula:

$$SV_\mu = \sum_{i=1}^{75} [\mu_i - \mu_S]^2$$

where μ_s is the statistic μ estimated for the population through use of the original sampling weights and μ_i is the same statistic estimated by using the weights for the i^{th} of 75 jackknife replicates. The standard error SE_μ for statistic μ, which reflects the uncertainty of the estimate due to sampling, 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 does not always include procedures for replication techniques.

For the jackknife replication of ICCS 2016 data, we used tailored SPSS software macros. These results can be replicated by using the IEA IDB Analyzer, which is generally recommended as a tool for analyzing IEA data.[17] 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).[18]

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 10 for a description of the scaling methodology for ICCS 2016 test items). Therefore, population statistics for ICCS 2016 civic knowledge scores should always be estimated using all five plausible values.

If θ is the international civic knowledge and μ_θ^P is the statistic of interest computed based on each of the P plausible values, then the statistic μ_θ based on all plausible values can be computed as follows:

$$\mu_\theta = \frac{1}{P} \sum_{p=1}^{P} \mu_\theta^p$$

The sampling variance SV_μ is calculated as the average of the sampling variance for each plausible value SV_μ^p:

$$SV_\mu = \frac{1}{P} \sum_{p=1}^{P} SV_\mu^p$$

Use of the P plausible values for data analysis also allows the amount of error associated with the measurement of civic knowledge to be estimated. The measurement variance or imputation variance IV_p is computed as:

$$IV_p = \frac{1}{P-1} \sum_{p=1}^{P} (\mu_\theta^p - \mu_\theta)^2$$

17 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.

18 The module is an add-in component running under SPSS and offers a number of features for applying different replication methods when estimating sampling and imputation variance. The application can be downloaded from https://iccs.acer.org/ICCS2016reports.

Here, μ_θ^p is the statistic of interest computed on each plausible value p and μ_θ is the mean statistic based on all P plausible values.

The estimate of the total variance TV_μ, 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_μ is equal to:

$$SE_\mu = \sqrt{V_\mu}$$

We calculated the average scale scores and their sampling and overall standard errors for each country (Table 12.2). The comparison between sampling and combined standard error shows that most of the error was due to sampling and that only a relatively small proportion could be attributed to measurement error.

Table 12.2: 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
Belgium (Flemish)	537	4.01	4.08	2931
Bulgaria	485	5.31	5.34	2966
Chile	482	3.06	3.11	5081
Chinese Taipei	581	2.98	3.03	3953
Colombia	482	3.35	3.39	5609
Croatia	531	2.42	2.48	3896
Denmark	586	2.95	2.98	6254
Dominican Republic	381	2.94	3.04	3937
Estonia	546	2.98	3.05	2857
Finland	577	2.20	2.26	3173
Italy	524	2.41	2.42	3450
Latvia	492	2.98	3.13	3224
Lithuania	518	2.89	3.04	3631
Malta	491	2.71	2.73	3764
Mexico	467	2.54	2.54	5526
Netherlands	523	4.47	4.54	2812
Norway	564	2.14	2.19	6271
Peru	438	3.52	3.54	5166
Russian Federation	545	4.17	4.28	7289
Slovenia	532	2.41	2.49	2844
Sweden	579	2.70	2.83	3264
Countries not meeting sampling requirements				
Hong Kong SAR	515	6.60	6.64	2653
Korea, Republic of	551	3.45	3.57	2601
Benchmarking participant not meeting sampling requirements				
North Rhine-Westphalia (Germany)	519	2.66	2.73	1451

Reporting of differences

We conducted significance tests for:

- Differences in population estimates between countries (pair-wise comparisons);
- Differences between a country and the international average;
- Differences in population estimates between subgroups within countries; and
- Differences between population estimates in ICCS 2016 and 2009 (estimation of changes over time) for countries that participated in both surveys.

We considered differences between two score averages (or percentages) *a* and *b* significant ($p < 0.05$) when the test statistic *t* was greater than the critical value, 1.96. We calculated *t* by dividing the difference by its standard error, SE_{dif_ab}:

$$t = \frac{(a\text{-}b)}{SE_{dif\_ab}}$$

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 $_b$ are the standard errors of the means from the two independent samples *a* and *b*.

When comparing subsamples within countries that are not independent samples (e.g., gender groups), we derived the difference between statistics for subgroups of interest and the standard error of the difference by using jackknife replication that involved the following formula:

$$SE_{dif\_ab} = \sqrt{\left[\sum_{i=1}^{75}((a^i - b^i) - (a - b))^2\right]}$$

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, we calculated the standard error of the differences with ($P = 5$) plausible values by using this formula:

$$SE_{dif\_ab} = \sqrt{\left[\sum_{p=1}^{P}\left(\sum_{i=1}^{75}((a_p^i - b_p^i) - (a_p - b_p))^2\right)/P\right] + \left[(1+\frac{1}{P})\frac{\sum_{p=1}^{P}((a_p - b_p) - (\bar{a}_p - \bar{b}_p))^2}{P-1}\right]}$$

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.

When comparing the country means c with the overall ICCS 2016 average *i*, we had to account for the fact that the country being considered had contributed to the international standard error. We did this by calculating the standard error SE_{dif_ic} of the difference between the overall ICCS 2016 average and an individual country average as:

$$SE_{dif\_ic} = \frac{\sqrt{((N-1)^2 - 1)SE_c^2 + \sum_{k=1}^{N} SE_k^2}}{N}$$

Here, SE_c is the sampling standard error for country c and SE_k is the sampling error for k^{th} of N participating countries. We used this formula to determine the statistical significance of differences due to sampling error between countries and the ICCS 2016 averages of the questionnaire scales throughout the ICCS 2016 reports.

While the above formula was sufficient for the questionnaire scale scores, we found it necessary to also take the imputation component of standard errors for countries into account when comparing the civic knowledge score averages of a country with the overall ICCS 2016 average. The imputation variance component of standard errors $SE^2_{i\_dif\_ic}$ was given as:

$$SE^2_{i\_dif\_ic} = \sqrt{(1+\frac{1}{P})^2 \operatorname{var}(d_1,\ldots,d_p,\ldots,d_5)}$$

where d_p is the difference between the overall ICCS 2016 and the country mean for the plausible value p.

We computed the final standard error ($SE_{a\_dif\_ic}$) of the difference between ICCS 2016 country test scores and the ICCS 2016 average as:

$$SE_{a\_dif\_ic} = \sqrt{SE^2{}_{dif\_ic} + SE^2_{i\_dif\_ic}}$$

The ICCS 2016 international report also included comparisons of results between ICCS 2009 and 2016. Because the process of equating the test or questionnaire scales across the cycles introduced some additional error into the calculation of any test statistic, we added an equating error term to the formula for the standard error of the difference between country averages.

When testing the difference of a statistic between the two assessments, we computed the standard error of the difference as follows:

$$SE(\mu_{ICCS16} - \mu_{ICCS09}) = \sqrt{SE_{(ICCS16)}{}^2 + SE_{(ICCS09)}{}^2 + EqErr^2}$$

Here, μ can be any statistic in units on the equated ICCS scale (mean, percentile, gender difference, but not percentages) and $SE_{(ICCS16)}$ and $SE_{(ICCS09)}$ 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 3.086 score points for the civic knowledge scale and ranged from 0.019 to 0.726 across the international and regional equated questionnaire scales (see Chapter 10 and 11 for the calculation of the respective equating errors). In Chapter 11, the equating errors for questionnaire scales are listed according to their respective reporting metrics (see Table 11.2).

The equating error could not be applied directly to report the significance of differences between percentages at or above Level B. Therefore, we applied the following replication method to estimate the equating error for percentages at at or above a certain level of proficiency (e.g. students with civic knowledge test scores corresponding to levels A and B).

To estimate the standard error of the percentage at or above the cut-point that defines the threshold between Levels C and B (479), within each participating country a number of n replicate cut-points were generated by adding a random error component with a mean of 0 and a standard deviation equal to the estimated equating error (3.086). Percentages of students at or above each replicate cut-point (ρ_n) were computed and an equating error for each participating country estimated as:

$$EquErr\ (\rho\_country) = \sqrt{\frac{(\rho_n - \rho_o)^2}{n}}$$

where ρ_n is the percentage of students at or above Level B (Table 12.3). We used 1000 replicate samples (n) to compute these estimates.

Table 12.3: Estimates of equating error for percentages at or above Level B

Country	Equating error estimates
Belgium (Flemish)	1.23
Bulgaria	0.83
Chile	1.23
Chinese Taipei	0.70
Colombia	1.45
Denmark	0.51
Dominican Republic	0.62
Estonia	1.08
Finland	0.61
Italy	1.12
Latvia	1.39
Lithuania	1.23
Malta	0.95
Mexico	1.36
Norway	0.61
Russian Federation	0.94
Slovenia	1.11
Sweden	0.62
ICCS 2016 average	0.98

Within each participating country, the standard errors for the differences between percentages at or above proficient standards were calculated as:

$$SE(\rho_{ICCS16} - \rho_{ICCS09}) = \sqrt{SE_{(\rho_{ICCS16})}^2 + SE_{(\rho_{ICCS09})}^2 + EqErr_{(\rho\_country)}^2}$$

Hierarchical linear modeling

To review which factors are associated with variation in civic knowledge within and across schools within participating countries, we estimated hierarchical (or multilevel) linear regression models (Raudenbush & Bryk, 2002) in which students were nested within schools. Predictor variables included variables reflecting students' personal and social background, the socioeconomic context of homes and schools, civic learning outside school, civic learning at school, and the school and community learning context.

A hierarchical regression model with *i* students nested in *j* clusters (schools) can be estimated as

$$Y_{ij} = \alpha_j + X_{ij}^n \beta_j + X_j^m \beta_j + U_{0j} + \varepsilon_{ij}$$

Here, Y_{ij} is the criterion variables, X_{ij}^n is a vector of student-level variables, with its corresponding vector of regression coefficients β_j, and X_j^m is a school-level variable with its corresponding vector of regression coefficients β_j. U_{0j} is the residual term at the level of the cluster (school), and ε_{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:

$$Y_{ij} = \alpha_j + U_{0j}^{null} + \varepsilon_{ij}^{null}$$

The residual term U_{0j}^{null} provides an estimate of the variance in Y_{ij} between j clusters, and ε_{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 our 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, we estimated the "null" model. From this model we excluded students with missing data for any of the analysis variables. We computed the explained variance at the school level EV_j as

$$EV_j = (1 - \frac{U_{0j}}{U_{0j}^{null}}) \times 100$$

and we computed the explained variance at the student level EV_{ij} as

$$EV_{ij} = (1 - \frac{\varepsilon_{ij}}{\varepsilon_{ij}^{null}}) \times 100$$

Because multilevel modeling takes the hierarchical structure of the cluster sample into account, we reported multilevel standard errors that reflect both sampling and imputation errors. Data were weighted (with normalized school-level and [within-school] student-level weights) following a recommended procedure for the analysis of IEA data with this particular sampling design (see Rutkowski, Gonzalez, Jonkers, & von Davier, 2010).

We used the software package Mplus (Version 7; see Muthén & Muthén, 2012) to estimate all hierarchical models. We excluded data from the benchmarking participant North Rhine-Westphalia (Germany) from the analyses due to their extremely low sample participation rate. Results from Hong Kong SAR and the Republic of Korea were reported separately and should be interpreted with caution.

As is customary when applying multivariate analyses, we observed increases in the proportions of missing data when including more variables in the model. For the multilevel analyses of civic knowledge, 93 percent of students on average across ICCS 2016 countries meeting sample participation requirements had valid data for all variables included in the model. In the Dominican Republic, only 81 percent of the weighted sample had valid data for inclusion in the analyses, and consequently the results for the Dominican Republic should be interpreted with due caution.

We calculated the respective unweighted and weighted percentages of students with valid data for all variables in the model (Table 12.4).

In most countries, one intact classroom per school was sampled, which made it impossible to disentangle classroom- and school-level variance. In one small country (Malta), all classrooms were sampled in each school; a few other countries had smaller numbers of schools where more than one classroom was sampled. These differences in sampling design need to be taken into account when interpreting the results of the multilevel analyses of ICCS 2016 data.

Table 12.4: ICCS 2016 students included in multilevel analyses of variation in civic knowledge

Country	Total number of assessed students	Total number of students in analysis	Unweighted percentage of students in analysis	Weighted percentage of students in analysis
Belgium (Flemish)	2931	2740	93.5	92.4
Bulgaria	2966	2751	92.8	91.6
Chile	5081	4823	94.9	94.7
Chinese Taipei	3953	3816	96.5	96.4
Colombia	5609	5294	94.4	93.7
Croatia	3896	3782	97.1	97.4
Denmark	6254	5573	89.1	89.2
Dominican Republic	3937	3230	82.0	81.0
Estonia	2857	2796	97.9	98.1
Finland	3173	3054	96.2	96.3
Italy	3450	3294	95.5	95.4
Latvia	3224	3004	93.2	93.8
Lithuania	3631	3422	94.2	95.0
Malta	3764	3389	90.0	89.8
Mexico	5526	5077	91.9	92.1
Netherlands	2812	2725	96.9	96.6
Norway	6271	5766	91.9	91.9
Peru	5166	4770	92.3	91.7
Russian Federation	7289	7016	96.3	95.1
Slovenia	2844	2728	95.9	96.0
Sweden	3264	2933	89.9	90.3
Countries not meeting sampling requirements				
Hong Kong SAR	2653	2382	89.8	89.2
Korea, Republic of	2601	2522	97.0	97.0
Minimum			82.0	81.0
Maximum			97.9	98.1

Note:
Data from the benchmarking participant North Rhine-Westphalia (Germany) are excluded from the analyses due to their extremely low sample participation rate.

Multiple regression modeling

When reporting ICCS 2016 data, we also applied single-level multiple regression models to explain variation in the questionnaire scale scores reflecting students' expected electoral and active political participation, respectively. For these we used the number of students in each country, students included in each of the multiple regression analyses, as well as the respective unweighted and weighted percentages of students with valid data for all variables in the model (Table 12.5). Predictor variables were student background variables, experience with civic participation, dispositions for civic engagement and beliefs.

When conducting multiple regression models (see, for example, Pedhazur, 1997), analysts regress a criterion variable Y_i on a set of k predictors $X_{1i} \dots X_{ki}$, with α being the intercept, ε_i the unexplained part of the model (residual), and k regression coefficients β:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots\dots + \beta_k X_{ki} + \varepsilon_i$$

We reported the unstandardized regression coefficients and the variance explained by the model to show the effects for each predictor and the overall explanatory power of the model. We employed jackknife replication using tailored SPSS macros to estimate the standard errors for the multiple regression model parameters (unstandardized coefficients and estimates of explained variance).

Across the participating countries, on average 92 percent of students in the sample had valid data for all variables. National average percentages of students with valid data for all variables ranged from 68 percent in the Dominican Republic to 98 percent in Chinese Taipei.

Mindful of these missing values, we compared our results with those from models that used an alternative approach to the treatment of missing values, where students with missing values on variables received mean scores and median values, and missing indicator variables were added for each variable (Cohen & Cohen, 1975). As we obtained almost identical results when applying these two approaches, we selected the simpler approach of "list-wise" exclusion of missing values, where only students with valid data for all variables in the models were included in the multiple regression analyses.

In order to estimate the unique contribution of each set of predictors to the variance explanation of the model, and the proportion of variance explained by more than one set of predictors, we computed different linear regression models. For each set of j with m predictor variables, we left one predictor variable 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. We computed the variance uniquely explained for predictor variable set j ($r^2_{u\_j}$) as:

$$r^2_{u\_j} = r^2_n - r^2_{n\text{-}m}$$

Here, r^2_n is the R square for the model and $r^2_{n\text{-}m}$ is the R square for the regression model without the m variables in predictor block j.

We then expressed the unique contribution of predictor set j to the explained variance in the predictor variable Y_i in percentages:

$$UVC_j = r^2_{u\_j} \times 100$$

The joint explained variance contribution reflects the proportion of variance explained by more than one of k sets of predictors. We computed the proportion of variance explained by more than one set of predictors JVC_j as

$$JVC_j = (r^2_n \times 100) - \sum_{k=1}^{k} UVC_j$$

Table 12.5: ICCS 2016 students included in multiple regression analyses of expected electoral and active political participation

Country	Total number of assessed students	Analysis of expected electoral participation:			Analysis of expected active political participation:		
		Total number of students in analysis	Unweighted percentage of students in analysis	Weighted percentage of students in analysis	Total number of students in students in analysis	Unweighted percentage of students in analysis	Weighted percentage of students in analysis
Belgium (Flemish)	2931	2803	95.6	95.7	2803	95.6	95.6
Bulgaria	2966	2669	90.0	88.8	2668	90.0	88.8
Chile	5081	4758	93.6	93.1	4760	93.7	93.2
Chinese Taipei	3953	3874	98.0	97.9	3875	98.0	98.0
Colombia	5609	4969	88.6	88.4	4957	88.4	88.1
Croatia	3896	3735	95.9	96.1	3729	95.7	96.1
Denmark	6254	5591	89.4	89.3	5589	89.4	89.3
Dominican Republic	3937	2685	68.2	67.6	2672	67.9	67.3
Estonia	2857	2769	96.9	97.1	2771	97.0	97.2
Finland	3173	3000	94.5	94.6	3001	94.6	94.6
Italy	3450	3341	96.8	96.7	3340	96.8	96.7
Latvia	3224	3012	93.4	94.0	3015	93.5	94.0
Lithuania	3631	3446	94.9	95.4	3446	94.9	95.4
Malta	3764	3331	88.5	88.1	3329	88.4	88.0
Mexico	5526	5022	90.9	90.9	5025	90.9	91.0
Netherlands	2812	2673	95.1	94.4	2672	95.0	94.4
Norway	6271	5595	89.2	89.3	5590	89.1	89.1
Peru	5166	4561	88.3	87.1	4553	88.1	87.0
Russian Federation	7289	7001	96.0	95.1	6999	96.0	95.0
Slovenia	2844	2764	97.2	97.3	2765	97.2	97.3
Sweden	3264	2971	91.0	90.7	2970	91.0	90.7
Countries not meeting sampling requirements							
Hong Kong SAR	2653	2453	92.5	91.9	2449	92.3	91.8
Korea, Republic of	2601	2504	96.3	96.3	2504	96.3	96.3

Note:
Data from the benchmarking participant North Rhine-Westphalia (Germany) are excluded from the analyses due to their extremely low sample participation rate.

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.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables. User's guide*. (Version 7). Los Angeles, CA: Muthén & Muthén.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (Third edition). Ford Worth, TX: Harcourt Brace College Publishers

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151.

Schulz, W. (2011). The reporting of ICCS results. In W. Schulz, J. Ainley, J., & J. Fraillon (Eds.). *ICCS 2009 technical report* (pp. 261–280). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.

Westat (2007). *WesVar® 4.3: User's guide* [computer software]. Rockville, MD: Author.

Wolter, K. M. (1985). *Introduction to variance estimation*. New York, NY: Springer.

Appendices

APPENDIX A: ORGANIZATIONS AND INDIVIDUALS INVOLVED IN ICCS 2016

International Study Center

The international Study Center is located at the Australian Council for Educational Research (ACER). ACER is responsible for designing and implementing the study in close cooperation with LPS (Laboratorio di Pedagogia Sperimentale at the Roma Tre University, Rome, Italy) on behalf of the IEA.

Staff at ACER

Wolfram Schulz, *research director*
Julian Fraillon, *coordinator of test development*
John Ainley, *project researcher*
Tim Friedman, *project researcher*
Nora Kovarcikova, *project researcher*
Naoko Tabata, *project researcher*
Judy Nixon, *test development*
Trisha Reimers, *test development*
Eveline Gebhardt, *coordinator of data analysis*
Louise Ockwell, *data analyst*
Jorge Fallas, *data analyst*
Leigh Patterson, *data analyst*
Dulce Lay, *data analyst*
Renee Kwong, *data analyst*

Staff at LPS

Bruno Losito, *associate research director*
Gabriella Agrusti, *project researcher*
Elisa Caponera, *project researcher*
Paola Mirti, *project researcher*
Valeria Damiani, *project researcher*
Francesco Agrusti, *project researcher*
Alessandro Sanzo, *project researcher*

International Association for the Evaluation of Educational Achievement (IEA)

IEA provides overall support and supervision for ICCS. The IEA Hamburg, Germany, as the international coordinating center for ICCS, is responsible for overall coordination of all activities, relations with participating countries, and sampling and data-processing. The IEA Amsterdam, the Netherlands, is responsible for translation verification and quality monitoring of the data collection.

Staff at the IEA Hamburg

Ralph Carstens, *project director*
Marta Kostek, *project coordinator*
Juliane Kobelt, *project coordinator*
Falk Brese, *international data manager*
Hannah Köhler, *international data manager*
Christine Busch, *deputy international data manager*
Sabine Weber, *research analyst (sampling)*
Sabine Tieck, *research analyst (sampling)*
Diego Cortes, *research analyst (sampling)*
Olaf Zuehlke, *research analyst (sampling)*
Duygu Savaşcı, *research analyst (sampling)*
Dirk Oehler, *research analyst*

Tim Daniel, *research analyst*
Michael Jung, *research analyst*
Alena Becker, *research analyst*
Parisa Aghakasiri, *research analyst*
Kamil Kowolik, *research analyst*
Svetoslav Velkov, *research analyst*
Ekaterina Mikheeva, *research analyst*
Clara Beyer, *research analyst*
Oriana Mora, *research analyst*
Maike Junod, *programmer*
Limiao Duan, *programmer*
Deepti Kalamadi, *programmer*
Bettina Wietzorek, *meeting and seminar coordinator*
Heiko Sibberns, *director of IEA Hamburg*

Staff at the IEA Amsterdam

Dirk Hastedt, *executive director*
Paulína Koršňáková, *director of IEA Amsterdam*
Andrea Netten, *director of IEA Amsterdam*
Gabriela Noveanu, *senior researcher*
Gillian Wilson, *senior publications officer*
Roel Burgers, *financial manager*
Isabelle Gemin, *financial officer*
Sive Finlay, *communications officer*
Mirjam Govaerts, *public outreach officer*

Project advisory committee (PAC)

The ICCS 2016 PAC has, from the beginning of the project, advised the international study center and its partner institutions during regular meetings

PAC members

Erik Amnå, *Örebro University, Sweden*
Cristián Cox, *Diego Portales University, Chile*
Barbara Malak-Minkiewicz, *The Netherlands*
Judith Torney-Purta, *University of Maryland, United States*
Wiel Veugelers, *University of Humanistic Studies Utrecht, The Netherlands*

Other project advisors

ICCS sampling referee

Marc Joncas from Statistics Canada in Ottawa was the sampling referee for the study. He provided invaluable advice on all sampling-related aspects of it.

Expert

Christian Monseur (Université de Liège) conducted a review of test and questionnaire scaling methodology. In addition, the international study center invited him to review the content of the international report.

ICCS 2016 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 the implementation of ICCS in participating countries.

Belgium (Flemish)
Ellen Claes
University of Leuven, Centre for Citizenship and Democracy

Bulgaria
Svetla Petrova
Center for Control and Assessment of the Quality in School Education

Chile
Elisa Salinas Valdivieso
Education Quality Assurance Agency

Chinese Taipei
Meihui Liu
National Taiwan Normal University

Colombia
Andrés Gutiérrez
Ximena Dueñas Herrera
Colombian Institute for the Assessment of Education (ICFES)

Croatia
Ines Elezović
Department for Quality Assurance in Education, National Centre for External Evaluation of Education

Denmark
Jens Bruun
Danish School of Education, Aarhus University

Dominican Republic
Massiel Cohen
Ancell Scheker
Ministry of Education

Estonia
Anu Toots
Tallin University

Finland
Jouko Mehtäläinen
Finnish Institute for Educational Research, University of Jyväskylä

Germany (North Rhine-Westphalia)
Hermann Josef Abs
University of Duisburg-Essen

Hong Kong SAR
Wai Man Lam
Centre for Civil Society and Governance, the University of Hong Kong

Italy
Laura Palmerio
INVALSI

Latvia
Ireta Chekse
University of Latvia

Lithuania
Mindaugas Stundža
Justė Grebliūnienė
National Examination Center

Malta
Karen Grixti
Directorate for Research and Policy Development

Mexico
María Antonieta-Díaz Gutiérrez
National Institute for the Evaluation of the Education (INEE, México)

The Netherlands
Anke Munniksma
University of Amsterdam

Norway
Lihong Huang
NOVA—Norwegian Social Research, Oslo and Akershus University College of Applied Sciences

Peru
Humberto Perez Leon
Office of Educational Quality Measurement

Republic of Korea
Geun Young Chang
Tae-Jun Kim
National Youth Policy Institute (NYPI)

Russian Federation
Petr Polozhevets
Publishing House "Teachers Weekly"

Slovenia
Eva Klemencic
Educational Research Institute

Sweden
Ellen Almgren
Swedish National Agency for Education

APPENDIX B: CHARACTERISTICS OF NATIONAL SAMPLES

This appendix describes, for each education system participating in ICCS 2016, 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 surveys could differ. It was particularly common for a 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 2016 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 Belgium (Flemish)

- School-level exclusions consisted of schools for children with special education needs.
- Explicit stratification was performed by school type (regular schools, eligible schools for students with special education needs [SEN], for a total of two explicit strata.
- Implicit stratification was applied by organizational type (public, private, none) and SES percentile (three levels), for a total of seven implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.1.1: 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	
Regular schools	163	0	130	22	8	3
Eligible SEN schools	2	0	2	0	0	0
Total	**165**	**0**	**132**	**22**	**8**	**3**

Table B.1.2: 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	
Regular schools	163	0	126	22	7	8
Eligible SEN schools	2	0	2	0	0	0
Total	**165**	**0**	**128**	**22**	**7**	**8**

Note:
Five schools were regarded as non-participating because the within-school participation rate was below 50%.

B.2 Bulgaria

- School-level exclusions consisted of schools for students with special education needs and of schools with less than five students in the target grade.
- Explicit stratification was performed by school type (general, profiled, vocational), for a total of three explicit strata.
- Implicit stratification was applied by region (region 01– 06) and size of settlement (small town, medium size town, large town, none), for a total of 23 implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.2.1: Allocation of student sample in Bulgaria

School participation sttatus—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
General	87	2	85	0	0	0
Profiled	32	0	32	0	0	0
Vocational	31	1	30	0	0	0
Total	**150**	**3**	**147**	**0**	**0**	**0**

Table B.2.2: Allocation of teacher sample in Bulgaria

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
General	87	2	83	0	0	2
Profiled	32	0	27	0	0	5
Vocational	31	1	30	0	0	0
Total	**150**	**3**	**140**	**0**	**0**	**7**

Note:
Seven schools were regarded as non-participating because the within-school participation rate was below 50%.

B.3 Chile

- School-level exclusions consisted of schools for students with special education needs, geographically inaccessible schools, and schools with less than eight students in the target grade.
- Explicit stratification was performed by school administration (public, private subsidized, private), for a total of three explicit strata.
- Private schools were oversampled.
- Implicit stratification was applied by national assessment performance groups (missing performance score, low performance, medium performance, high performance), for a total of four implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.3.1: Allocation of student sample in Chile

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Public	71	1	68	2	0	0
Private subsidized	79	1	74	2	2	0
Private	30	0	22	6	2	0
Total	**180**	**2**	**164**	**10**	**4**	**0**

Table B.3.2: Allocation of teacher sample in Chile

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Public	71	1	65	2	0	3
Private subsidized	79	1	70	2	2	4
Private	30	0	20	6	2	2
Total	**180**	**2**	**155**	**10**	**4**	**9**

Note:
Eight schools were regarded as non-participating because the within-school participation rate was below 50%.

B.4 Chinese Taipei

- School-level exclusions consisted of schools for students with special education needs and schools with less than three classes in the target grade.
- Explicit stratification was performed by region (North, Middle, South, East) and school type (private, public), for a total of eight explicit strata.
- Implicit stratification was applied by urbanization (rural, urban), for a total of two implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.4.1: 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	7	0	7	0	0	0
North – Public	56	0	52	0	0	4
Middle – Private	6	0	4	0	0	2
Middle – Public	40	0	38	1	0	1
South – Private	3	0	2	0	0	1
South – Public	30	0	29	0	0	1
East – Private	2	0	2	0	0	0
East – Public	6	0	6	0	0	0
Total	**150**	**0**	**140**	**1**	**0**	**9**

Table B.4.2: 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	7	0	7	0	0	0
North – Public	56	0	54	0	0	2
Middle – Private	6	0	6	0	0	0
Middle – Public	40	0	37	1	0	2
South – Private	3	0	3	0	0	0
South – Public	30	0	28	0	0	2
East – Private	2	0	2	0	0	0
East – Public	6	0	6	0	0	0
Total	**150**	**0**	**143**	**1**	**0**	**6**

B.5 Colombia

- School-level exclusions consisted of schools with less than five students in the target grade.
- Explicit stratification was performed by school type (public, private), hemisphere (calendar A, calendar B), and municipality type (certified, not certified), for a total of four explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.5.1: 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	
Private – Calendar A	25	0	23	2	0	0
Private – Calendar B	3	0	3	0	0	0
Public – Certified municipality	59	0	58	1	0	0
Public – Not certified municipality	63	0	61	2	0	0
Total	**150**	**0**	**145**	**5**	**0**	**0**

Table B.5.2: 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	
Private – Calendar A	25	0	22	2	0	1
Private – Calendar B	3	0	3	0	0	0
Public – Certified municipality	59	0	53	0	0	6
Public – Not certified municipality	63	0	55	1	0	7
Total	**150**	**0**	**133**	**3**	**0**	**14**

Note:
Ten schools were regarded as non-participating because the within-school participation rate was below 50%.

B.6 Croatia

- School-level exclusions consisted of schools for students with special education needs and schools with less than five students in the target grade.
- Explicit stratification was performed by region and schools with special civic education (CE) program, for a total of seven explicit strata.
- Schools with CE program were oversampled.
- No implicit stratification was applied.
- One classroom per school was sampled in the explicit strata 1 to 6, while in stratum 7, all classes in all schools with special CE curriculum were sampled (census).
- Small schools were sampled with equal probabilities.

Table B.6.1: Allocation of student sample in Croatia

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Region 1	32	0	30	2	0	0
Region 2	28	0	28	0	0	0
Region 3	16	0	16	0	0	0
Region 4	16	0	16	0	0	0
Region 5	30	0	27	1	0	2
Region 6	26	0	25	0	0	1
Former civic education schools	30	0	30	0	0	0
Total	**178**	**0**	**172**	**3**	**0**	**3**

Note:
Four schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.6.2: Allocation of teacher sample in Croatia

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Region 1	32	0	30	2	0	0
Region 2	28	0	27	0	0	1
Region 3	16	0	16	0	0	0
Region 4	16	0	16	0	0	0
Region 5	30	0	29	1	0	0
Region 6	26	0	25	0	0	1
Former civic education schools	30	0	30	0	0	0
Total	**178**	**0**	**173**	**3**	**0**	**2**

Note:
Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.7 Denmark

- School-level exclusions consisted of public and private schools for students with special education needs and private schools with less than four students in the target grade.
- No explicit stratification was performed.
- Implicit stratification was applied by region, for a total of 18 implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.7.1: Allocation of student sample in Denmark

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Denmark	240	21	114	50	20	34	1
Total	**240**	**21**	**114**	**50**	**20**	**34**	**1**

Table B.7.2: Allocation of teacher sample in Denmark

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Denmark	240	21	38	14	7	159	1
Total	**240**	**21**	**38**	**14**	**7**	**159**	**1**

Note:
Three schools were regarded as non-participating because the within-school participation rate was below 50%.

B.8 Dominican Republic

- School-level exclusions consisted of schools for students with special education needs, foreign language schools, and schools with less than seven students in the target grade.
- Explicit stratification was performed by organizational type (private, public, semi-official) and urbanization (urban, rural), for a total of six explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.8.1: Allocation of student sample in the Dominican Republic

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Private – Urban	26	0	25	1	0	0
Private – Rural	2	0	2	0	0	0
Public – Urban	88	8	77	2	1	0
Public – Rural	29	1	27	1	0	0
Semi official – Urban	3	0	3	0	0	0
Semi official – Rural	2	0	2	0	0	0
Total	**150**	**9**	**136**	**4**	**1**	**0**

Table B.8.2: Allocation of teacher sample in the Dominican Republic

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Private – Urban	26	0	21	1	0	4
Private – Rural	2	0	2	0	0	0
Public – Urban	88	8	69	2	0	9
Public – Rural	29	1	27	1	0	0
Semi official – Urban	3	0	3	0	0	0
Semi official – Rural	2	0	2	0	0	0
Total	**150**	**9**	**124**	**4**	**0**	**13**

Note:
Six schools were regarded as non-participating because the within-school participation rate was below 50%.

B.9 Estonia

- School-level exclusions consisted of schools for students with special education needs, schools with less than five students in the target grade, and international curriculum or EU schools.
- Explicit stratification was performed by language (Estonian, Russian, Russian & Estonian), for a total of three explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.9.1: 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	134	6	121	4	0	3
Russian	25	2	23	0	0	0
Russian & Estonian	16	0	16	0	0	0
Total	**175**	**8**	**160**	**4**	**0**	**3**

Table B.9.2: 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	134	6	36	2	0	90
Russian	25	2	6	0	0	17
Russian & Estonian	16	0	5	0	0	11
Total	**175**	**8**	**47**	**2**	**0**	**118**

Note:
One hundred and six schools were regarded as non-participating because the within-school participation rate was below 50%.

B.10 Finland

- School-level exclusions consisted of schools for students with special education needs, foreign language schools (instructional language not Finnish or Swedish), and schools in Åland (independent autonomous area).
- Explicit stratification was performed by region (five levels) and urbanization (urban, semi-urban, rural), for a total of 15 explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.10.1: 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	
Helsinki-Uusimaa – Urban	40	0	35	4	0	1
Helsinki-Uusimaa – Semi-urban	6	0	6	0	0	0
Helsinki-Uusimaa – Rural	2	0	2	0	0	0
Southern Finland – Urban	26	0	25	1	0	0
Southern Finland – Semi-urban	7	0	6	1	0	0
Southern Finland – Rural	5	0	4	1	0	0
Western Finland – Urban	23	0	19	3	0	1
Western Finland – Semi-urban	11	0	10	1	0	0
Western Finland – Rural	8	0	5	3	0	0
Northern and Eastern Finland – Urban	25	1	23	1	0	0
Northern and Eastern Finland – Semi-urban	9	1	6	2	0	0
Northern and Eastern Finland – Rural	12	0	9	2	0	1
Swedish speaking – Urban	6	1	5	0	0	0
Swedish speaking – Semi-urban	2	0	2	0	0	0
Swedish speaking – Rural	3	0	3	0	0	0
Total	**185**	**3**	**160**	**19**	**0**	**3**

Table B.10.2: 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	
Helsinki-Uusimaa – Urban	40	0	34	4	0	2
Helsinki-Uusimaa – Semi-urban	6	0	6	0	0	0
Helsinki-Uusimaa – Rural	2	0	2	0	0	0
Southern Finland – Urban	26	0	24	1	0	1
Southern Finland – Semi-urban	7	0	6	1	0	0
Southern Finland – Rural	5	0	4	1	0	0
Western Finland – Urban	23	0	17	3	0	3
Western Finland – Semi-urban	11	0	9	1	0	1
Western Finland – Rural	8	0	5	3	0	0
Northern and Eastern Finland – Urban	25	1	21	1	0	2
Northern and Eastern Finland – Semi-urban	9	1	6	2	0	0
Northern and Eastern Finland – Rural	12	0	7	2	0	3
Swedish speaking – Urban	6	1	5	0	0	0
Swedish speaking – Semi-urban	2	0	2	0	0	0
Swedish speaking – Rural	3	0	3	0	0	0
Total	185	3	151	19	0	12

Note:
Nine schools were regarded as non-participating because the within-school participation rate was below 50%.

B.11 Hong Kong, SAR

- School-level exclusions consisted of schools with non-local curriculum.
- Explicit stratification was performed by school type (government, aided, DSS & other private), for a total of three explicit strata.
- Implicit stratification was applied by banding (one, two, three) and region (Hong Kong Island, Kowloon, New Territories), for a total of nine implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.
- Unapproved teacher sampling procedures made it necessary to remove Hong Kong SAR from the teacher survey.

Table B.11: 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	8	0	0	2
Aided	118	0	63	6	0	49
DSS & other private	22	0	12	2	0	8
Total	**150**	**0**	**83**	**8**	**0**	**59**

B.12 Italy

- School-level exclusions consisted of schools with students taught in Slovenian or Ladin, and German schools.
- Explicit stratification was performed by geographic area (North West, North East, Center, South, South Islands), for a total of five explicit strata.
- Implicit stratification was applied by school type (public, private), for a total of two implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.12: Allocation of student and teacher samples in Italy

School participation status—Student and teacher surveys

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
1 North West	42	0	39	2	1	0
2 North East	30	0	26	4	0	0
3 Center	32	0	29	3	0	0
4 South	36	0	34	2	0	0
5 South Islands	30	0	29	1	0	0
Total	**170**	**0**	**157**	**12**	**1**	**0**

B.13 Korea, Republic of

- School-level exclusions consisted of schools for students with special education needs, non-mainstream schools, and schools with less than 20 students in the target grade.
- Explicit stratification was performed by region (five regions) and urbanization (rural, urban), for a total of nine explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.13.1: Allocation of student sample in the Republic of Korea

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Seoul – Urban	25	0	10	3	0	12
Jungbu – Rural	7	0	5	0	0	2
Jungbu – Urban	44	1	18	1	0	24
Chungcheong – Rural	4	0	3	0	0	1
Chungcheong – Urban	12	0	9	0	0	3
Youngnam – Rural	7	0	6	0	0	1
Youngnam – Urban	32	0	23	0	0	9
Honam – Rural	4	0	4	0	0	0
Honam – Urban	15	0	11	0	0	4
Total	150	1	89	4	0	56

Table B.13.2: Allocation of teacher sample in the Republic of Korea

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Seoul – Urban	25	0	13	4	0	8
Jungbu – Rural	7	0	5	0	0	2
Jungbu – Urban	44	1	26	1	0	16
Chungcheong – Rural	4	0	3	0	0	1
Chungcheong – Urban	12	0	9	0	0	3
Youngnam – Rural	7	0	7	0	0	0
Youngnam – Urban	32	0	23	0	0	9
Honam – Rural	4	0	4	0	0	0
Honam – Urban	15	0	11	0	0	4
Total	150	1	101	5	0	43

B.14 Latvia

- School-level exclusions consisted of schools for students with special education needs and schools where the language of instruction is neither Latvian nor Russian.
- Explicit stratification was performed by grade (grade 8 only, grades 4 and 8), urbanization (Riga, city, town-rural), language of instruction, and school type (gymnasium-secondary, basic-beginners), for a total of 10 explicit strata.
- No implicit stratification was applied.
- One classroom per school was sampled in schools with less than 70 students in the target grade and two classrooms were sampled in schools with 70 and more students in the target grade.
- Small schools were sampled with equal probabilities.

Table B.14.1: 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	
Gr8 only – Riga – None –None	8	0	7	0	0	1
Gr8 only – City – None –None	8	0	6	1	0	1
Gr 8 only – Town-rural – None – None	10	0	9	1	0	0
Gr4 and Gr8 – Riga – Latvian – None	18	0	16	0	0	2
Gr4 and Gr8 – Riga – Russian – None	24	0	21	1	0	2
Gr4 and Gr8 – City – Latvian – None	10	0	10	0	0	0
Gr4 and Gr8 – City – Russian – None	12	0	11	0	0	1
Gr4 and Gr8 – Town-rural – Latvian – Gymnasium-secondary	34	0	32	0	0	2
Gr4 and Gr8 – Town-rural – Latvian – Basic-beginner	24	0	24	0	0	0
Gr4 and Gr8 – Town-rural – Russian – None	8	0	8	0	0	0
Total	156	0	144	3	0	9

Note:
One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.14.2: 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	
Gr8 only – Riga – None – None	8	0	5	0	0	3
Gr8 only – City – None – None	8	0	6	1	0	1
Gr 8 only – Town-rural – None – None	10	0	9	1	0	0
Gr4 and Gr8 – Riga – Latvian – None	18	0	16	0	0	2
Gr4 and Gr8 – Riga – Russian – None	24	0	21	1	0	2
Gr4 and Gr8 – City – Latvian – None	10	0	10	0	0	0
Gr4 and Gr8 – City – Russian – None	12	0	11	0	0	1
Gr4 and Gr8– Town-rural – Latvian – Gymnasium-secondary	34	0	32	0	0	2
Gr4 and Gr8 – Town-rural – Latvian – Basic-beginner	24	0	23	0	0	1
Gr4 and Gr8 – Town-rural – Russian – None	8	0	8	0	0	0
Total	156	0	141	3	0	12

Note:
Three schools were regarded as non-participating because the within-school participation rate was below 50%.

B.15 Lithuania

- School-level exclusions consisted of schools for students with special education needs, schools with students not taught in Lithuanian, Polish or Russian, and schools with less than six students in the target grade.
- Explicit stratification was performed by language of instruction (Lithuanian, Russian, Polish, two languages) and urbanization (capital, other major cities, medium-size cities, small cities and villages), for a total of seven explicit strata.
- Russian and Polish schools were oversampled.
- Implicit stratification was applied by school type (basic, secondary, progymnasium, gymnasium, none), for a total of five implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.15.1: Allocation of student sample in Lithuania

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Lithuanian – Capital	16	1	15	0	0	0
Lithuanian – Other major cities	29	0	29	0	0	0
Lithuanian – Medium-size cities	52	1	51	0	0	0
Lithuanian – Small cities and villages	30	0	29	0	0	1
Russian – None	26	2	24	0	0	0
Polish – None	26	0	26	0	0	0
Two languages – None	8	0	8	0	0	0
Total	**187**	**4**	**182**	**0**	**0**	**1**

Table B.15.2: Allocation of teacher sample in Lithuania

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Lithuanian – Capital	16	1	15	0	0	0
Lithuanian – Other major cities	29	0	29	0	0	0
Lithuanian – Medium-size cities	52	1	51	0	0	0
Lithuanian – Small cities and villages	30	0	29	1	0	0
Russian – None	26	2	24	0	0	0
Polish – None	26	0	26	0	0	0
Two languages – None	8	0	8	0	0	0
Total	**187**	**4**	**182**	**1**	**0**	**0**

B.16 Malta

- School-level exclusions consisted of schools with less than twelve students in the target grade, schools for students with special education needs, and language schools.
- All eligible schools were sampled for ICCS 2016.
- No explicit stratification was performed.
- Implicit stratification was applied by organizational type (state, church, independent) and gender (males, females, co-education) for a total of seven implicit strata.
- All classrooms were sampled per school.

Table B.16: Allocation of student and teacher samples in Malta

School participation status—Student and teacher surveys

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Malta	47	0	47	0	0	0
Total	**47**	**0**	**47**	**0**	**0**	**0**

B.17 Mexico

- School-level exclusions consisted of workers secondary schools, communitarian secondary schools, lower secondary schools for the deaf, and migrant lower secondary schools.
- Explicit stratification was performed by school size (very small, moderate small, large), funding (public, private), and school program (general lower secondary, technical lower secondary, TV lower secondary), for a total of nine explicit strata.
- Very small schools were oversampled.
- Implicit stratification was applied by urbanization (urban, rural), for a total of two implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.17.1: Allocation of student sample in Mexico

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Very small	35	0	31	1	0	2	1
Moderate small – Public – General lower secondary	8	0	7	0	0	1	0
Moderate small – Public – Technical lower secondary	8	0	6	1	0	1	0
Moderate small – Public – TV lower secondary	16	0	13	0	0	3	0
Moderate small – Private	10	0	10	0	0	0	0
Large – Public – General lower secondary	71	0	67	4	0	0	0
Large – Public – Technical lower secondary	45	0	43	2	0	0	0
Large – Public – TV lower secondary	18	0	15	1	0	2	0
Large – Private	12	0	12	0	0	0	0
Total	**223**	**0**	**204**	**9**	**0**	**9**	**1**

Table B.17.2: Allocation of teacher sample in Mexico

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Very small	35	0	30	0	0	4	1
Moderate small – Public – General lower secondary	8	0	7	0	0	1	0
Moderate small – Public – Technical lower secondary	8	0	6	1	0	1	0
Moderate small – Public – TV lower secondary	16	0	12	0	0	4	0
Moderate small – Private	10	0	10	0	0	0	0
Large – Public – General lower secondary	71	0	67	4	0	0	0
Large – Public – Technical lower secondary	45	0	43	2	0	0	0
Large – Public – TV lower secondary	18	0	15	1	0	2	0
Large – Private	12	0	12	0	0	0	0
Total	**223**	**0**	**202**	**8**	**0**	**12**	**1**

B.18 The Netherlands

- School-level exclusions consisted of schools for students with special education needs.
- Explicit stratification was performed by school type (general, vocational, mixed), for a total of three explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.18.1: 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	
General	67	0	34	15	5	13
Vocational	77	0	42	18	3	14
Mixed	6	0	5	1	0	0
Total	150	0	81	34	8	27

Note:
One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.18.2: 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	
General	67	0	31	14	5	17
Vocational	77	0	39	17	2	19
Mixed	6	0	4	0	0	2
Total	150	0	74	31	7	38

Note:
Ten schools were regarded as non-participating because the within-school participation rate was below 50%.

B.19 North Rhine-Westphalia (Germany)

- School-level exclusions consisted of schools for students with special education needs and Waldorf schools.
- Explicit stratification was performed by school type (Hauptschule, Realschule, Gesamtschule, Gymnasium, Förderschule), and proportion of immigrants (two levels , high and low), for a total of nine explicit strata.
- Gymnasia with a high proportion of immigrants were oversampled.
- During the main survey, it was decided to exclude all schools for students with special education needs (Förderschulen).
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.19.1: Allocation of student sample in North Rhine-Westphalia (Germany)

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Hauptschule – High prop. imm.	16	4	0	1	1	10	0
Hauptschule – Low prop. imm.	14	4	1	2	2	5	0
Realschule – High prop. imm.	18	2	4	0	2	10	0
Realschule – Low prop. imm.	20	4	1	3	1	11	0
Gesamtschule – High prop. imm.	15	1	2	0	1	11	0
Gesamtschule – Low prop. imm.	15	0	3	1	2	9	0
Gymnasium – High prop. imm.	30	0	5	4	2	19	0
Gymnasium – Low prop. imm.	40	0	9	6	6	19	0
Förderschule	6	0	0	0	0	0	6
Total	**174**	**15**	**25**	**17**	**17**	**94**	**6**

Notes:
Two schools were regarded as non-participating because the within-school participation rate was below 50%.
High prop. imm. = high proportion of immigrants; low prop. imm. = low proportion of immigrants.

Table B.19.2: Allocation of teacher sample in North Rhine-Westphalia (Germany)

School participation status—Teacher survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Hauptschule – High prop. imm.	16	4	0	0	0	12	0
Hauptschule – Low prop. imm.	14	4	1	2	1	6	0
Realschule – High prop. imm.	18	2	2	0	1	13	0
Realschule – Low prop. imm.	20	4	0	0	0	16	0
Gesamtschule – High prop. imm.	15	1	1	0	0	13	0
Gesamtschule – Low prop. imm.	15	0	1	0	0	14	0
Gymnasium – High prop. imm.	30	0	2	0	1	27	0
Gymnasium – Low prop. imm.	40	0	1	1	2	36	0
Förderschule	6	0	0	0	0	0	6
Total	**174**	**15**	**8**	**3**	**5**	**137**	**6**

Notes:
Thirty-five schools were regarded as non-participating because the within-school participation rate was below 50%.
High prop. imm. = high proportion of immigrants; low prop. imm. = low proportion of immigrants.

B.20 Norway

- School-level exclusions consisted of Sami schools, schools for students with special education needs, foreign language schools, and schools with less than five students in the target grade.
- Explicit stratification was performed by language (Bokmål, Nynorsk), for a total of two explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.20.1: Allocation of student sample in Norway

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Bokmål	129	1	122	3	2	0	1
Nynorsk	21	0	21	0	0	0	0
Total	**150**	**1**	**143**	**3**	**2**	**0**	**1**

Table B.20.2: Allocation of teacher sample in Norway

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools	Excluded schools
			Sampled	First replacement	Second replacement		
Bokmål	129	1	117	3	2	5	1
Nynorsk	21	0	21	0	0	0	0
Total	**150**	**1**	**138**	**3**	**2**	**5**	**1**

Note:
Five schools were regarded as non-participating because the within-school participation rate was below 50%.

B.21 Peru

- School-level exclusions consisted of CEBA-schools, geographically inaccessible schools (Datem del Marañon), and schools with less than five students in the target grade.
- Explicit stratification was performed by school type (public, private) and urbanization (urban, rural), for a total of four explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.21: Allocation of student and teacher samples in Peru

School participation status—Student and teacher surveys

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Public – Urban	129	0	129	0	0	0
Public – Rural	29	1	28	0	0	0
Private – Urban	50	2	48	0	0	0
Private – Rural	2	1	1	0	0	0
Total	210	4	206	0	0	0

B.22 Russian Federation

- School-level exclusions consisted of schools for students with special education needs and schools with less than four students in target grade.
- Explicit stratification was performed by region, for a total of 42 explicit strata.
- No implicit stratification was applied.
- An extra sampling stage (regions) was required prior to sampling schools. 28 of 69 regions were selected with probability proportional to the region size and 14 bigger regions were selected with certainty. While each certainty region itself is an explicit stratum, the other sampled regions make on explicit stratum. In the large explicit stratum, a sample of schools was selected within each region.
- Within regions, schools were selected with probability proportional to (school) size systematic sampling. Schools were sorted (serpentine) by location (up to seven levels) before being sorted by school size.
- Within each region, a minimum of four schools was selected.
- Schools in the Republic of Tatarstan were oversampled.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.
- Unapproved teacher sampling procedures made it necessary to remove the Russian Federation from the teacher survey.

Table B.22: Allocation of student sample in the Russian Federation

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Republic of Tatarstan	6	0	6	0	0	0
Republic of Tatarstan (oversample)	150	0	150	0	0	0
The City of Sankt-Petersburg	6	0	6	0	0	0
The City of Moscow	12	0	12	0	0	0
Moscow region	10	0	10	0	0	0
Perm territory	4	0	4	0	0	0
Samara region	4	0	4	0	0	0
Nizhni Novgorod region	4	0	4	0	0	0
Republic of Bashkortostan	8	0	8	0	0	0
Krasnodar territory	8	0	8	0	0	0
Rostov region	6	0	6	0	0	0
Chelyabinsk region	6	0	6	0	0	0
Sverdlovsk region	6	0	6	0	0	0
Krasnoyarsk territory	4	0	4	0	0	0
Republic of Dagestan	6	0	6	0	0	0
Novgorod region	4	0	4	0	0	0
Kaliningrad region	4	0	4	0	0	0
Arkhangelsk region	4	0	4	0	0	0
Voronezh region	4	0	4	0	0	0
Belgorod region	4	0	4	0	0	0
Vladimir region	4	0	4	0	0	0
Lipetzk region	4	0	4	0	0	0
Yaroslavl region	4	0	4	0	0	0
Kaluga region	4	0	4	0	0	0
Kostroma region	4	0	4	0	0	0
Ulyanovsk region	4	0	4	0	0	0
Chuvashi Republic	4	0	4	0	0	0
Orenburg region	4	0	4	0	0	0
Saratov region	4	0	4	0	0	0
Volgograd region	4	0	4	0	0	0
Yamalo-Nenets autonomous district	4	0	4	0	0	0
Tyumen region	4	0	4	0	0	0
Irkutsk region	4	0	4	0	0	0
Kemerovo region	4	0	4	0	0	0

Table B.22: Allocation of student sample in the Russian Federation (contd.)

School participation status—Student survey

Explicit strata	Total sampled schools	Ineligible schools	Participating schools			Non-participating schools
			Sampled	First replacement	Second replacement	
Novosibirsk region	4	0	4	0	0	0)
Altai territory	4	0	4	0	0	0
Omsk region	4	0	4	0	0	0
Tomsk region	4	0	4	0	0	0
Kamchatka territory	4	0	4	0	0	0
Khabarovsk territory	4	0	4	0	0	0
Primorsky territory	4	0	4	0	0	0
Stavropol territory	4	0	4	0	0	0
Kabardino-Balkarian Republic	4	0	4	0	0	0
Total	**352**	**0**	**352**	**0**	**0**	**0**

B.23 Slovenia

- School-level exclusions consisted of schools for students with special education needs, schools with multi-grade classrooms, and schools with different curriculum (Waldorf and Montessori schools).
- No explicit stratification was performed.
- Implicit stratification was applied by NUTS 3 regions, for a total of twelve implicit strata.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.23.1: 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	
Slovenia	150	0	135	10	0	5
Total	**150**	**0**	**135**	**10**	**0**	**5**

Table B.23.2: 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	
Slovenia	150	0	133	10	0	7
Total	**150**	**0**	**133**	**10**	**0**	**7**

Note:
Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.24 Sweden

- School-level exclusions consisted of schools for students with special education needs, international schools, and schools with less than five students in the target grade.
- Explicit stratification was performed by grade average (missing, low, low–medium, medium, medium–high, high, very high), for a total of seven explicit strata.
- No implicit stratification was applied.
- One classroom was sampled per school.
- Small schools were sampled with equal probabilities.

Table B.24.1: 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	
Missing	16	1	15	0	0	0
Low	23	0	21	0	0	2
Low–medium	27	0	26	1	0	0
Medium	30	0	30	0	0	0
Medium–high	23	0	23	0	0	0
High	19	0	19	0	0	0
Very high	20	0	20	0	0	0
Total	**158**	**1**	**154**	**1**	**0**	**2**

Table B.24.2: 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	
Missing	16	1	14	0	0	1
Low	23	0	19	0	0	4
Low–medium	27	0	23	1	0	3
Medium	30	0	25	0	0	5
Medium–high	23	0	21	0	0	2
High	19	0	14	0	0	5
Very high	20	0	18	0	0	2
Total	**158**	**1**	**134**	**1**	**0**	**22**

Note:
Seventeen schools were regarded as non-participating because the within-school participation rate was below 50%.

APPENDIX C: DESCRIPTIONS OF CIVIC KNOWLEDGE TEST ITEMS AND ALLOCATIONS TO PROFICIENCY LEVELS

Table C1: Descriptions of civic knowledge test items and allocations to proficiency levels

Item	Item Label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge scale difficulty	Proficiency level
CI2ASM1	Interview help farmers get more money	2009	MC	3	2	495	Level B
CI2ASM2	Farmers seeking	2009	MC	2	2	441	Level C
CI2BCM1	Best reason elect leader	2009	MC	3	1	524	Level B
CI2BPM1	Important find out policies before vote	2009	MC	3	2	508	Level B
CI2BPM2	Use only friends information	2009	MC	3	2	481	Level B
CI2CCM1	Affection by damage to the environment	2009	MC	4	2	471	Level C
CI2CCM2	What advisor wants people to understand	2009	MC	4	2	534	Level B
CI2CEM1	Censorship government responsibility	2009	MC	2	2	430	Level C
CI2CEM2	Best reason for opposing censorship	2009	MC	2	2	556	Level B
CI2CNM1	Best reason live next door	2009	MC	2	2	489	Level B
CI2CNM2	Need to understand in order to accept	2009	MC	4	2	382	Level D
CI2DLM1	Why laws about information of donations	2009	MC	2	2	605	Level A
CI2ECM1	Most likely reason to ask for advice	2009	MC	3	2	430	Level C
CI2ECM2	Risk by asking for advice	2009	MC	3	2	501	Level B
CI2ETM2	Argument against protecting business	2009	MC	2	2	530	Level B
CI2FDM1	Adult citizens expected to decide	2009	MC	1	1	485	Level B
CI2FSM1	Most likely behavior against law	2009	MC	2	2	491	Level B
CI2GFM1	Reason best supports opinion playground	2009	MC	1	2	450	Level C
CI2GLM1	Factory benefit local people	2009	MC	1	2	461	Level C
CI2GLM2	Factory harm local people	2009	MC	1	2	494	Level B
CI2HRM1	Entitled to protection of human rights	2009	MC	2	1	464	Level C
CI2JOM1	Journalists research report news freely	2009	MC	2	2	443	Level C
CI2ORM1	Not teach official religion	2009	MC	2	2	495	Level B
CI2PCM1	Best reason for choosing card b	2009	MC	3	2	535	Level B
CI2PCM2	Availability of card on the internet	2009	MC	3	2	412	Level C
CI2PGM1	Way pressure groups contribute democracy	2009	MC	1	2	497	Level B
CI2PGM2	Advantage of independent pressure groups	2009	MC	1	2	591	Level A

Table C1: Descriptions of civic knowledge test items and allocations to proficiency levels (contd.)

Item	Item Label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge scale difficulty	Proficiency level
CI2PJM1	Best reason people in jail cannot vote	2009	MC	2	2	542	Level B
CI2PJM2	Best reason people in jail allowed vote	2009	MC	1	2	407	Level C
CI2PRM1	Best reason against violent protest	2009	MC	3	2	576	Level A
CI2RCM1	Need to introduce new emblems	2009	MC	2	2	500	Level B
CI2REM2	Help for refugees	2009	MC	2	1	476	Level C
CI2REM3	Responsibility of refugees	2009	MC	2	1	430	Level C
CI2SCM1	Reason best support district decision	2009	MC	2	2	526	Level B
CI2SCM2	How become better players because rule	2009	MC	2	2	518	Level B
CI2VOM1	Best reason for voluntary voting	2009	MC	1	2	463	Level C
CI2VOM2	Best support responsibility to vote	2009	MC	1	2	544	Level B
CI2VOM3	Reason for using secret ballots	2009	MC	1	2	370	Level D
CI303M1	Traffic laws	2016	MC	2	1	306	Below Level D
CI307M1	Constitution recognition	2016	MC	2	1	540	Level B
CI308M1	Equal opportunity employment	2016	MC	2	2	436	Level C
CI312M1	Democratic government action	2016	MC	1	1	488	Level B
CI314M1	Law into own hands	2016	MC	2	1	419	Level C
CI3CAM1	Art as political expression	2016	MC	2	2	567	Level A
CI3CPO1	Town meeting	2016	CR	3	2	516	Level B
CI3CPO2	Survey people	2016	CR	3	2	571	Level A
CI3CRM1	Art and culture	2016	MC	4	2	419	Level C
CI3CRM2	Corporate citizenship	2016	MC	1	2	564	Level A
CI3CSM1	School in community	2016	MC	1	2	459	Level C
CI3DBM1	Medical care as right	2016	MC	4	2	370	Level D
CI3DBM2	NGO independence	2016	MC	1	2	485	Level B
CI3DDM1	Debate and discuss	2016	MC	3	2	512	Level B
CI3EPM1	Informed voters	2016	MC	1	2	433	Level C
CI3GMM1	Government and military	2016	MC	1	1	488	Level B
CI3GTM1	Government taxes	2016	MC	1	2	415	Level C
CI3ICM1	International charity	2016	MC	1	2	456	Level C
CI3IEM1	International environmental agreement	2016	MC	1	1	474	Level C

Table C1: Descriptions of civic knowledge test items and allocations to proficiency levels (contd.)

Item	Item Label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge scale difficulty	Proficiency level
CI3IVM1	Informed voting	2016	MC	3	2	443	Level C
CI3LPM1	Online advocacy	2016	MC	1	1	401	Level C
CI3LSM1	Legal support	2016	MC	2	2	451	Level C
CI3LTM1	Regular elections	2016	MC	2	2	587	Level A
CI3MAM1	Role of military	2016	MC	1	1	374	Level D
CI3MDM1	Government controlled media	2016	MC	1	2	506	Level B
CI3MPM1	Misuse of power example	2016	MC	2	1	451	Level C
CI3NPM1	Reason for rules	2016	MC	1	1	301	Below Level D
CI3NPM2	Benefits of rules	2016	MC	1	1	354	Level D
CI3NWM1	Noisy workplace laws	2016	MC	1	2	363	Level D
CI3PAM1	Political parties	2016	MC	3	2	475	Level C
CI3PEM1	Open budget	2016	MC	1	2	482	Level B
CI3PLM1	Pet leash laws	2016	MC	1	2	321	Level D
CI3REM1	Right to education	2016	MC	2	1	284	Below Level D
CI3RRM1	Representative responsibilities	2016	MC	1	1	414	Level C
CI3SCM1	Democratic selection	2016	MC	3	1	395	Level C
CI3SDM1	Secret information	2016	MC	1	2	536	Level B
CI3SMM1	Social media	2016	MC	1	2	470	Level C
CI3SPM1	Speeding minister	2016	MC	2	1	322	Level D
CI3SWM1	Social welfare justification	2016	MC	1	2	407	Level C
CI3UHM1	Union help	2016	MC	1	1	456	Level C
CI3ULM1	Unjust law	2016	MC	2	2	526	Level B
CI3VGM1	Volunteer group	2016	MC	3	2	356	Level D

Table C1: Descriptions of civic knowledge test items and allocations to proficiency levels (contd.)

Item	Item Label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge scale difficulty	Proficiency level
Partial Credit Thresholds							
CI2BIO1_1	Understanding histories and cultures	2009	CR	2	2	614	Level A
CI2BIO1_2	Understanding histories and cultures	2009	CR	2	2	826	Level A
CI2ETO1_1	Help protect farmers business	2009	CR	3	2	593	Level A
CI2ETO1_2	Help protect farmers business	2009	CR	3	2	794	Level A
CI2WFO1_1	Reasons giving money unemployed people	2009	CR	2	2	489	Level B
CI2WFO1_2	Reasons giving money unemployed people	2009	CR	2	2	688	Level A
CI2WFO2_1	Argument in favor of requiring people	2009	CR	1	2	572	Level A
CI2WFO2_2	Argument in favor of requiring people	2009	CR	1	2	608	Level A
CI3CBO1_1	Citizenship book	2016	CR	2	2	443	Level C
CI3CBO1_2	Citizenship book	2016	CR	2	2	606	Level A
CI3MPO2_1	Misuse of power prevention	2016	CR	1	1	509	Level B
CI3MPO2_2	Misuse of power prevention	2016	CR	1	1	670	Level A
CI3PRO1_1	Consult electorate	2016	CR	3	2	466	Level C
CI3PRO1_2	Consult electorate	2016	CR	3	2	663	Level A

Notes:
The response probability (*rp*) for reporting item parameters was set at *rp* = 0.62.
MC = multiple choice; CR = constructed response; NGO = nongovernmental organization.

APPENDIX D: MAPPING OF QUESTIONNAIRE ITEMS COMMON TO ICCS 2009 AND 2016

Table D.1: Mapping of student questionnaire items between 2009 and 2016

ICCS 2016 question	Variable name(s)	Mapping against ICCS 2009
Q1	IS3G01A, IS3G01B	Question is the same as ICCS 2009 Q1
Q2	IS3G02	Question is the same as ICCS 2009 Q2
Q2B	IS3G02B	Question is the same as ICCS 2009 Q2B
Q3	IS3G03	Question has been modified from ICCS 2009 Q3 (ISCED level categories have been changed)
Q3B	IS3G03BA to IS3G03BG	Question is the same as ICCS 2009 Q11B
Q4	IS3G04A to IS3G04C	Unchanged from ICCS 2009 Q4
Q5	IS3G05	Unchanged from ICCS 2009 Q5
Q6	IS3G06A, IS3G06B	Unchanged from ICCS 2009 Q6
Q7	IS3G07	Question has been modified from ICCS 2009 Q7 (ISCED level categories have been changed)
Q8	IS3G08A, IS3G08B	Unchanged from ICCS 2009 Q8
Q9	IS3G09	Qtuestion has been modified from ICCS 2009 Q9 (ISCED level categories have been changed)
Q10	IS3G10A to IS3G10C	Question has been modified from ICCS 2009 Q10 (New item added)
Q11	IS3G11	Question has been modified from ICCS 2009 Q11 (changes to instructions and response options)
Q12	IS3G12A to IS3G12C	Question new for ICCS 2016
Q13	IS3G13	Question new for ICCS 2016
Q14	IS3G14A to IS3G14I	Question modified from ICCS 2009 Q13 (changes to stem, item modifications and new items added). S_POLDISC scale is comparable across cycles
Q15	IS3G15A to IS3G15J	Question modified from ICCS 2009 Q14 (item modifications and new items added).
Q16	IS3G16A to IS3G16G	Question modified from ICCS 2009 Q15 (item modifications and new items added).
Q17	IS3G17A to IS3G17F	Unchanged from ICCS 2009 Q16. S_OPDISC scale is comparable across cycles
Q18	IS3G18A to IS3G18G	Question new for ICCS 2016
Q19	IS3G19A to IS3G19J	Question modified from ICCS 2009 Q18 (changes to stem, item modifications and new items added). S_STUTREL scale is comparable across cycles
Q20	IS3G20A to IS3G20F	Question new for ICCS 2016
Q21	IS3G21A to IS3G21E	Question modified from ICCS 2009 Q19 (new item added). S_VALPARTS scale is comparable across cycles
Q22	IS3G22A to IS3G22I	Question new for ICCS 2016
Q23	IS3G23A to IS3G23Q	Question modified from ICCS 2009 Q21 (new items added). S_CITCON and S_CITSOC scales are comparable across cycles
Q24	IS3G24A to IS3G24G	Unchanged from ICCS 2009 Q24. S_GENEQL is comparable across cycles
Q25	IS3G25A to IS3G25E	Unchanged from ICCS 2009 Q25. S_ETHRGHT is comparable across cycles
Q26	IS3G26A to IS3G26O	Question modified from ICCS 2009 Q27 (changes made to stem, item added). S_INTRUST scale is comparable across cycles
Q27	IS3G27A to IS3G27E	Question modified from ICCS 2009 Q28 (some items dropped)
Q28	IS3G28A to IS3G28M	Question new for ICCS 2016
Q29	IS3G29A to IS3G29G	Unchanged from ICCS 2009 Q30
Q30	IS3G30A to IS3G30K	Question modified from ICCS 2009 Q31/Q33 (changes to stem, item modifications and new items added)
Q31	IS3G31A to IS3G31L	Question modified from ICCS 2009 Q32/Q33 (changes to stem and new items added). S_ELECPART and S_POLPART scales are comparable across cycles

Table D.1: Mapping of student questionnaire items between 2009 and 2016 (contd.)

ICCS 2016 question	Variable name(s)	Mapping against ICCS 2009
Q32	IS3G32A to IS3G32E	Question new for ICCS 2016
Q33	IS3G33	Unchanged from ICCS 2009 Q34
Q34	IS3G34	Unchanged from ICCS 2009 Q35
Q35	IS3G35A to IS3G35G	Question modified from ICCS 2009 Q36 (new items added). S_RELINF scale is comparable across cycles

Table D.2: Mapping of European questionnaire items between 2009 and 2016

ICCS 2016 question	Variable name(s)	Mapping against ICCS 2009
Q1	ES3G01A to ES3G01F	Unchanged from ICCS 2009 Q1. E_EUIDENT scale is comparable across cycles
Q2	ES3G02A to ES3G02D	Question modified from ICCS 2009 Q3 (changes to stem, response options and items)
Q3	ES3G03A to ES3G03F	Question modified from ICCS 2009 Q8 (changes to stem and items)
Q4	ES3G04A to ES3G04E	Question modified from ICCS 2009 StQ Q26 (minor change to stimulus, one less item). E_IMMRGHT scale is comparable across cycles
Q5	ES3G05A to ES3G05H	Question new for ICCS 2016
Q6	ES3G06A to ES3G06G	Question new for ICCS 2016
Q7	ES3G07A to ES3G07H	Question new for ICCS 2016
Q8	ES3G08A to ES3G08E	Question new for ICCS 2016
Q9	ES3G09A to ES3G09F	Question new for ICCS 2016
Q10	ES3G10A to ES3G10K	Question new for ICCS 2016
Q11	ES3G11A to ES3G11E	Question new for ICCS 2016

Table D.3: Mapping of Latin American questionnaire items between 2009 and 2016

ICCS 2016 question	Variable name(s)	Mapping against ICCS 2009
P1	LS3G01B to LS3G01F	Unchanged from ICCS 2009 P2. LAUTGOV is comparable across cycles
P2	LS3G02A to LS3G02E	Unchanged from ICCS 2009 P3
P3	LS3G03A to LS3G03F	Unchanged from ICCS 2009 P4. L_ATTCORR is comparable across cycles
P4	LS3G04A to LS3G04J	Modified from ICCS 2009 P7/P8 (questions combined, new items added, L_ATTVIOL scale comparable across cycles)
P5	LS3G05A to LS3G05J	Unchanged from ICCS 2009 P5. L_DISLAW is comparable across cycles
P6	LS3G06A to LS3G06H	Modified from ICCS 2009 P6 (changes to response options, items)
P7	LS3G07A to LS3G07K	Modified from ICCS 2009 P9 (changes to response options, items)
P8	LS3G08A to LS3G08E	Modified from ICCS 2009 P11 (changes to items)
P9	LS3G09A to LS3G09K	Question new for ICCS 2016

Table D.4: Mapping of teacher questionnaire items between 2009 and 2016

ICCS 2016 question	Variable name(s)	Mapping against ICCS 2009
Q1	IT3G01A to IT3G01F	Question modified from ICCS 2009 Q1 (item modifications and new items added)
Q2	IT3G02	Question modified from ICCS 2009 Q3 (changes to stem)
Q3	IT3G03	Unchanged from ICCS 2009 Q7
Q4	IT3G04	Unchanged from ICCS 2009 Q4
Q5	IT3G05	Unchanged from ICCS 2009 Q8
Q6	IT3G06A to IT3G06E	Question modified from ICCS 2009 Q11 (changes to stem, item modifications and new items added).
Q7	IT3G07A to IT3G07I	Question modified from ICCS 2009 Q14 (changes to stem, item modifications)
Q8	IT3G08A to IT3G08I	Question modified from ICCS 2009 Q15 (changes to stem, item modifications and new items added).
Q9	IT3G09A to IT3G09F	Unchanged from ICCS 2009 Q17
Q10	IT3G10A to IT3G10D	Unchanged from ICCS 2016 Q20.
Q11	IT3G11A to IT3G11H	Question new for ICCS 2016
Q12	IT3G12A to IT3G12G	Question new for ICCS 2016
Q13	IT3G13A to IT3G13E	Question new for ICCS 2016
Q14	IT3G14A to IT3G14J	Question modified from ICCS 2009 Q21 (item modifications)
Q15	IT3G15	Unchanged from ICCS 2009 Q23
Q16	IT3G16A to IT3G16H	Question modified from ICCS 2009 Q24 (changes to stem, item modifications and new items added).
Q17	IT3G17A to IT3G17H	Question modified from ICCS 2009 Q25 (changes to stem, item modifications and new items added).
Q18	IT3G18A to IT3G18L	Question modified from ICCS 2009 Q28 (changes to stem, response options, item modifications and new items added).
Q19	IT3G19A to IT3G19L	Question new for ICCS 2016
Q20	IT3G20A to IT3G20E	Question new for ICCS 2016
Q21	IT3G21A to IT3G21E	Question modified from ICCS 2009 Q27 (changes to stem and item modifications).
Q22	IT3G22A to IT3G22N	Question modified from ICCS 2009 Q29 (modified set of items).

Table D.5: Mapping of school questionnaire items between 2009 and 2016

ICCS 2016 question	Variable name(s)	Mapping against ICCS 2009
Q1	IC3G01	Question modified from ICCS 2009 Q1 (stem and response options modified)
Q2	IC3G02A to IC3G02E	Question modified from ICCS 2009 Q5 (modified and reduced item set)
Q3	IC3G03A to IC3G03H	Question modified from ICCS 2009 Q12 (stem, items and response options modified)
Q4	IC3G04A to IC3G04I	Question modified from ICCS 2009 Q6 (changes to items)
Q5	IC3G05A, IC3G05B	Question modified from ICCS 2009 Q7 (changes to items)
Q6	IC3G06A to IC3G06F	Question new for ICCS 2016
Q7	IC3G07A to IC3G07H	Question new for ICCS 2016
Q8	IC3G08A to IC3G08F	Question new for ICCS 2016
Q9	IC3G09A to IC3G09E	Question new for ICCS 2016
Q10	IC3G10A to IC3G10A	Question new for ICCS 2016
Q11	IC3G11A to IC3G11J	Question modified from ICCS 2009 Q13 (changes to stem, new item)
Q12	IC3G12A to IC3G12L	Question modified from ICCS 2009 Q14 (changes to stem, modification of item)
Q13	IC3G13A to IC3G13E	Question modified from ICCS 2009 Q16 (changes to items)
Q14	IC3G14A to IC3G14I	Question modified from ICCS 2009 Q4 (changes to stem, new items)
Q15	IC3G15	Question modified from ICCS 2009 Q18 (changes to stem, response options)
Q16	IC3G16A to IC3G16J	Question modified from ICCS 2009 Q17 (changes to items)
Q17	IC3G17	Unchanged from ICCS 2009 Q19
Q18	IC3G18A, IC3G18B	Unchanged from ICCS 2009 Q20
Q19	IC3G19A, IC3G19B	Unchanged from ICCS 2009 Q21
Q20	IC3G20	Question modified from ICCS 2009 Q23 (changes to stem)
Q21	IC3G21A, IC3G21B	Question new for ICCS 2016

APPENDIX E: CODING INFORMATION FOR THE ITEMS IN THE QUESTIONNAIRES AND THE TEST

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
S_Age	Age of student		Imputed from year and month	Copy value (Missing = mean)	Valid = 0 Missing = 1
SGENDER	Gender	0. boy 1. girl	0 1	Copy value (Missing = 0)	Valid = 0 Missing = 1
IS3G02B	Student's ethnic background	1. <A> 2. <B> 3. <C> 4. <D>		Length of dummy code varies across countries, Dummy coding for national options, largest group is reference category (Missing = reference category)	Valid = 0 Missing = 1
S_ISCED	Highest level of education you expect to complete	1 = <ISCED level 6, 7 or 8> 2 = <ISCED level 4 or 5> 3 = <ISCED level 3> 4 = <ISCED level 2> or below		Dummy coding for S_ISCED variable, option with the highest percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
FAMSTRUC	Family structure	Derived index (IS3G03BA - G)		Dummy coding for FAMSTRUC variable, option with the highest percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
IS3G04A	Student's country of birth	1. <Country of test> 2. <Other country/Group A> 3. <Other country/Group B> 4. <Another country>		Length of dummy code varies across countries, option with the highest percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
IS3G04B	Mother's or <female guardian's> country of birth				
IS3G04C	Father's or <male guardian's> country of birth				
IS3G05	Language use at home	1. <Language of test> 2. <Other language 1> 3. <Other language 2> 4. <Another language>		Length of dummy code varies across countries, option with the highest percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
S_MISEI	Mother's occupational status	ISEI score (IS3G06A)		Copy value (Missing = mean)	Valid = 0 Missing = 1
S_MISCED	Highest level of education completed by your mother or <female guardian>	1 = <ISCED level 6, 7 or 8> 2 = <ISCED level 4 or 5> 3 = <ISCED level 3> 4 = <ISCED level 2> 5 = She did not complete <ISCED level 2>		Dummy coding for S_MISCED variable, Option with the highest is reference category (Missing = reference category)	Valid = 0 Missing = 1

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
S_FISEI	Father's occupational status	ISEI score (IS3G08A)		Copy value (Missing = mean)	Valid = 0 Missing = 1
S_FISCED	Highest level of education completed by your father or <male guardian>	1 = <ISCED level 6, 7 or 8> 2 = <ISCED level 4 or 5> 3 = <ISCED level 3> 4 = <ISCED level 2> 5 = He did not complete <ISCED level 2>		Dummy coding for S_FISCED variable, Option with the highest percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
IS3G10A	Interest in political and social issues - Student	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
IS3G10B	Interest in political and social issues – Mother or <female guardian>				
IS3G10C	Interest in political and social issues – Father or <male guardian>				
IS3G11	Books at home	1. 0–10 books 2. 11–25 books 3. 26–100 books 4. 101–200 books 5. More than 200 books		Dummy coding for IS3G11 variable, Option with the highest percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
IS3G12A	Number of regularly used devices – desktop or portable computers (laptop, notebook or netbook)	1. None 2. One 3. Two 4. Three or more	0 1 2 3	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
IS3G12B	Number of regularly used devices – tablet devices or e-readers (e.g. <iPad> or <Kindle>)				
IS3G12C	Number of regularly used devices – mobile phones with internet access (e.g. <smart phones>)				
IS3G13	Internet connection at home	1. Yes 2. No	0 1	Copy value (Missing = 0)	Valid = 0 Missing = 1
IS3G14A	Talk about political or social issues – your parent(s)	1. Never or hardly ever 2. Monthly (at least once a month) 3. Weekly (at least once a week) 4. Daily or almost daily	0 1 2 3	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
IS3G14D	Talk about political or social issues – your friends				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G14E	Talk about what is happening in other countries – your parent(s)				
IS3G14F	Talk about what is happening in other countries – your friends				
IS3G14G	Political or social issues – using the internet to find information	1. Never or hardly ever 2. Monthly (at least once a month) 3. Weekly (at least once a week) 4. Daily or almost daily	0 1 2 3	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
IS3G14H	Political or social issues – posting a comment or image on the internet or social media				
IS3G14I	Political or social issues – sharing or commenting on another person's online post				
IS3G14B	National and international news – watching TV	1. Never or hardly ever 2. Monthly (at least once a month) 3. Weekly (at least once a week) 4. Daily or almost daily	0 1 2 3	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
IS3G14C	National and international news – reading newspaper				
IS3G15A	A youth organization affiliated with a political party or union	1. Yes, I have done this within the last twelve months 2. Yes, I have done this but more than a year ago 3. No, I have never done this	2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean) (* Optional variables)	Valid = 0 Missing = 1
IS3G15B	An environmental action group or organization				
IS3G15C	A human rights organization				
IS3G15D	A voluntary group doing something to help the community				
IS3G15E	An organization collecting money for a social cause				
IS3G15F	A group of young people campaigning for an issue				
IS3G15G	An animal rights or animal welfare group				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G15H	A religious group or organization	1. Yes, I have done this within the last twelve 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) (* Optional variables)	Valid = 0 Missing = 1
IS3G15I	A community youth group (such as <boys/girls scouts, YMCA>)				
IS3G15J	A community youth group (such as a sports team)				
IS3G16A	Active participation in an organized debate	1. Yes, I have done this within the last twelve months 2. Yes, I have done this but more than a year ago 3. No, I have never done this	2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G16B	Voting for <class representative> or <school parliament>				
IS3G16C	Taking part in decision - making about how the school is run				
IS3G16D	Taking part in discussions at a <student assembly>				
IS3G16E	Becoming a candidate for <class representative> or <school parliament>				
IS3G16F	Participating in an activity to make the school more <environmentally friendly>				
IS3G16G	Voluntary participation in school based music or drama activities outside of regular classes	1. Yes, I have done this within the last twelve 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) (* Optional variables)	Valid = 0 Missing = 1
IS3G17A	Teachers encourage students to make up their own minds	1. Never 2. Rarely 3. Sometimes 4. Often	0 1 2 3	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G17B	Teachers encourage students to express their opinions				
IS3G17C	Students bring up current political events for discussion in class				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G17D	Students express opinions in class even when their opinions are different				
IS3G17E	Teachers encourage to discuss the issues with people having different opinions				
IS3G17F	Teachers present several sides of the issues when explaining them in class				
IS3G18A	How citizens can vote in local or national elections	1. To a large extent 2. To a moderate extent 3. To a small extent 4. Not at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G18B	How laws are introduced and changed in <country of test>				
IS3G18C	How to protect the environment				
IS3G18D	How to contribute to solving problems in the <local community>				
IS3G18E	How citizen rights are protected in <country of test>				
IS3G18F	Political issues and events in other countries				
IS3G18G	How the economy works				
IS3G19A	Most of my teachers treat me fairly	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G19B	Students get along well with most teachers.				
IS3G19C	Most teachers are interested in students' well-being.				
IS3G19D	Most of my teachers listen to what I have to say.				
IS3G19E	If I need extra help, I receive it from my teachers.				
IS3G19F	Most teachers would stop students from being bullied.				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G19G	Most students at my school treat each other with respect.	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
IS3G19H	Most students at my school get along well with each other.				
IS3G19I	My school is a place where students feel safe.				
IS3G19J	I am afraid of being bullied by other students	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
IS3G20A	A student called you by an offensive nickname	1. Not at all 2. Once 3. 2 to 4 times 4. 5 times or more	0 1 2 3	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G20B	A student said things about you to make others laugh				
IS3G20C	A student threatened to hurt you				
IS3G20D	You were physically attacked by another student				
IS3G20E	A student broke something belonging to you on purpose				
IS3G20F	A student posted offensive pictures or text about you on the Internet				
IS3G21A	Student participation in how schools are run can make schools better	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G21B	Lots of positive changes can happen in schools when students work together				
IS3G21C	Organizing groups of students to express their opinions could help solve problems				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G21D	Students can have more influence on what happens in schools if they act together				
IS3G21E	Voting in student elections can make a difference to what happens at schools				
IS3G22A	Political leaders give government jobs to their family members	1. Good for democracy 2. Neither good nor bad for democracy 3. Bad for democracy	0 1 2	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
IS3G22B	One company or the government owns all newspapers in a country				
IS3G22F	The police have right to hold people in jail without trial				
IS3G22H	The government influences decisions by courts of justice				
IS3G22C	People are allowed to publicly criticize the government	1. Good for democracy 2. Neither good nor bad for democracy 3. Bad for democracy	2 1 0	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
IS3G22D	All adult citizens have the right to elect their political leaders				
IS3G22E	People are able to protest if they think a law is unfair				
IS3G22G	Differences in income between poor and rich people are small				
IS3G22I	All <ethnic/racial> groups in the country have the same rights				
IS3G23A	Voting in every national election	1. Very important 2. Quite important 3. Not very important 4. Not important at all	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
IS3G23B	Joining a political party				
IS3G23C	Learning about the country's history				
IS3G23B	Joining a political party				
IS3G23C	Learning about the country's history				
IS3G23D	Following political issues in the newspaper, on the radio, on TV or on the Internet				
IS3G23F	Engaging in political discussions				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G23E	Showing respect for government representatives	1. Very important 2. Quite important 3. Not very important 4. Not important at all	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
IS3G23K	Working hard				
IS3G23L	Always obeying the law				
IS3G23M	Ensuring the economic welfare of their families				
IS3G23O	Respecting the rights of others to have their own opinions				
IS3G23P	Supporting people who are worse off than you				
IS3G23G	Participating in peaceful protests against laws believed to be unjust	1. Very important 2. Quite important 3. Not very important 4. Not important at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G23H	Participating in activities to benefit people in the <local community>				
IS3G23I	Taking part in activities promoting human rights				
IS3G23J	Taking part in activities to protect the environment				
IS3G23N	Making personal efforts to protect natural resources				
IS3G23Q	Engaging in activities to help people in less developed countries				
IS3G24A	Men and women should have equal opportunities to take part in government	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 (C, D, F = 0) 2 (C, D, F = 1) 1 (C, D, F = 2) 0 (C, D, F = 3)	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G24B	Men and women should have the same rights in every way				
IS3G24C	Women should stay out of politics				
IS3G24D	Not many jobs available, men should have more right to a job than women				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G24E	Men and women should get equal pay when they are doing the same jobs				
IS3G24F	Men are better qualified to be political leaders than women				
IS3G24G	Women's first priority should be 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
IS3G25A	All <ethnic/racial groups> should have an equal chance to get a good education	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G25B	All <ethnic/racial groups> should have an equal chance to get good jobs				
IS3G25C	Schools should teach students to respect <members of all ethnic/racial groups>				
IS3G25D	<Members of all ethnic/racial groups> should be encouraged to run in elections				
IS3G25E	<Members of all ethnic/racial groups> should have the same rights and responsibilities				
IS3G26A	The <national government> of <country of test>	1. Completely 2. Quite a lot 3. A little 4. Not at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G26B	The <local government> of your town or city				
IS3G26C	Courts of justice				
IS3G26D	The police				
IS3G26E	Political parties				
IS3G26F	<National Parliament>				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G26G	Trust institutions – Media	1. Completely 2. Quite a lot 3. A little 4. Not at all	3 2 1 0	Recode as per scoring (Missing = median) (* M, N, O - Optional variables)	Valid = 0 Missing = 1
S3G26H	Trust institutions – Social media				
IS3G26I	Trust institutions – The Armed Forces				
IS3G26J	Trust institutions – Schools				
IS3G26K	Trust institutions – The United Nations				
IS3G26L	Trust institutions – People in general				
IS3G26M	Trust institutions – <state/province> government				
IS3G26N	Trust institutions – European Commission				
IS3G26O	Trust institutions – European Parliament				
IS3G27A	The <flag of country of test> is important to me.	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G27B	I have great respect for <country of test>.				
IS3G27C	In <country of test> we should be proud of what we have achieved.				
IS3G27D	I am proud to live in <country of test>.				
IS3G27E	<Country of test> is a better country to live in than most other countries.				
IS3G28A	Threat to the world's future - Pollution	1. To a large extent 2. To a moderate extent 3. To a small extent 4. Not at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G28B	Threat to the world's future - Energy shortages				
IS3G28C	Threat to the world's future - Global financial crises				
IS3G28D	Threat to the world's future - Crime				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G28E	Threat to the world's future – Water shortages				
IS3G28F	Threat to the world's future – Violent conflict				
IS3G28G	Threat to the world's future – Poverty				
IS3G28H	Threat to the world's future – Food shortages				
IS3G28I	Threat to the world's future – Climate change				
IS3G28J	Threat to the world's future – Unemployment				
IS3G28K	Threat to the world's future – Overpopulation				
IS3G28L	Threat to the world's future – Infectious diseases (e.g. <bird flu>, <AIDS>)				
IS3G28M	Threat to the world's future – Terrorism				
IS3G29A	Discuss a newspaper article about a conflict between countries	1. Very well 2. Fairly well 3. Not very well 4. Not at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G29B	Argue your point of view about a controversial political or social issue				
IS3G29C	Stand as a candidate in a <school election>				
IS3G29D	Organize a group of students in order to achieve changes at school				
IS3G29E	Follow a television debate about a controversial issue				
IS3G29F	Write a letter or email to a newspaper giving your view on a current issue				
IS3G29G	Speak in front of your class about a social or political issue				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G30A	Express opinion – Talk to others about your views	1. I would certainly do this 2. I would probably do this 3. I would probably not do this 4. I would certainly not do this	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
IS3G30B	Express opinion – Contact an <elected representative>				
IS3G30C	Express opinion – Take part in a peaceful march or rally				
IS3G30D	Express opinion – Collect signatures for a petition				
IS3G30E	Express opinion – Contribute to an online discussion forum				
IS3G30F	Express opinion – Organize an online group to take a stance				
IS3G30G	Express opinion – Participate in an online campaign				
IS3G30H	Express opinion – Choose to buy certain products in support of social justice				
IS3G30I	Express opinion – Spray - paint protest slogans on walls	1. I would certainly do this 2. I would probably do this 3. I would probably not do this 4. I would certainly not do this	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G30J	Express opinion – Stage a protest by blocking traffic				
IS3G30K	Express opinion – Occupy public buildings as a sign of protest				
IS3G31A	Adult – Vote in <local elections>	1. I would certainly do this 2. I would probably do this 3. I would probably not do this 4. I would certainly not do this	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G31B	Adult – Vote in <national elections>				
IS3G31C	Adult – Get information about candidates before voting				
IS3G31D	Adult – Help a candidate or party during an election campaign	1. I would certainly do this 2. I would probably do this 3. I would probably not do this 4. I would certainly not do this	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G31E	Adult – Join a political party				
IS3G31F	Adult – Join a trade union				
IS3G31G	Adult – Stand as a candidate in <local elections>				
IS3G31H	Adult – Join an organization for a political or social cause				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G31I	Adult – Volunteer time to help people in <local community>	1. I would certainly do this 2. I would probably do this 3. I would probably not do this 4. I would certainly not do this	3 2 1 0	Recode as per scoring (Missing = median) (* K, L - Optional variables)	Valid = 0 Missing = 1
IS3G31J	Adult – Make personal efforts to help the environment				
IS3G31K	Adult – Vote in <state, province elections>				
IS3G31L	Adult – Vote in European elections				
IS3G32A	Vote in a school election of <class representatives> or <school parliament>	1. Very likely 2. Quite likely 3. Not very likely 4. Not at all likely	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
IS3G32B	Join a group of students campaigning for an issue you agree with				
IS3G32C	Become a candidate for <class representative> or <school parliament>				
IS3G32D	Take part in discussions in a <student assembly>				
IS3G32E	Participate in writing articles for a school newspaper or website				
IS2P33	Student's religion	1. No religion 2. [A] 3. [B] 4. [C] 5. [D] 6. Other religion		Length of dummy code varies across countries, first option is reference category (Missing = 1) (* Optional variables)	Valid = 0 Missing = 1
IS3G34	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	0 1 2 3 4	Recode as per scoring (Missing = median) (* Optional variables)	Valid = 0 Missing = 1

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS3G35A	Religion is more important to me than what is happening in national politics.	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean) (* Optional variables)	Valid = 0 Missing = 1
IS3G35B	Religion helps me to decide what is right and what is wrong.				
IS3G35C	Religious leaders should have more power in society.				
IS3G35D	Religion should influence people's behavior towards others.				
IS3G35E	Rules of life based on religion are more important than civil laws.				
IS3G35G	Religious people are better citizens.				
IS3G35F	All people should be free to practice the religion they choose.	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing = median) (* Optional variables)	Valid = 0 Missing = 1
European regional questionnaire					
ES3G01A	I see myself as European	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G01B	I am proud to live in Europe				
ES3G01C	I feel part of Europe				
ES3G01D	I see myself first as a citizen of Europe and then as a citizen of the world				
ES3G01E	I feel part of the European Union	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing = median) (* Optional variables)	Valid = 0 Missing = 1
ES3G01F	I am proud that my country is a member of the European Union				
ES3G02A	Political and economic systems of other European countries	1. To a large extent 2. To a moderate extent 3. To a small extent 4. Not at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G02B	The history of Europe				
ES3G02C	Political and social issues in other European countries				
ES3G02D	Political and economic integration between European countries				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
ES3G03A	Good for the European economy	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G03B	Allowed to work anywhere in Europe				
ES3G03C	Helps to reduce unemployment				
ES3G03D	Only if their skills are needed	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G03E	Jobs that no one in the other wants				
ES3G03F	Only a limited number of people				
ES3G04A	<Immigrants> should have the opportunity to continue speaking their own language	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
ES3G04B	<Immigrant> children should have the same opportunities for education	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G04C	<Immigrants> who live in a country for several years should have the opportunity to vote				
ES3G04D	<Immigrants> should have the opportunity to continue their own customs and lifestyle				
ES3G04E	<Immigrants> should have the same rights that everyone else in the country has				
ES3G05A	European countries should cooperate to protect the environment	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G05B	European countries should cooperate to guarantee high levels of employment				
ES3G05C	European countries should cooperate to strengthen their economies				
ES3G05D	European countries should recognize all educational qualifications achieved in any other European country				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
ES3G05E	European countries should have a European army for peace keeping missions				
ES3G05F	European countries should cooperate to prevent and combat terrorism				
ES3G05G	European countries should cooperate to combat illegal entry from non - European countries				
ES3G05H	European countries should cooperate to provide shelter to people escaping persecution				
ES3G06A	In <country of test> it is common that women have lower salaries	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G06B	In <country of test> <immigrants> are more exposed to unfair treatment than other groups				
ES3G06C	In <country of test> gay and lesbian people are often <bullied>				
ES3G06G	In <country of test> young people are often discriminated against				
ES3G06D	In <country of test> there is less discrimination than in other European countries	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
ES3G06E	There is only a limited amount of discrimination in <country of test>				
ES3G06F	There is less discrimination in Europe than in other parts of the world				
ES3G07A	There will be stronger cooperation among European countries	1. Very likely 2. Likely 3. Unlikely 4. Very unlikely	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
ES3G07B	There will be greater peace across Europe				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
ES3G07F	There will be less air and water pollution in Europe				
ES3G07H	Democracy will be strengthened across Europe				
ES3G07C	Terrorism will be more of a threat all across Europe	1. Very likely 2. Likely 3. Unlikely 4. Very unlikely	0 1 2 3	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G07D	Europe will be more influenced by non - European powers				
ES3G07E	The economy will be weaker in all European countries				
ES3G07G	There will be a rise in poverty and unemployment in Europe				
ES3G08A	I will find a steady job	1. Very likely 2. Likely 3. Unlikely 4. Very unlikely	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G08B	My financial situation will be better than that of my parents				
ES3G08C	I will find a job I like				
ES3G08D	I will have the opportunity to travel abroad for leisure				
ES3G08E	I will earn enough money to start a family				
ES3G09A	People should not buy goods coming from non - democratic countries	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G09B	People should not buy goods produced by companies using child labor				
ES3G09C	People should not buy products whose production has a negative impact on the environment				
ES3G09D	People should not buy goods produced by a company violating social rights of their employees				
ES3G09E	People should buy only products that can be recycled afterwards				
ES3G09F	People should buy <green products> even if they are more expensive				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
ES3G10A	Buy beer	1. Less than 14 years old 2. 14 years old 3. 16 years old 4. 18 years old 5. 20 years old or more		Dummy coding, largest group is reference category (Missing = reference category)	Valid = 0 Missing = 1
ES3G10B	Buy spirits				
ES3G10C	Vote in national elections				
ES3G10D	Get a car driving license				
ES3G10E	Get a credit card				
ES3G10F	Buy cigarettes				
ES3G10G	Get a job				
ES3G10H	Get a motorbike driving license				
ES3G10I	Join a social network (e.g. Facebook, Twitter)				
ES3G10J	Vote in local elections				
ES3G10K	Get married				
ES3G11A	<EU> guarantees respect for human rights all over Europe	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
ES3G11B	<EU> makes Europe a safe place to live				
ES3G11C	<EU> takes care of the environment				
ES3G11D	<EU> is good for the economy of individual countries				
ES3G11E	<EU> is good because countries share a common set of rules and laws				
Latin American regional questionnaire					
LS3G01A	It is better for government leaders to make decisions without consulting anybody	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
LS3G01B	People in government must enforce their authority even if it means violating the rights				
LS3G01C	People in government lose part of their authority when they admit their mistakes				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
LS3G01D	People whose opinions different than those of the government must be considered its enemies				
LS3G01E	The most important opinion of a country should be that of the president				
LS3G01F	It is fair that the government does not comply with law when it thinks it is not necessary				
LS3G02A	Concentration of power in one person guarantees order				
LS3G02B	The government should close communication media that are critical				
LS3G02C	If the president does not agree with <Congress>, he/she should dissolve it				
LS3G02D	Dictatorships are justified when they bring order and safety	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
LS3G02E	Dictatorships are justified when they bring economic benefits				
LS3G03A	It is acceptable for a civil servant to accept bribes if his salary is too low	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
LS3G03B	It is acceptable for a civil servant to use the resources of the institution				
LS3G03C	Good candidates grant personal benefits to voters in return for their votes				
LS3G03D	Paying an additional amount to a civil servant in order to obtain a personal benefit				
LS3G03E	It is acceptable that a civil servant helps his/her friends by giving them employment				
LS3G03F	Since public resources belong to everyone, acceptable that those who can keep part of them				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
LS3G04A	Peace is only achieved through dialogue and negotiation	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
LS3G04B	To achieve peace, the means justify the end				
LS3G04C	If the authorities fail to act, the citizens should organize themselves to punish criminals				
LS3G04D	Hitting is a justified punishment when someone commits a crime against my family				
LS3G04E	He who does me harm will have to pay for it	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
LS3G04F	Watching fights between classmates is fun				
LS3G04G	If you can't succeed by doing good things, <try> the bad ones				
LS3G04H	You have to fight so people do not think you are a coward				
LS3G04I	Revenge is sweet				
LS3G04J	Aggression serves to achieve what one wants				
LS3G05A	When it is the only alternative left for achieving important objectives	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score	Valid = 0 Missing = 1
LS3G05B	When it is the only way one has to help one's family				
LS3G05C	When others who disobeyed it were not punished				
LS3G05D	When others do it				
LS3G05E	When one distrusts the enacting body				
LS3G05F	When one is sure nobody will realize				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
LS3G05G	When it is the only way of fighting publicly against an unfair law	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	0 1 2 3	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
LS3G06A	Persons with different skin color than yours	1. Yes 2. No	0 1	Compute total score if at least two of the variables have valid score (missing = mean)	Valid = 0 Missing = 1
LS3G06B	Persons of a different social class than yours				
LS3G06C	Persons of a different religion than yours				
LS3G06D	Persons who come from another region of the country				
LS3G06E	Persons with physical disabilities				
LS3G06F	Persons with mental disorders				
LS3G06G	Persons from a different country				
LS3G06H	Persons of indigenous origin				
LS3G07A	A classmate falls and gets hurt	1. I think it is fun 2. I don't care 3. It bothers me	0 1 2	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
LS3G07B	A classmate gets beaten up				
LS3G07C	A classmate gets unfairly reprimanded				
LS3G07D	A classmate gets unfairly punished				
LS3G07E	A classmate gets something stolen from him/her				
LS3G07F	A classmate gets ridiculed				
LS3G07G	A classmate gets insulted				
LS3G07H	A classmate looks very sad				
LS3G07I	A classmate gets bad grades				
LS3G07J	A classmate has nobody to play with				
LS3G07K	There is a fight between classmates				

Table E.1: List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge (contd.)

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
LS3G08A	Persons of the same sex should have the right to get married	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Compute total score if at least two of the variables have valid score. (Missing = mean)	Valid = 0 Missing = 1
LS3G08B	Two persons of the same sex should have the right to adopt children				
LS3G08C	Homosexuals should have the same rights as all other citizens				
LS3G08D	All schools should accept homosexuals				
LS3G08E	Homosexuals should have the right to hold any political or public position				
LS3G09A	People discriminate against women	1. A lot 2. To some extent 3. A little 4. Not at all	3 2 1 0	Compute total score if at least two of the variables have valid score (Missing = mean)	Valid = 0 Missing = 1
LS3G09B	People discriminate against young persons				
LS3G09C	People discriminate against homosexual persons				
LS3G09D	People discriminate against unemployed persons				
LS3G09E	People discriminate against persons with a disability				
LS3G09F	People discriminate against persons of African origin				
LS3G09G	People discriminate against religious minorities				
LS3G09H	People discriminate against poor persons				
LS3G09I	People discriminate against older persons				
LS3G09J	People discriminate against <Immigrants>	1. A lot 2. To some extent 3. A little 4. Not at all	3 2 1 0	Recode as per scoring (Missing = median) (* Optional variables)	Valid = 0 Missing = 1
LS3G09K	People discriminate against persons of indigenous origin				

The IEA's International Civic and Citizenship Education Study (ICCS) investigates the ways in which young people are prepared to undertake their roles as citizens in a range of countries in the second decade of the 21st century. ICCS 2016 is the second cycle of a study initiated in 2009.

This technical report follows the publication of several international and regional reports that presented the results of ICCS 2016. It comprises detailed information on the development of the instruments, including their translation and translation verification, sampling design and implementation, field operations and quality control of data collection, data management, sampling weights and participation rates, and the scaling, analysis and reporting of ICCS 2016 data. The technical report enables researchers to evaluate published reports and articles based on data from this study and, used in conjunction with the ICCS 2016 User Guide for the International Database, will provide guidance for their own analyses.

Over the past 50 years, the IEA has conducted comparative research studies in a range of domains focusing on educational policies, practices, and outcomes in many countries around the world. The association conducted its first survey of civic education in 1971. The rich and robust comparative ICCS 2016 database will allow participating education systems to evaluate the strengths of their educational policies, both internationally and within a regional context, and to measure their progress toward achieving critical components of the United Nations' 2030 agenda for sustainable development.

IEA