

ICCS 2022 Technical Report

EDITORS:
Wolfram Schulz
Tim Friedman
Julian Fraillon

Wolfram Schulz

The Australian Council for Educational Research
Camberwell, Victoria
Australia

Tim Friedman

The Australian Council for Educational Research
Camberwell, Victoria
Australia

Julian Fraillon

International Association for the Evaluation of Educational Achievement
Amsterdam, The Netherlands

Edited by:

Wolfram Schulz
Tim Friedman
Julian Fraillon

Contributors:

John Ainley
Gabiella Agrusti
Valeria Damiani
Lauren Musu
Kateřina Hartmanová
Jan-Philipp Wagner

Laila Helou
Diego Cortés
Umut Atasever
Hannah Kowolik
Falk Brese
Yasin Afana
Greg Macaskill

IEA Secretariat

Keizersgracht 311
1016 EE Amsterdam
The Netherlands
Tel.: +31 (0)20 625 3625
Fax: +31 20 420 7136
secretariat@iea.nl



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.

© IEA International Association for the Evaluation of Educational Achievement 2024. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

This work is subject to copyright. All commercial rights are reserved by the author(s), whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Regarding these commercial rights a non-exclus. Regarding these commercial rights a non-exclusive license has been granted to the publisher.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Wolfram Schulz • Tim Friedman • Julian Fraillon

ICCS 2022 Technical Report

IEA International Civic and Citizenship
Education Study 2022

Jointly prepared by:

Australian Council for Educational Research (ACER – Melbourne, Australia)
Libera Università Maria Santissima Assunta (LUMSA Università – Rome, Italy)
International Association for the Evaluation of Educational Achievement
(IEA – Amsterdam, The Netherlands, and Hamburg, Germany)

Central contact: iccs@acer.org



Contents

List of Tables and Figures	8
Chapter 1: Overview of ICCS 2022	16
1.1 Background	16
1.2 New developments in ICCS 2022	16
1.3 Research questions	17
1.4 Assessment framework	18
1.5 Participating countries, study design, and implementation	19
1.6 Report structure	20
References	22
Chapter 2: ICCS Civic Knowledge Test Development	24
2.1 Overview	24
2.2 Test scope and format	26
2.3 Test development process	28
2.4 Field trial test design and content	31
2.5 Main survey test design and content	37
2.6 Released test items	41
References	42
Chapter 3: Questionnaire Development	43
3.1 Introduction	43
3.2 Conceptual framework for questionnaire development	43
3.3 Student questionnaire development	45
3.4 Teacher questionnaire development	47
3.5 School questionnaire development	48
3.6 European student questionnaire development	50
3.7 National context survey development	50
References	53
Chapter 4: ICCS Computer-based Systems	54
4.1 Overview	54
4.2 ICCS 2022 computer-based components and architecture	54
4.3 Procedures and software user manuals	54
4.4 ICCS 2022 system architecture	55
4.5 Functions of the computer-based delivery platform for ICCS 2022	55
4.6 Monitoring data completeness from computer-based delivery in ICCS 2022	59
Chapter 5: Preparation and Verification of the ICCS 2022 Study Instruments	62
5.1 Introduction	62
5.2 Adaptation and translation of the ICCS 2022 study instruments	63
5.3 International verification processes	69
Chapter 6: Sampling Design and Implementation	73
6.1 Overview	73
6.2 Target population definitions	73
6.3 Coverage and exclusions	74
6.4 Sample size requirements	76
6.5 School sampling design	79
6.6 Within school sampling design	81
References	82

Chapter 7: Field Operations Procedures	83
7.1 Overview	83
7.2 Field operations personnel	83
7.3 Field operations resources	84
7.4 Field operations processes	86
7.5 Scoring the ICCS assessment and checking scorer reliability	93
7.6 Field trial procedures	94
Chapter 8: Quality Assurance in ICCS 2022	95
8.1 International quality assurance program	95
8.2 Findings from the ICCS 2022 administration	96
8.3 Findings from the national materials review	105
8.4 Findings from the survey activities questionnaire	105
8.5 Summary	115
Chapter 9: Data Management, Cleaning, and Creation of the ICCS International Database	117
9.1 Introduction	117
9.2 Data sources	117
9.3 Data processing and data cleaning	120
9.4 Interim data products	126
9.5 Final product: the ICCS 2022 international database	127
Reference	127
Chapter 10: Weighting Procedures	128
10.1 Overview	128
10.2 Within-school participation requirements	128
10.3 Weighting components for student survey	129
10.4 Weighting components for teacher survey	131
10.5 Weighting components for school survey	133
10.6 Calculating participation rates	134
10.7 ICCS standards for sampling participation	140
References	143
Chapter 11: Scaling Procedures for Civic Knowledge Test Items	144
11.1 Overview	144
11.2 Scaling model	144
11.3 Scaling outcomes	145
11.4 Final scaling and equating procedures	161
References	171
Chapter 12: Scaling Procedures for ICCS 2022 Questionnaire Items	173
12.1 Introduction	173
12.2 Simple Indices	173
12.3 Scaling procedures	176
12.4 Scaled indices	185
12.5 Reviewing psychometric differences across delivery modes	278
References	280

Chapter 13: The Reporting of ICCS 2022 Results	283
13.1 Overview	283
13.2 Estimation of sampling variance	283
13.3 Estimation of imputation variance for civic knowledge scores	285
13.4 Reporting of differences	286
13.5 Hierarchical linear modeling	291
13.6 Multiple regression modeling	292
References	295
Appendix A: Organizations and Individuals Involved in ICCS 2022	296
International study center	296
International association for the evaluation of educational achievement (IEA)	296
ICCS 2022 project advisory committee (PAC)	297
Other project advisors	297
ICCS 2022 National research coordinators (NRCs)	297
Appendix B: Characteristics of National Samples	300
B.1. Bulgaria - Core and bridge samples	300
B.2. Brazil – Core sample	301
B.3. Colombia - Core and bridge samples	302
B.4. Chinese Taipei - Core and bridge samples	304
B.5. Croatia – Core sample	306
B.6. Cyprus – Core sample	307
B.7. Denmark – Core sample	308
B.8. Estonia - Core and bridge samples	309
B.9. France – Core sample	310
B.10. Italy - Core and bridge samples	311
B.11. Lithuania - Core and bridge samples	312
B.12. Latvia - Core and bridge samples	313
B.13. Malta - Core and bridge samples	314
B.14. the Netherlands – Core sample	315
B.15. Norway (9) - Core and bridge samples	316
B.16. Poland – Core samples	317
B.17. Romania – Core sample	318
B.18. Serbia – Core sample	319
B.19. Slovak Republic – Core sample	320
B.20. Slovenia - Core and bridge samples	321
B.21. Spain – Core sample	322
B.22. Sweden - Core and bridge samples	323
B.23. North Rhine-Westphalia (Germany) – Core sample	324
B.24. Schleswig-Holstein (Germany) – Core samples	326
Appendix C: Civic Knowledge Test Items with Associated Framework References, Scaled Item Difficulties and Proficiency Levels	327
Appendix D: Mapping of Questionnaire Items in ICCS 2022 and 2016	331
Appendix E: Tables with Coding Information for Conditioning Variables	336
Appendix F: Tables Comparing Questionnaire Scale Means Across Delivery Modes	345

List of Tables and Figures

Tables

Table 2.1:	Test development process timeline	25
Table 2.2:	ICCS 2022 test development specifications	28
Table 2.3:	Composition of field trial test instrument by item format and origin	32
Table 2.4:	Field trial mapping to assessment framework	33
Table 2.5:	Field trial cluster composition	34
Table 2.6:	Field trial test form design (computer-based and paper-based delivery)	35
Table 2.7:	Field trial test form design (mode-effect study)	35
Table 2.8:	Composition of main survey test instrument by item format and origin	38
Table 2.9:	Main survey item mapping to assessment framework	38
Table 2.10:	Main survey cluster composition	39
Table 2.11:	Main survey test form design (computer-based delivery)	39
Table 2.12:	Main survey test form design (paper-based delivery)	40
Table 2.13:	Main survey test form design (bridging study)	40
Table 2.14:	Characteristics and mapping of ICCS 2022 released test items	41
Table 3.1:	Mapping of variables to contextual framework (examples)	45
Table 4.1:	ICCS computer-based or computer-supported systems and operations	55
Table 5.1:	Languages and delivery modes used for the ICCS 2022 survey instruments	64
Table 6.1:	Population coverage and exclusion rates	76
Table 6.2:	School, student, and teacher sample sizes in ICCS	78
Table 6.3:	School and student sample sizes in the ICCS bridging study	79
Table 7.1:	Hierarchical identification (ID) system	87
Table 7.2:	Time allowed for administering the ICCS student instruments	93
Table 8.1:	Quality of ICCS 2022 administration preparation—Across all 378 observed ICCS administration sessions (percent of IQO responses, rounded to a whole number)	97
Table 8.2:	Quality of ICCS 2022 administration preparation—across PBA administration sessions (percent of IQO responses)	97
Table 8.3:	Quality of ICCS 2022 administration preparation—across CBA administration sessions (percent of IQO responses).	98
Table 8.4:	Test administration sessions —378 ICCS administration sessions (percent of IQO responses)	99
Table 8.5:	Student cooperation during test administration—378 ICCS administration sessions (percent of IQO responses)	99
Table 8.6:	Student questionnaire administration—378 ICCS administration sessions (percent of IQO responses)	100
Table 8.7:	Regional student questionnaire Administration—378 ICCS administration sessions (percent of IQO responses)	100
Table 8.8:	General impressions on the quality of administration—378 ICCS administration sessions (percent of IQO responses)	100
Table 8.9:	General observations of the administration—378 ICCS administration sessions (percent of IQO responses)	101
Table 8.10:	Interview with the school coordinator, overview— (percent of SC responses)	102
Table 8.11:	Interview with the school coordinator, details (percent of SC responses)	103
Table 8.12:	Interview with the test administrator, overview— (percent of TA responses)	104
Table 8.13:	Interview with the test administrator, details (percent of TA responses)	104
Table 8.14:	Contacting schools and recruiting school coordinators (Number of NRC responses, 20)	max 106

Table 8.15:	School recruitment challenges (Number of NRC responses, max 20)	106
Table 8.16:	Recruiting school coordinators (number of NRC responses, max 20)	107
Table 8.17:	Adapting and translating the ICCS assessment materials (Number of NRC responses, max 20)	108
Table 8.18:	Assembling and printing the ICCS assessment materials (Number of NRC responses, max 20)	109
Table 8.19:	Sampling (Number of NRC responses, max 20)	110
Table 8.20:	Administering the booklets and questionnaires (Number of NRC responses, max 20)	111
Table 8.21:	Preparing and administering online questionnaire (number of NRC responses, max 20)	112
Table 8.22:	National quality assurance program (number of NRC responses, max 20)	113
Table 8.23:	Scoring open-ended response items and coding occupational data (number of NRC responses, max 20)	113
Table 8.24:	Entering and submitting data (number of NRC responses)	114
Table 8.25:	Other experiences (number of NRC responses, max 20)	115
Table 10.1:	Unweighted participation rates in the student survey (core)	135
Table 10.2:	Unweighted participation rates in the student survey (bridge)	135
Table 10.3:	Weighted participation rates in the student survey (core)	137
Table 10.4:	Weighted participation rates in the student survey (bridge)	137
Table 10.5:	Unweighted participation rates in the teacher survey	139
Table 10.6:	Weighted participation rates in the teacher survey	140
Table 11.1:	Percentages of omitted and invalid responses by delivery mode	147
Table 11.2:	Average percentages of omitted and invalid responses by delivery mode and item format	151
Table 11.3:	Item facilities, total-score correlations, and weighted item fit for international sample	152
Table 11.4:	Percentages of scorer agreement for open-ended ICCS test items	153
Table 11.5:	Gender DIF estimates for test items	155
Table 11.6:	Summary of problematic items provided to national centers	161
Table 11.7:	National items adjudication	161
Table 11.8:	Final set of ICCS 2022 item parameters used to scale international test items	162
Table 11.9:	Average percentage correct for CBA and PBA data	165
Table 11.10:	ICCS 2022 item parameters from joint calibration with ICCS 2016 data	167
Table 12.1:	Transformation parameters for new ICCS 2022 questionnaire scales (means and standard deviations of original IRT logit scores)	180
Table 12.2:	Transformation parameters and linking error for equated questionnaire scales	183
Table 12.3:	Factor loadings and reliabilities for the national index of students' socioeconomic background in ICCS 2022	186
Table 12.4:	Reliabilities for scales measuring students' participation in out-of-school activities	188
Table 12.5:	Partial credit model item parameters for scales measuring students' activities outside school	189
Table 12.6:	Reliabilities for scales measuring students' participation in civic activities in the community and at school	191
Table 12.7:	Partial credit model item parameters for scales measuring students' participation in civic activities in the community and at school	191
Table 12.8:	Reliabilities for scales measuring students' perceptions of open classroom climates and civic learning	194
Table 12.9:	Partial credit model item parameters for scales measuring students' perceptions of open classroom climates and civic learning	194
Table 12.10:	Reliabilities for scales measuring students' perceptions of school interactions	197
Table 12.11:	Partial credit model item parameters for scales measuring students' perceptions of school interactions	197
Table 12.12:	Reliabilities for scales measuring students' perceptions of political systems and institutions	199
Table 12.13:	Partial credit model item parameters for scales measuring students' perceptions of political systems and institutions	199
Table 12.14:	Reliabilities for scales measuring students' beliefs about threats to democracy and	

	restrictions in national emergencies	201
Table 12.15:	Partial credit model item parameters for scales measuring students' beliefs about threats to democracy and restrictions in national emergencies	201
Table 12.16:	Reliabilities for scales measuring students' perceptions of the importance of good citizenship behavior	204
Table 12.17:	Partial credit model item parameters for scales measuring students' perceptions of the importance of good citizenship behavior	204
Table 12.18:	Reliabilities for scales measuring students' support for equal rights for immigrants and gender equality	207
Table 12.19:	Partial credit model item parameters for scales measuring students' activities outside school	207
Table 12.20:	Reliabilities for scale measuring students' endorsement of equal rights for all ethnic groups in society	209
Table 12.21:	Item parameters for scale measuring students' endorsement of equal rights for all ethnic groups in society	209
Table 12.22:	Reliabilities for scales measuring students' support for environmental protection and their environmental concerns	211
Table 12.23:	Partial credit model item parameters for scales measuring students' activities outside school	211
Table 12.24:	Reliabilities for scales measuring students' citizenship self-efficacy and willingness to participate at school	213
Table 12.25:	Item parameters for scale measuring students' citizenship self-efficacy and willingness to participate at school	213
Table 12.26:	Reliabilities for scales measuring students' expected participation in activities to express their opinions	215
Table 12.27:	Partial credit model item parameters for scales measuring students' expected participation in activities to express their opinions	215
Table 12.28:	Reliabilities for scales measuring students' expected political participation	217
Table 12.29:	Item parameters for scale measuring students' expected political participation	217
Table 12.30:	Reliabilities for scale measuring students' positive attitudes toward religious influence in society	219
Table 12.31:	Item parameters for scale measuring students' positive attitudes toward religious influence in society	219
Table 12.32:	Reliabilities for scales measuring students' European identity and perceptions of learning about Europe at school	221
Table 12.33:	Item parameters for scale measuring students' European identity and perceptions of learning about Europe at school	221
Table 12.34:	Reliabilities for scales measuring students' attitudes toward freedom of movement in Europe	223
Table 12.35:	Item parameters for scale measuring students' attitudes toward freedom of movement in Europe	223
Table 12.36:	Reliabilities for scales measuring students' support for European cooperation	225
Table 12.37:	Item parameters for scale measuring students' support for European cooperation	225
Table 12.38:	Reliabilities for scales measuring students' expectations for Europe's future	227
Table 12.39:	Item parameters for scale measuring students' expectations for Europe's future	227
Table 12.40:	Reliabilities for scale measuring students' expectations for their own future	229
Table 12.41:	Item parameters for scale measuring students' expectations for their own future	229
Table 12.42:	Reliabilities for scales measuring students' reports on ethical consumerism and sustainable behaviors	231
Table 12.43:	Item parameters for scale measuring students' reports on ethical consumerism and sustainable behaviors	231
Table 12.44:	Reliabilities for scales measuring students' attitudes toward the EU	233
Table 12.45:	Item parameters for scale measuring students' attitudes toward the EU	233
Table 12.46:	Reliabilities for scale measuring teachers' perceptions of teacher participation at school	235

Table 12.47:	Item parameters for scale measuring teachers' perceptions of teacher participation at school	235
Table 12.48:	Reliabilities for scale measuring teachers' perceptions of teachers' perceptions of social problems at school	237
Table 12.49:	Item parameters for scale measuring teachers' perceptions of social problems at school	237
Table 12.50:	Reliabilities for scale measuring teachers' perceptions of teachers' reports of student activities in the community	239
Table 12.51:	Item parameters for scale measuring teachers' reports of student activities in the community	239
Table 12.52:	Reliabilities for scale measuring teachers' perceptions of classroom climate	241
Table 12.53:	Item parameters for scale measuring teachers' perceptions of classroom climate	241
Table 12.54:	Reliabilities for scale measuring teachers' perceptions of teachers' reports on student involvement in activities	243
Table 12.55:	Item parameters for scale measuring teachers' reports on student involvement in activities	243
Table 12.56:	Reliabilities for scale measuring teachers' perceptions of teachers' reports on activities dealing with diversity	245
Table 12.57:	Item parameters for scale measuring teachers' reports on activities dealing with diversity	245
Table 12.58:	Reliabilities for scales measuring teachers' perceptions of teachers' opinions on the influence of cultural and ethnic diversity on teaching activities	247
Table 12.59:	Partial credit model item parameters for scales measuring teachers' opinions on the influence of cultural and ethnic diversity on teaching activities	247
Table 12.60:	Reliabilities for scales measuring teachers' opinions on the influence of socioeconomic differences on teaching activities	249
Table 12.61:	Partial credit model item parameters for scales measuring teachers' opinions on the influence of socioeconomic differences on teaching activities	249
Table 12.62:	Reliabilities for scale measuring teachers' reports on activities related to digital technologies	251
Table 12.63:	Item parameters for scale measuring teachers' reports on activities related to digital technologies	251
Table 12.64:	Reliabilities for scales measuring teachers' perceptions of the importance of good citizenship	253
Table 12.65:	Partial credit model item parameters for scales measuring teachers' perceptions of the importance of good citizenship	253
Table 12.66:	Reliabilities for scale measuring teachers' reports on civic-related activities in class	255
Table 12.67:	Item parameters for scale measuring teachers' reports on civic-related activities in class	255
Table 12.68:	Reliabilities for scale measuring teachers' reports on activities related to global issues	257
Table 12.69:	Item parameters for scale measuring teachers' reports on activities related to global issues	257
Table 12.70:	Reliabilities for scales measuring teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics	259
Table 12.71:	Partial credit model item parameters for scales measuring teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics	259
Table 12.72:	Reliabilities for scales measuring teachers' reports on their PD activities	261
Table 12.73:	Partial credit model item parameters for scales measuring teachers' reports on their PD activities	261
Table 12.74:	Reliabilities for scale measuring school principals' perceptions of teacher participation at school	263
Table 12.75:	Item parameters for scale measuring school principals' perceptions of teacher participation at school	263
Table 12.76:	Reliabilities for scales measuring school principals' perceptions of student and parent involvement at school	265
Table 12.77:	Partial credit model item parameters for scales measuring school principals' perceptions of student and parent involvement at school	265
Table 12.78:	Reliabilities for scales measuring school principals' perceptions of school activities in the community	267

Table 12.79:	Partial credit model item parameters for scales measuring school principals' perceptions of school activities in the community	267
Table 12.80:	Reliabilities for scale measuring school principals' reports on environment-friendly practices at school	269
Table 12.81:	Item parameters for scale measuring school principals' reports on environment-friendly practices at school	269
Table 12.82:	Reliabilities for scale measuring school principals' reports on training activities related to the use of digital technologies	271
Table 12.83:	Item parameters for scale measuring school principals' reports on training activities related to the use of digital technologies	271
Table 12.84:	Reliabilities for scale measuring school principals' reports on the availability of resources in local community	273
Table 12.85:	Item parameters for scale measuring school principals' reports on the availability of resources in local community	273
Table 12.86:	Reliabilities for scales measuring school principals' perceptions of social tensions in the community	275
Table 12.87:	Partial credit model item parameters for scales measuring school principals' perceptions of social tensions in the community	275
Table 12.88:	Item parameters for Latin American questionnaire scales	277
Table 12.89:	Linear transformations applied to set Latin American questionnaire scale scores to reporting metric	278
Table 12.90:	Correlations between IRT parameters from computer- and paper-based	278
Table 13.1:	Example of the computation of replicate weights	284
Table 13.2:	National averages for civic knowledge with standard deviations, sampling, and overall errors	286
Table 13.3:	Estimates of equating errors (scale scores)	289
Table 13.4:	Estimates of equating error for percentages	290
Table 13.5:	ICCS 2022 students included in multilevel analyses of variation in civic knowledge	292
Table 13.6:	ICCS 2022 students included in multiple regression analyses of expected electoral and active political participation	294
Table B.1.1:	Allocation of core student sample in Bulgaria	300
Table B.1.2:	Allocation of teacher sample in Bulgaria	300
Table B.1.3:	Allocation of bridge student sample in Bulgaria	301
Table B.2.1:	Allocation of core student sample in Brazil	301
Table B.2.2:	Allocation of teacher sample in Brazil	302
Table B.3.1:	Allocation of core student sample in Colombia	302
Table B.3.2:	Allocation of teacher sample in Colombia	303
Table B.3.3:	Allocation of bridge student sample in Colombia	303
Table B.4.1:	Allocation of core student sample in Chinese Taipei	304
Table B.4.2:	Allocation of teacher sample in Chinese Taipei	305
Table B.4.3:	Allocation of bridge student sample in Chinese Taipei	305
Table B.5.1:	Allocation of core student sample in Croatia	306
Table B.5.2:	Allocation of teacher sample in Croatia	306
Table B.6.1:	Allocation of core student sample in Cyprus	307
Table B.6.2:	Allocation of teacher sample in Cyprus	308
Table B.7.1:	Allocation of core student sample in Denmark	308
Table B.7.2:	Allocation of teacher sample in Denmark	308
Table B.8.1:	Allocation of core student sample in Estonia	309
Table B.8.2:	Allocation of teacher sample in Estonia	309
Table B.8.3:	Allocation of bridge student sample in Estonia	309
Table B.9.1:	Allocation of core student sample in France	310
Table B.9.2:	Allocation of teacher sample in France	310
Table B.10.1:	Allocation of core student sample in Italy	311

Table B.10.2:	Allocation of teacher sample in Italy	311
Table B.10.3:	Allocation of bridge student sample in Italy	311
Table B.11.1:	Allocation of core student sample in Lithuania	312
Table B.11.2:	Allocation of teacher sample in Lithuania	312
Table B.11.3:	Allocation of bridge student sample in Lithuania	312
Table B.12.1:	Allocation of core student sample in Latvia	313
Table B.12.2:	Allocation of teacher sample in Latvia	313
Table B.12.3:	Allocation of bridge student sample in Latvia	314
Table B.13.1:	Allocation of core student sample in Malta	314
Table B.13.2:	Allocation of teacher sample in Malta	314
Table B.13.3:	Allocation of bridge student sample in Malta	315
Table B.14.1:	Allocation of core student sample in the Netherlands	315
Table B.14.2:	Allocation of teacher sample in the Netherlands	315
Table B.15.1:	Allocation of core student sample in Norway (9)	316
Table B.15.2:	Allocation of teacher sample in Norway (9)	316
Table B.15.3:	Allocation of bridge student sample in Norway (9)	316
Table B.16.1:	Allocation of core student sample in Poland	317
Table B.16.2:	Allocation of teacher sample in Poland	317
Table B.17.1:	Allocation of core student sample in Romania	318
Table B.17.2:	Allocation of teacher sample in Romania	318
Table B.18.1:	Allocation of core student sample in Serbia	319
Table B.18.2:	Allocation of teacher sample in Serbia	319
Table B.19.1:	Allocation of core student sample in Slovak Republic	320
Table B.19.2:	Allocation of teacher sample in Slovak Republic	320
Table B.20.1:	Allocation of core student sample in Slovenia	321
Table B.20.2:	Allocation of teacher sample in Slovenia	321
Table B.20.3:	Allocation of bridge student sample in Slovenia	321
Table B.21.1:	Allocation of core student sample in Spain	322
Table B.21.2:	Allocation of teacher sample in Spain	322
Table B.22.1:	Allocation of core student sample in Sweden	323
Table B.22.2:	Allocation of teacher sample in Sweden	323
Table B.22.3:	Allocation of bridge student sample in Sweden	323
Table B.23.1:	Allocation of core student sample in North Rhine-Westphalia	324
Table B.23.2:	Allocation of teacher sample in North Rhine-Westphalia	325
Table B.24.1:	Allocation of core student sample in Schleswig-Holstein	326
Table B.24.2:	Allocation of teacher sample in Schleswig-Holstein	326
Table C.1:	Civic knowledge items used in ICCS 2022	327
Table D.1:	Mapping of student questionnaire items in ICCS 2022 and ICCS 2016	331
Table D.2:	Mapping of teacher questionnaire items in ICCS 2022 and ICCS 2016	333
Table D.3:	Mapping of school questionnaire items in ICCS 2022 and ICCS 2016	334
Table D.4:	Mapping of European regional student questionnaire items in ICCS 2022 and ICCS 2016	335
Table D.5:	Mapping of Latin American regional student questionnaire items in ICCS 2022 and ICCS 2016	335
Table E.1:	List of international and regional student level variables with their respective coding used to generate plausible values of ICCS civic knowledge	336
Table F.1:	National average scale scores for student questionnaire scales across different delivery modes	345
Table F.2:	National average scale scores for student questionnaire scales across different delivery modes	346
Table F.3:	National average scale scores for student questionnaire scales across different delivery modes	347
Table F.4:	National average scale scores for student questionnaire scales across different delivery modes	348
Table F.5:	National average scale scores for student questionnaire scales across different delivery modes	349
Table F.6:	National average scale scores for student questionnaire scales across different delivery modes	350

Figures

Figure 2.1:	Example decontextualized test item	27
Figure 2.2:	Example unit comprising stimulus material and corresponding test items	27
Figure 3.1:	Contexts for the development of learning outcomes related to civic and citizenship education in ICCS 2022	44
Figure 5.1:	Instrument preparation workflow for the ICCS 2022 Main Survey	62
Figure 6.1:	Systematic selection of schools with probability proportional to their size	80
Figure 7.1:	Procedures for working with schools to prepare for test administration	89
Figure 9.1:	Overview of the data cleaning process	122
Figure 10.1:	Categories into which countries were placed with respect to student sampling participation rates	141
Figure 10.2:	Categories into which countries were placed with respect to teacher sampling participation rates	142
Figure 11.1:	Mapping of student abilities and item thresholds	145
Figure 11.2:	Item characteristic curve by category for item C14SLM2	152
Figure 11.3:	Example of item-by-country interaction graph for Item C14SLM2	158
Figure 11.4:	Example of item statistics provided to national centers	160
Figure 11.5:	Scatterplot of item parameters for CBA and PBA bridging study items	164
Figure 11.6:	Scatterplot for link item parameter estimates from ICCS 2022 and ICCS 2016	166
Figure 11.7:	Proficiency level cut-points and percentage of students at each level	170
Figure 12.1:	Summed category probabilities for fictitious item	184
Figure 12.2:	Example of questionnaire item-by-score map	185
Figure 12.3:	Confirmatory factor analysis of items measuring students' participation in out-of-school activities	187
Figure 12.4:	Confirmatory factor analysis of items measuring students' participation in civic activities in the community and at school	190
Figure 12.5:	Confirmatory factor analysis of items measuring students' perceptions of open classroom climates and civic learning	193
Figure 12.6:	Confirmatory factor analysis of items measuring students' perceptions of school interactions	196
Figure 12.7:	Confirmatory factor analysis of items measuring students' perceptions of political systems and institutions	198
Figure 12.8:	Confirmatory factor analysis of items measuring students' beliefs about threats to democracy and restrictions in national emergencies	200
Figure 12.9:	Confirmatory factor analysis of items measuring students' perceptions of the importance of good citizenship behavior	203
Figure 12.10:	Confirmatory factor analysis of items measuring students' participation in out-of-school activities	206
Figure 12.11:	Confirmatory factor analysis of items measuring students' endorsement of equal rights for all ethnic groups in society	208
Figure 12.12:	Confirmatory factor analysis of items measuring students' support for environmental protection and their environmental concerns	210
Figure 12.13:	Confirmatory factor analysis of items measuring students' citizenship self-efficacy and willingness to participate at school	212
Figure 12.14:	Confirmatory factor analysis of items measuring students' expected participation in activities to express their opinions	214
Figure 12.15:	Confirmatory factor analysis of items measuring students' expected political participation	216
Figure 12.16:	Confirmatory factor analysis of items measuring students' positive attitudes toward religious influence in society	218

Figure 12.17:	Confirmatory factor analysis of items measuring students' European identity and perceptions of learning about Europe at school	220
Figure 12.18:	Confirmatory factor analysis of items measuring students' attitudes toward freedom of movement in Europe	222
Figure 12.19:	Confirmatory factor analysis of items measuring students' support for European cooperation	224
Figure 12.20:	Confirmatory factor analysis of items measuring students' expectations for Europe's future	226
Figure 12.21:	Confirmatory factor analysis of items measuring students' expectations for their own future	228
Figure 12.22:	Confirmatory factor analysis of items measuring students' reports on ethical consumerism and sustainable behaviors	230
Figure 12.23:	Confirmatory factor analysis of items measuring students' attitudes toward the EU	232
Figure 12.24:	Confirmatory factor analysis of items measuring teachers' perceptions of teacher participation at school	234
Figure 12.25:	Confirmatory factor analysis of items measuring teachers' perceptions of social problems at school	236
Figure 12.26:	Confirmatory factor analysis of items measuring teachers' reports of student activities in the community	238
Figure 12.27:	Confirmatory factor analysis of items measuring teachers' perceptions of classroom climate	240
Figure 12.28:	Confirmatory factor analysis of items measuring teachers' reports on student involvement in activities	242
Figure 12.29:	Confirmatory factor analysis of items measuring teachers' reports on activities dealing with diversity	244
Figure 12.30:	Confirmatory factor analysis of items measuring teachers' opinions on the influence of cultural and ethnic diversity on teaching activities	246
Figure 12.31:	Confirmatory factor analysis of items measuring teachers' opinions on the influence of socioeconomic differences on teaching activities	248
Figure 12.32:	Confirmatory factor analysis of items measuring teachers' reports on activities related to digital technologies	250
Figure 12.33:	Confirmatory factor analysis of items measuring teachers' perceptions of the importance of good citizenship	252
Figure 12.34:	Confirmatory factor analysis of items measuring teachers' reports on civic-related activities in class	254
Figure 12.35:	Confirmatory factor analysis of items measuring teachers' reports on activities related to global issues	256
Figure 12.36:	Confirmatory factor analysis of items measuring teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics	258
Figure 12.37:	Confirmatory factor analysis of items measuring teachers' reports on their PD activities	260
Figure 12.38:	Confirmatory factor analysis of items measuring school principals' perceptions of teacher participation at school	262
Figure 12.39:	Confirmatory factor analysis of items school principals' perceptions of student and parent involvement at school	264
Figure 12.40:	Confirmatory factor analysis of items school principals' perceptions of school activities in the community	266
Figure 12.41:	Confirmatory factor analysis of items measuring school principals' reports on environment-friendly practices at school	268
Figure 12.42:	Confirmatory factor analysis of items measuring school principals' reports on training activities related to the use of digital technologies	270
Figure 12.43:	Confirmatory factor analysis of items measuring school principals' reports on the availability of resources in local community	272
Figure 12.44:	Confirmatory factor analysis of items school principals' perceptions of social tensions in the community	274

Chapter 1:

Overview of ICCS 2022

John Ainley and Wolfram Schulz

1.1 Background

The International Association for the Evaluation of Educational Achievement (IEA) has conducted cross-national surveys of civic and citizenship education for over fifty years (Malak-Minkiewicz & Torney-Purta, 2021). As an extension of its previous research on this topic, it established the International Civic and Citizenship Education Study (ICCS) with a first data collection in 2009 (Schulz et al., 2010). ICCS 2009 was built on previous IEA research, especially the IEA Civic Education Study of 1999 known as CIVED (Torney-Purta et al., 2001). ICCS 2009 expanded the assessment of civic knowledge to 80 items (there had been 38 items in CIVED) by using a rotated block design that provided a wider coverage of the field and included a cluster of items that had been used in CIVED (Schulz, et al., 2011). This was followed by a second implementation in 2016 (Schulz et al., 2018) and now a third implementation in 2022 (Schulz et al., 2024).

Each cycle of ICCS collects and analyses data on students' civic knowledge and understanding, as well as their attitudes, perceptions, and activities related to civics and citizenship both to measure changes over time and to respond to new developments of relevance for civic and citizenship education. The International Civic and Citizenship Education Study (ICCS) 2022 focusses on the ways in which young people are prepared to undertake their roles as citizens in a range of countries in the third decade of the 21st century. ICCS also gathers and analyzes information on education systems as well as schools, teachers, and students. Based on nationally representative samples of students, it also examined differences among countries in relation to these outcomes of civic and citizenship education, changes in these outcomes over time, and the relationship of these outcomes to student characteristics, school, and community contexts, as well as national characteristics.

ICCS 2022 is a continuation, and an extension, of previous cycles ICCS. It is 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 2022 extends the scope of previous assessments but is designed to maintain continuity with ICCS 2009 and ICCS 2016. It is designed to enable the measurement of enduring aspects of civic and citizenship educational contexts, processes, and outcomes and supporting the comparison of outcomes and contexts across 2009, 2016, and 2022. Key variables are statistically linked to enable changes to be investigated. For example, ICCS 2022, as was the case for ICCS 2016, uses a rotated-block assessment design that enables the inclusion of secure assessment material from previous cycles to provide the basis for analyses of trends as well as the inclusion of new assessment material to provide relevance to contemporary curricular developments.

1.2 New developments in ICCS 2022

Although ICCS 2022 instruments included many measures from ICCS 2009 or ICCS 2016 to collect data about changes over time, the survey also set out to address topics that had recently been included or become more prominent in civic and citizenship education, or considerably broaden the measurement of aspects that had already been included in previous cycles. To this end, through extensive discussions and deliberations with experts and national study centers we identified five focus areas for ICCS 2022 for inclusion in the assessment framework (see Schulz et al., 2023). Each focus area was developed in response to recent developments that were viewed as relevant to the area of civic and citizenship education. The following focus areas were identified to broaden the scope of ICCS for this cycle:

- *Sustainability*: Education for sustainable development (ESD), which is frequently considered in conjunction with Global Citizenship Education (GCED), had received increased attention as an educational area of interest in view of many demographic, environmental (including the potential impact of climate change), economic, and social challenges, aspects of which have been increasingly included in national curricula (Benavot, 2014).

- *Engagement through digital technologies*: Over the last two decades, there has been an enormous increase amongst young people in their daily interactions via digital communication and social media. Digital technologies for engagement have further opened new possibilities for mobilization, organization, and interaction of wider audiences and of young people (Brennan, 2018; Cho, 2020).
- *Diversity*: As result of recent economic, demographic, and technological changes, increasing levels of migration with growing diversity have been observed in many countries (Sandoval- Hernández et al., 2018; United Nations, Department of Economic and Social Affairs, Population Division, 2017) with implications for schools and other educational institutions (Malak-Minkiewicz & Torney-Purta, 2021) including the building of more multicultural and inclusive schools (Banks, 2020).
- *Young people's views of the political system*: Recent years have witnessed growing concerns about a global "democratic recession" (Diamond, 2021) connected to a surge in authoritarian government practices as well as electoral success of extreme political movements threatening the stability of established democratic systems (Boogards, 2017) in connection with growing alienation also among young voters (Estellés & Castellví, 2020). The recent COVID-19 pandemic has also highlighted challenges to forms of democratic governments in view of restrictions of personal freedom and divisions in society about these responses (Marzocchi, 2020).
- *Global Citizenship* was identified as an overarching area that deserved more explicit recognition regarding aspects already present in ICCS or in relation to the other focus areas. GCED has received considerable attention in recent debates about global education targets giving growing cross-national interconnectedness and globalization of political, social, economic, and environmental issues (Veugelers, 2011), but there have been inconsistent definitions across the literature (Pashby et al., 2020).

One major change in ICCS 2022 was the introduction of a computer-based delivery for student instruments, as an option for countries. Eighteen participants opted for this new mode while six countries continued delivering ICCS student test and questionnaire material in a print-based format. One important aspect of introducing computer-based delivery was the development of three clusters with test item material that made use of interactive possibilities for measurement and were only available in a digital mode.

It was necessary to provide a basis for investigating and adjusting mode effects through print-based data collections in additional representative samples. These bridging studies were conducted in all 11 countries that had opted for computer-based delivery and had also participated in ICCS 2016. The additional samples provided invaluable insights in potential mode differences for student outcomes reported in this cycle as well as a way of evaluating whether a mode effect adjustment was needed for civic knowledge test scores (see Chapter 11).

1.3 Research Questions

ICCS 2022 addressed five research questions which are elaborated more fully in the assessment framework (see Schulz et al., 2023):

1. *How is civic and citizenship education implemented in participating countries?* This includes the aims and principles of civic and citizenship education, curricular approaches, changes, and developments since the 2009 and 2016 cycles, and perceptions the roles of civic and citizenship education.
2. *What is the extent and variation of students' civic knowledge within and across participating countries?* This includes variations in civic knowledge associated with student characteristics and background variables, contextual factors explaining variation in students' civic knowledge, and changes in civic knowledge since the previous survey cycles.
3. *What is the extent of students' engagement in different spheres of society and which factors within or across countries are related to it?* This research question is related to students' beliefs about their capacity for civic participation, the extent and variation of students' civic participation in and out of school, students' expectations regarding future civic and political participation, and changes in the extent and forms of student engagement since the previous ICCS cycles.
4. *What beliefs do students in participating countries hold regarding important civic issues in modern society and what are the factors influencing their variation?* This research question concerns students' beliefs about the importance of principles underlying a democratic society, students' attitudes toward civic

institutions and society, students' perceptions of social cohesion and diversity in the societies in which they live, and changes in student beliefs observed since previous ICCS cycles.

5. *How is schooling in participating countries organized with regard to civic and citizenship education and what is its association with students' learning outcomes?* This research question is related to the extent to which schools have participatory processes that facilitate civic engagement, the extent to which schools and communities interact to foster students' civic engagement and learning, and the extent to which schools offer programs or activities related to civic learning.

1.4 Assessment Framework

The assessment framework for ICCS 2022 (Schulz, et al., 2023) provided a conceptual underpinning for the international instrumentation. It identified and defined those aspects of cognitive and affective-behavioral content that were considered important learning outcomes of civic and citizenship education, as well as contextual factors for students' civic learning. It should be noted that the way students develop civic knowledge and understanding, as well as affective-behavioral dispositions towards civic and citizenship issues, depends on many factors, including those beyond the learning environment at schools (see Amnå et al., 2009; Neundorff et al., 2016; Pancer, 2015).

The ICCS 2022 assessment framework includes a civic knowledge framework that describes the aspects of students' civic knowledge in terms of the content and the associated cognitive processes that are measured in the student test items. It distinguishes four content domains related to knowledge and understanding developed through civic and citizenship education:

- I. *Civic institutions and systems* (three sub-domains): (i) State institutions, (ii) economic systems, and (iii) civic society.
- II. *Civic principles* (five sub-domains): (i) Equity, (ii) freedom, (iii) rule of law, (iv) sustainability, and (v) solidarity.
- III. *Civic participation* (three sub-domains): (i) Decision-making, (ii) influencing, and (iii) community participation.
- IV. *Civic roles and identities* (three sub-domains): (i) Citizens, (ii) civic self-image, and (iii) civic connectedness.

The civic knowledge framework further distinguishes between processes involved in understanding complex sets of factors influencing civic actions and planning for and evaluating strategic solutions and outcomes. It differentiates two cognitive domains:

- I. *Knowing* refers to the learned civic and citizenship information that students use when engaging in the more complex cognitive tasks that help them make sense of their civic worlds.
- II. *Reasoning and applying* 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 and to make use of these in real-world contexts.

The ICCS 2022 assessment framework also includes a *civic attitudes and engagement framework* that outlines affective-behavioral constructs measured through student questionnaire items included in international and regional instruments (see a more detailed description in Chapter 3). These are regarded as important learning outcomes and have a similar standing in the process of development, analysis, and reporting as cognitive measures of students' civic knowledge. Student attitudes, perceptions, and behaviors relevant to civic and citizenship issues are organized according to the two affective-behavioral areas attitudes (e.g., judgements in relation to ideas, people, objects, events, or situations) and engagement (e.g., interest in, and expectations of, civic engagement through civic action and future political participation).

The ICCS 2022 assessment framework also includes a *contextual framework* which distinguishes four contextual levels (wider community, schools and classrooms, home and peer environments, individuals), at which factors influencing learning outcomes may be located (see a more detailed description in Chapter 3). Further, the ICCS 2022 contextual framework distinguishes between variables that are antecedents (pre-existing variables) and processes (variables related to civic learning).

1.5 Participating Countries, Study Design, and Implementation

Twenty-four education systems, including the national systems of 22 countries and the education systems of two states from another country, participated in ICCS 2022. Most participants were from Europe with just two Latin American countries and one Asian country being part of ICCS 2022. Various factors, including the consequences of the COVID-19 pandemic, impacted country participation that was less than the initial interest from a broader range of countries. ICCS 2022 gathered and analyzed data from approximately 82,000 students in their eighth year of schooling, 40,000 teachers of that grade and 3,400 school principals in these countries. In the nested sampling design used in ICCS 2022, the surveyed principals were the principals of the schools sampled for the student and teacher surveys.

Fourteen countries and one benchmarking participant that participated in ICCS 2022 had also participated in ICCS 2016, and 16 of the countries participated in both ICCS 2022 and ICCS 2009. Twelve countries participated in ICCS 2022, ICCS 2016, and ICCS 2009. Four of the ICCS 2022 countries had participated in ICCS 2009 but not in ICCS 2016. Four countries and one benchmarking participant participated for the first time in ICCS 2022.

Populations and samples

The student and teacher population definitions and sampling methods for ICCS 2022 were the same as those used in ICCS 2009 and 2016. The student population was 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.¹ 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.²

Stratified two-stage cluster samples were selected for the study. In the first stage of sampling, PPS procedures (probability proportional to size as measured by the number of students enrolled in a school) were used to sample schools within the participating countries or entities. In the second stage, random procedures were used to select intact classrooms for the student survey, and individual teachers of the target grade were selected at random for the teacher survey.

Instruments

Data for ICCS 2022 were collected using the following survey instruments:

- The *international student test* consisted of items measuring students' civic knowledge and ability to analyze and reason on topics related to civic and citizenship education. The instrument was either administered on paper or using the computer-based delivery platform. While computer-based tests included 14 item clusters (including three clusters with computer-enhanced item material), paper-based tests administered in countries that chose the paper mode included 11 item clusters (with items that were common across modes). Eight of the common item clusters were used to assess students who participated in the bridging study.
- The *international student questionnaire* consisted of items measuring student background variables, school contexts, and student perceptions and beliefs. The instrument was either administered on paper or using the computer-based delivery platform.
- *Regional student questionnaires* were optional for national centers and administered only in countries participating in European and Latin American data collections. The instruments were either administered on paper or using the computer-based delivery platform.
- The teacher questionnaire gathered information about teacher background variables and teachers' perceptions of factors related to the context of civic and citizenship education in their respective schools. It was administered to sampled teachers teaching any subject in the target grade, while an additional section was targeted at teachers teaching civic-related subjects (as defined by national centers). Participating national centers primarily delivered this instrument using online administration, with an option of offering paper-based delivery.

¹ If the average age of students in Grade 8 was below 13.5 years, Grade 9 became the target population.

² The teacher population included only those teachers who were teaching the target grade during the testing period and who had been employed at school since the beginning of the school year.

- The *school questionnaire* captured school characteristics and school-level variables related to civic and citizenship participation as provided by principals (or designates) at sampled schools. As with the teacher questionnaire, the school questionnaire was completed primarily online by respondents, with an option by the national center to administer using paper-based delivery.
- The *national contexts survey* was completed online by experts coordinated by the national centers. It was designed to gather data about the structure of the education systems, the status of civic and citizenship education in the national curricula and policy documents, and recent developments. Data obtained from this survey supplement published information sources about countries and their education systems to assist with the interpretation of the results from the student, school, and teacher questionnaires, and in describing national contexts for civic and citizenship education.

Common items as a basis for measuring changes over time

Of the *student test* items used in ICCS 2022, almost half were secure items from ICCS 2016. The inclusion of these items meant that student achievement in ICCS 2022, 2016, and 2009 could be reported on the same scale and compared across cycles. The ICCS 2022 questionnaire instruments (for students, teachers, schools, and national centers) also included selected sets of questions from the corresponding ICCS 2016 instruments, thus allowing for comparisons across the cycles in these selected areas. Importantly, the ICCS 2022 international report presents only the changes for those countries where data collections met the technical standards associated with sampling, instrument preparation, field operations, scoring, and data management during both cycles. This stipulation means that the reporting of changes over time does not cover all 15 countries that participated in ICCS 2022 and ICCS 2016 or all 17 countries that participated in both ICCS 2022 and ICCS 2009. Consequently, the number of countries included in comparisons of data collected by the various questions and instruments vary.

Participation rates

Data collection started in late February 2022 in the Northern Hemisphere and ended with the data collection in Brazil in September 2022. While most countries met IEA sample participation requirements for the student survey, two countries (Brazil and Denmark) and the benchmarking participant Schleswig-Holstein (Germany) failed to meet these requirements. For the teacher survey, there was a much lower number of 13 countries that met IEA sample participation requirements. It needs to be acknowledged that participation rates for ICCS 2022 were also affected by the COVID-19 pandemic which caused disruptions across many education systems. In keeping with established IEA standards, results from ICCS 2022 distinguish national data from surveys meeting IEA sample participation requirements from those that did not. By reporting data from surveys not meeting participation requirements in a different section of the reporting tables, readers are advised to interpret the corresponding findings with due caution regarding the possible impact of sample participation bias. Further, comparisons across different cycles of ICCS are only presented for national data where sample participation requirements for all data points.

1.6 Report Structure

This report is structured to provide comprehensive technical information about the conceptual, methodological, and analytical aspects of ICCS 2022. Following this overview there is a series of chapters that provide detailed information about ICCS 2022 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. It needs to be acknowledged that, given the cyclical nature of a study like ICCS, in many chapters there are sections that are highly similar in content to descriptions of technical aspects included in the technical report from the previous ICCS cycles (see Schulz et al., 2011; Schulz et al., 2018). This is especially the case for descriptions of technical aspects that did not change from previous cycles.

Chapters 2, 3, 4, and 5 are concerned with the development, adaptation, and translation of instruments. Chapter 2 provides information about the development and properties of the ICCS 2022 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, and the national contexts survey. In addition, ICCS 2022 also surveyed students in the two Latin American countries using items that had been used previously in ICCS 2016. Chapter 4 describes

the development of the computer-based assessment systems including the adaptation of print-based items to computer-based formats and the development of the computer-enhanced item material that was only delivered by computer. Chapter 5 outlines the preparation, translation, and verification of the ICCS 2022 study instruments.

Chapters 6 and 10 are concerned with aspects of sampling as well as weighting before and after collection. Chapter 6 describes the sampling design, national variations, and implementation, while Chapter 10 documents the weighting and adjudication procedures that were used to ensure the results from ICCS 2022 represented the defined populations in each country.

Chapters 7, 8, and 9 are concerned with survey implementation. Chapter 7 describes the field operation procedures and the process of preparing data files at the country level. Chapter 8 documents the quality control protocols and observation procedures that were used during the data collection. Chapter 9 provides an account of data management and the iterative building of the international databases.

Chapters 11 and 12 document the psychometric and statistical analyses used in this study. Chapter 11 reports on the scaling procedures for the cognitive test and how responses to the items were used to construct scores on the civic knowledge scale. This includes the procedures for equating items from ICCS 2022 on the previously established civic knowledge scale and for adjusting mode effects. Chapter 12 describes the methods and procedures used to derive scales from sets of questionnaire items.

Chapter 13 describes how the study results were reported, including the types of statistical analyses that were used to produce the reporting tables.

Finally, the technical report also contains appendices including:

- Lists of the organizations and individuals involved in ICCS 2022 (Appendix A);
- Characteristics of national samples (Appendix B);
- Descriptions of civic knowledge test items and their allocations to proficiency levels, noting that many items will remain secure for use in future ICCS cycles (Appendix C);
- A mapping of questionnaire items common to ICCS 2022 and 2016 (Appendix D);
- Coding information for the items in the questionnaires and the test (Appendix E); and
- Information about questionnaire scale averages and score-point differences across delivery modes (paper- vs. computer-mode) in 11 countries (Appendix F).

References

- Amnå, E., Ekström, M., Kerr, M., & Stattin, H. (2009). Political socialization and human agency. The development of civic engagement from adolescence to adulthood. *Statsvetenskaplig Tidskrift*, 111(1), 27–40. https://www.researchgate.net/publication/270790092_Political_Socialization_and_Human_Agency_The_Development_of_Civic_Engagement_from_Adolescence_to_Adulthood
- Banks, J. A. (2020). *Diversity, transformative knowledge, and civic education*. Routledge. <https://www.routledge.com/Diversity-Transformative-Knowledge-and-Civic-Education-Selected-Essays/Banks/p/book/9780367863197>
- Benavot, A. (2014). *Education for Sustainable Development in Primary and Secondary Education*. <https://doi.org/10.13140/RG.2.1.1978.9283>
- Boogards, M. (2017). Lessons from Brexit and Trump: populism is what happens when political parties lose control. *Zeitschrift für Vergleichende Politikwissenschaft*, 11(4), 513–518. <https://doi.org/10.1007/s12286-017-0352-y>
- Brennan, G. (2018). How digital media reshapes political activism: mass protests, social mobilization, and civic engagement. *Geopolitics, History, and International Relations*, 10(2), 76–81. <https://www.jstor.org/stable/26802343>
- Cho, A. (2020). *Digital civic engagement by young people*. Office of Global Insight and Policy, UNICEF. <https://www.unicef.org/globalinsight/reports/digital-civic-engagement-young-people>
- Diamond, L. (2021). Democratic regression in comparative perspective: scope, methods, and causes. *Democratization*, 28(1), 22–42. <https://doi.org/10.1080/13510347.2020.1807517>
- Estellés, M., & Castellví, J. (2020). The educational implications of populism, emotions and digital hate speech: A dialogue with scholars from Canada, Chile, Spain, the UK, and the US. *Sustainability*, 12(15), 6034. <https://doi.org/10.3390/su12156034>
- Malak-Minkiewicz, B., & Torney-Purta, J. (Eds.). (2021). *Influences of the IEA civic and citizenship education studies: practice, policy, and research across countries and regions*. Springer. <https://doi.org/10.1007/978-3-030-71102-3>
- Marzocchi, O. (2020). *The impact of COVID-19 measures on democracy, the rule of law and fundamental rights in the EU* (European Parliament Briefing). European Parliament. <https://data.europa.eu/doi/10.2861/47207>
- Neundorff, A., Niemi, R., & Smets, K. (2016). The compensation effect of civic education on political engagement: How civics classes make up for missing parental socialization. *Political Behavior*, 38(4), 1–29. <https://doi.org/10.1007/s11109-016-9341-0>
- Pancer, S. M. (2015). *The psychology of citizenship and civic engagement*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199752126.001.0001>
- Pashby, K., da Costa, M., Stein, S., & Andreotti, V. (2020) A meta-review of typologies of global citizenship education. *Comparative Education*, 56(2), 144–164. <https://doi.org/10.1080/03050068.2020.1723352>
- Sandoval-Hernández, A., Isac, M. M., & Miranda, D. (Eds.) (2018). *Teaching Tolerance in a Globalized World*. (IEA Research for Education; Vol. 4). Springer. <https://doi.org/10.1007/978-3-319-78692-6>
- Sant, E. (2019). Democratic education: A theoretical review (2006–2017). *Review of Educational Research*, 89(5), 655–696. <https://doi.org/10.3102/0034654319862493>
- Schulz, W., Ainley, J., & Fraillon, J. (Eds.) (2011). *ICCS 2009 technical report*. 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*. International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., Valeria, D., & Friedman, T. (2024). *Education for citizenship in times of global challenge*. IEA International Civic and Citizenship Education Study 2022 international report. Springer.

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. Springer. <https://doi.org/10.1007/978-3-319-73963-2>

Schulz, W., Fraillon, J., Losito, B., Agrusti, G., Ainley, J., Damiani, V., & Friedman, T. (2023). *IEA International Civic and Citizenship Education Study 2022 assessment framework*. Springer. <https://doi.org/10.1007/978-3-031-20113-4>

Schulz, W., Carstens, R., Losito, B., & Fraillon, J. (Eds.). (2018). *ICCS 2016 Technical Report*. Amsterdam: IEA. International Association for the Evaluation of Educational Achievement (IEA).

Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. International Association for the Evaluation of Educational Achievement (IEA).

United Nations, Department of Economic and Social Affairs, Population Division (2017). *International Migration Report 2017: Highlights (ST/ESA/SER.A/404)*. United Nations. https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2017_Highlights.pdf

Veugelers, W. (2011b). A Humanist Perspective on Moral Development and Citizenship Education. Empowering Autonomy, Humanity and Democracy. In W. Veugelers (Ed.), *Education and Humanism Linking Autonomy and Humanity* (pp. 9–34). <https://www.sensepublishers.com/media/299-education-and-humanism.pdf>

Chapter 2:

ICCS civic knowledge test development

Julian Fraillon

2.1 Overview

The International Civic and Citizenship Education Study (ICCS) civic knowledge test was developed between November 2017 and April 2020. The ICCS 2022 civic knowledge 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 2022 field trial and main survey.

Table 2.1: Test development process timeline

Original dates	Revised dates/mode resulting from the impact of the COVID-19 pandemic	Milestone/Activity
November 2017	No change	Preliminary establishment of test development process and test specifications
March 2018	No change	NRC Meeting 1 (Hamburg) Discussion of test development process and test specifications and subsequent review/revision
April 2018 to March 2019	No change	Item development planning including test platform discussion and computer-enhanced item prototyping
April 2019	No change	PAC Meeting 1 (Berkeley) Discussion of draft framework and draft test items and subsequent review/revision
April to May 2019	No change	Planning for NRC item development workshop
May 2019	No change	NRC Meeting 2 (Rome) Discussion of draft framework, items, and item development workshop and subsequent review/revision
May to December 2019	No change	Open submission of item suggestions by NRCs and PAC members
May to April 2020	No change	Ongoing item development at the ISC
January to June 2020	No change	Field trial instrument preparation
March 2020	No change	Online review of test items by PAC and NRCs
March to April 2020	Reduced number of countries	Small scale pilot of test items and subsequent review/revision
April 2020	Changed to virtual meeting	PAC Meeting 2 (virtual) Discussion of proposed field trial instruments and subsequent review/revision
May 2020	Changed to virtual meeting	NRC Meeting 3 (virtual) Discussion of proposed field trial instruments and subsequent review/revision
May 2020	No change	Final online review of proposed field trial instruments by NRCs
May to July 2020	No change	Finalization of field trial instruments
July 2020	No change	Field trial instrument release
September 2020	Changed to virtual meeting	Field trial scorer training (virtual) subsequent review/finalization of scoring guides
October to December 2020	October 2020 to April 2021	Field trial data collection
January to April 2021	April to July 2021	Field trial analysis
April 2021	August/September 2021, changed to virtual meeting	PAC Meeting 2 (virtual) Review of field trial outcomes and proposals for final instruments
May 2021	September 2021	Confirmation of proposed main survey instrument based on field trial outcomes
May 2021	September 2021, changed to virtual meeting	NRC Meeting 4 (virtual) Review of field trial outcomes and final instruments
May to June 2021	September to October 2021	Finalization of main survey instruments
July 2021	October 2021	Main survey instrument release
July 2021	February 2022, changed to virtual meeting	Main survey scorer training (virtual) with subsequent review/finalization of scoring guides

2.2 Test scope and format

ICCS Assessment framework

The cognitive test items were developed with reference to the ICCS 2022 assessment framework (Schulz et al., 2023) and designed to measure a single trait labelled civic knowledge in the international report on ICCS 2022 (Schulz et al., 2024). The way civic knowledge was expressed through the ICCS test items required students to apply the cognitive processes to the civic and citizenship content as described in the assessment framework.

Each test item developed for ICCS 2022 was mapped to both the civics and citizenship content and cognitive process required by students to respond correctly to the item.

The ICCS 2022 assessment framework includes four content and two cognitive domains. The four content domains are:

- Civic institutions and systems;
- Civic principles;
- Civic participation;
- Civic roles and identities.

The two cognitive domains are:

- Knowing;
- Reasoning and applying.

In ICCS 2022, the increased focus on global citizenship education (GCED) and education for sustainable development (ESD) required an increase in the number of items relative to previous cycles. Some new test items were developed specifically to address these two areas of increased focus. The degree to which all test items (both newly developed and secure trend items) represented content associated with each of GCED and ESD was also evaluated as part of the test development process.

Test item formats

In the two previous ICCS cycles, all students completed the tests on paper. Approximately 90 percent of items were four option multiple-choice (MC), with students required to select one option, and the remaining 10 percent of items were short constructed response (CR), with students expected to respond by writing between a few words and a few sentences.

ICCS 2022 initiated the transition in ICCS from paper-based to computer-based assessment. Countries could choose whether all students in the country would complete the tests on computer or on paper. The introduction of computer-based test delivery to ICCS afforded the opportunity to broaden the set of item formats available to students. However, it was also necessary to ensure that student test data collected on computer in ICCS 2022 were comparable to those data collected on paper in ICCS 2022, and in previous cycles of ICCS.

The ICCS 2022 test instrument included a set of items that were common to the paper-based and computer-based delivery modes. This included the secure trend items from previous ICCS cycles and newly developed items. All these items were either MC or CR as used in previous ICCS cycles. Hereafter in this chapter, MC and CR items are collectively referred to as conventional items.

The conventional items that were common to paper-based and computer-based delivery modes included a small number of decontextualized questions of knowledge or understanding (Figure 2.1¹) although most of the items were presented in units. Each unit typically included some brief contextual stimulus (an image or some text) followed by items relating to the context established by that stimulus (Figure 2.2²).

¹ The test item shown in Figure 2.1 is presented as sample item 4 in the ICCS 2022 international report.

² The test items shown in Figure 2.2 are presented as sample items 5 and 9 in the ICCS 2022 international report.

Figure 2.1: Example decontextualized test item

What is the most likely reason a person in a democracy would join a political party?

- ☐ to help the police with the enforcement of laws
- ☐ to decide on sentences for people found guilty of crimes
- ☐ to decide on the location of polling booths for national elections
- ☐ to participate in activities that influence government policy

Figure 2.2: Example unit comprising stimulus material and corresponding test items

Fake news is a term used to describe false or misleading information that is deliberately presented as news.

How can the presence of fake news harm a democracy?

Give **two ways**.

1. _____

2. _____

What can a person do to confirm whether information in a news article is real or fake?

- ☐ check whether or not the same information is presented by many different sources
- ☐ check whether or not the government agrees or disagrees with the news
- ☐ check whether or not their friends believe the news article
- ☐ check whether or not the news is [trending] on social media

The test instrument also included a set of items developed for use on computer only. This included a range of computer-enhanced items, such as drag and drop items and sliders that dynamically adjusted the relative proportions of allocations in a fictitious budget.

The items presented in the computer-based delivery mode only were presented in computer-enhanced clusters. These were fixed sequences of items (some computer-enhanced and some MC or CR), each with a form of narrative structure providing a thematic link across the items within that cluster. In total there were 20 items distributed across three computer-enhanced clusters in the ICCS 2022 main survey computer-based test instrument.

2.3 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 number of items planned for use in the main survey, incorporating the available secure trend items from previous cycles of ICCS, and with a view to increasing the emphasis on GCED and ESD in the ICCS test.

Under the planned test design for ICCS 2022, and consistent with previous cycles of ICCS, each test item was planned to be allocated to one *cluster* of items. Clusters comprising conventional items only were planned to include 11±1 unique items and were anticipated to take students not longer than 15 minutes to complete. The computer-enhanced clusters were also planned to take students not longer than 15 minutes to complete, but the number of items was planned to be less than in the clusters comprising conventional items only. This is because it was anticipated that some of the computer-enhanced items may take longer to complete than the conventional test items.

Each *test form* was planned to comprise three clusters. These were planned to be allocated using a partially balanced incomplete block design. Details of these allocations are included in the sections relating to the field trial and main survey test designs later in this chapter.

The early phase commitment to developing the computer-enhanced test items was to create themes and initial item plans for between six and eight computer-based clusters for consideration by stakeholders, with a view to reducing this set to five to be instantiated in the online test environment for inclusion in the field trial and main survey.

Table 2.2: ICCS 2022 test development specifications

	Pilot study	Field trial	Main survey
Items to be used in both the computer-based assessment (CBA) and paper-based assessment (PBA)			
Number of test forms	4+ (not rotated)	14	12-13
Trend items (civic knowledge)		22	44-55 (4-5 clusters)
New MC civic knowledge items	70-80	60-70	40-50
New CR civic knowledge items	7-8	6-7	4-5
New MCQ GCED/ESD items	60-70	40-50	30
New CR GCED/ESD items	6-7	4-5	3
Items to be used in CBA only			
New computer-enhanced clusters	4-5	4-5	3-4
New items within computer-enhanced clusters	25-35	25-35	15-20
Total new items			
MC	130-150	100-120	70-80
CR	14-15	11-12	7-8
Computer-enhanced	25-35	25-35	15-20

By design, the ICCS test instrument planning assumed that the ICCS 2022 test would measure similar civic and citizenship education content and cognitive processes to those measured in previous cycles. However, the ICCS 2022 test was planned to include a higher proportion of GCED/ESD items than in previous cycles, and it was accepted that the inclusion of the computer-enhanced clusters may also affect the relative proportions of ICCS test content mapped to the different content domains by delivery mode (CBA versus PBA).

The test specifications plan was presented for consideration at the first meeting of ICCS NRCs in March 2018. It was further reviewed by the ICCS PAC and finalized following the second meeting of NRCs in May 2019.

NRC item development workshop

An item development workshop was conducted as part of the second meeting of NRCs in May 2019 in Rome. At this workshop, national representatives were provided with information about the framework and procedures for ICCS 2022 test development, as well as the test specifications. Participants drafted ideas 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 civic knowledge construct;
- 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 linguistic, cultural, or gender bias);
- Confirmation of item formats, scoring guide formats, and test development systems, including the expected features of the online test authoring and delivery system;
- An overview of the plans for computer-enhanced test clusters, including review of an example cluster with a discussion of the principles underpinning the development of the clusters;
- An introduction to the general principles underpinning all ICCS cognitive test-item development; and
- Test-item development in small groups.

At the workshop, NRCs were also provided with an open invitation to submit materials to support the development of ICCS 2022 test items. These could include proposed stimulus materials, topics, test items, or test item ideas.

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 testing 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 test materials. This rigorous quality-control mechanism is employed during the development of test 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 have 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 2022. 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 2022 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, context, and translatability

- 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?
- Were there any characteristics of the item format that created hindrances to test-takers' capacity to access the core content of the item, and if so, how could they be corrected?
- Did the content of the computer-enhanced clusters make use of the computer-based delivery environment in a way that extended beyond what could be delivered on paper?

Test-takers

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

Scoring

- Was the key (the correct answer to a multiple-choice question) irrefutably correct?
- Were the distractors (the incorrect options to a multiple-choice question) plausible but also irrefutably incorrect?
- 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 scorers 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 based on 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 2022 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 March 2020, after which the draft test material was further revised in accordance with the feedback, and in preparation for the small-scale piloting of ICCS test material.

Pilot testing

The original plan for ICCS 2022 was for pilot testing to be conducted in all participating countries. The impact of the COVID-19 pandemic meant that most countries were not able to participate in the pilot testing. Small-scale pilot testing with convenience samples of students was conducted in five countries (Chinese Taipei, Colombia, Norway, Spain, and Sweden). Six non-overlapping test forms were created with a total of 155 newly developed items representing parallel content that could be delivered on computer and on paper.

Because of the reduced number of countries participating in the pilot, and the challenges within those countries to find schools available to participate, the computer-enhanced cluster content was not included in the pilot instruments. The focus of the piloting was narrowed to collect feedback on the large set of items that potentially were to be presented to all students in ICCS, as well as to collect information on student experiences of the computer-based delivery in comparison to the paper-based delivery of these common 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 sufficiently high numbers of students to support a robust scaling analysis, but response frequencies for multiple-choice options and proportions of correct responses for all items were checked as part of the review. We also conducted Rasch IRT analyses of the pooled (from all countries) multiple-choice data for each of the six test forms. 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 in preparation for further review by the ICCS PAC and ICCS NRCs.

Final item review and revision – post pilot

A final proposed item set, that incorporated feedback and information collected in the external item review and the pilot was presented to ICCS NRCs for discussion at the third NRC meeting, conducted remotely, in May 2020. Following the meeting, the full set of proposed items and scoring guides were made available to NRCs to conduct a final online review in May. Feedback from this final review was used to complete a final revision of the items and confirmation of the test instrument for the ICCS field trial.

2.4 Field trial test design and content

Overview

In ICCS 2022, the field trial test needed to accommodate countries completing computer-based testing and paper-based testing.

The test instrument included a pool of 195 items that can be described as belonging to two categories. The majority of items (154) were conventional items, developed in order to be administered in parallel in the paper-based and computer-based ICCS 2022 tests. These were items in which the stimulus, content and

formats were identical in the paper-based and computer-based tests. Twenty-two secure trend items from ICCS 2016 were included in this set of 154 items.

The remaining 41 items were developed to be used within five computer-enhanced clusters. Eighteen of these items used formats that could not be delivered on paper. Adapted (but not identical) versions of 11 of the remaining 23 items were also included in the paper-based test forms. These 11 items had formats that were suitable for delivery on paper and did not require the broader contexts provided by the computer-enhanced clusters to be accessible to students. The remaining 12 items had formats that might also have been delivered on paper but could not be included in the paper-based tests without extensive adaptation to compensate for the lack of context provided by the computer-enhanced clusters. Therefore, these items were delivered only on computer.

Table 2.3 shows a summary of the item types in the ICCS 2022 field trial instrument. The computer-enhanced clusters included three drag and drop items and 15 dynamic interactive items. These are items with a range of formats that shared the common attribute of being associated with some form of dynamic interactive feedback to students within the test.

Table 2.3: Composition of field trial test instrument by item format and origin

Item format	New items (n)	Trend items (n)	Total (n)
MC	122	19	141
CR	33	3	36
Drag and drop	3		3
Dynamic interactive	15		15
Total	173	22	195

Coverage of the assessment framework

The target mapping of test content to the content and cognitive domains in the ICCS assessment frameworks for ICCS 2009 and 2016 was as follows:

Content domains

- Domain 1 (civic society and systems): 40%
- Domain 2 (civic principles): 30%
- Domain 3 (civic participation): 20%
- Domain 4 (civic identities): 10%

Cognitive domains

- Domain 1 (knowing): 30%
- Domain 2 (reasoning and applying): 70%

The ICCS 2022 assessment framework was still under review at the time the test development was completed. Therefore, during the ICCS test instrument development phase, it was not confirmed whether the changes to the assessment framework and the introduction of the newly developed computer-enhanced clusters would require changes in the proportions of content by domain in the ICCS test instrument.

Our expectation was that data collected using the ICCS 2022 test would be equated to and reported on the single ICCS civic knowledge scale. Some variation in the proportions of content from the four content domains was not anticipated to compromise this expectation, given that content from the four domains has historically been reported on the single achievement scale. However, we left open the possibility that developing a second GCED/ESD scale may be feasible. As such, all test items were also classified according to whether they were consistent with GCED/ESD content.

The 195 items used in the ICCS 2022 field trial, and particularly the 173 new items, provided good overall coverage of the content and cognitive domains in the ICCS assessment framework. The intended profile of the main survey test instrument had not been confirmed at the time of the field trial. However, the field trial

instrument was developed to include sufficient content coverage across the framework (including items associated with either or both of GCED/ESD) to accommodate a range of relative proportions of content by domain for in the main survey instrument (Table 2.4).

Table 2.4: Field trial mapping to assessment framework

Domain type	Domain	New items (n)	Trend items (n)	Total items (n)	All items (%)	GCED (n)	ESD (n)
Content	1. Civic institutions and systems	40	5	45	23%	16	6
	2. Civic principles	47	8	55	28%	14	17
	3. Civic participation	69	5	74	38%	7	7
	4. Civic roles and identities	17	4	21	11%	11	5
	Total	173	22	195	100%	48	35
Cognitive	1. Knowing	49	8	57	29%	15	18
	2. Reasoning and applying	124	14	138	71%	33	17
	Total	173	22	195	100%	48	35

Test design

Allocation of item content to clusters

The items in the field trial test item pool were allocated to clusters. These were unique sets of items, anticipated to be completed by students in up to 15 minutes. Each item was allocated to one cluster only. Each test form comprised three clusters, and students were allocated a total of 45 minutes to complete the test.

Clusters of items that could be delivered either on paper or on computer typically included 11 conventional items. The computer-enhanced clusters included predetermined content based on their cluster themes and included between five and eight items. The computer-enhanced clusters were structurally equivalent to the item clusters in the test design.

The ICCS 2022 field trial included the following three testing scenarios:

- Computer-based delivery (used in all countries selecting the computer-based delivery option);
- Paper-based delivery (used in all countries selecting the paper-based delivery option);
- Mode-effect study (paper-based delivery to an additional sample of students in selected countries participating in the computer-based delivery option).

The test items were allocated to clusters in a test design comprising 14 clusters across 14 test forms in each of the paper-based and computer-based delivery testing scenarios. Each of clusters 3, 5, 7, 9, and 13 had alternative versions. These comprised the computer-enhanced clusters (denoted as _V(C) in Table 2.5) and the alternative paper-based clusters (denoted as V_(P) in Table 2.5). The test item content was different between each of the computer-enhanced clusters and their alternative paper-based clusters. The alternative clusters were included to maximize the item content that could be included in the field trial and to create a common design across all participating countries regardless of delivery mode.

The content of the computer-enhanced clusters was established by the item content associated with the narrative theme of each cluster. The conventional item clusters were established so that the clusters were equivalent with respect to: item format (MCQ and CR), reading load (in English), CCE content and cognitive processes (as specified by the ICCS 2022 assessment framework). Further checks were completed to ensure that there was no interference of item content across items within test forms (each form comprised three clusters). Where any interference was noted as possible (such as where two items addressing similar content were in a test form) the cluster content was adjusted accordingly. Two clusters (C1 and C11) were constructed

using secure trend items only. The remaining clusters were constructed using newly developed items only.

Table 2.5: Field trial cluster composition

Cluster	Item origin	Items (n)	Item type	Delivery mode
C1	Trend	11	Conventional	Computer and paper
C2	New	11	Conventional	Computer and paper
C3_V(C)	New	7	Computer-enhanced	Computer only
C3_V(P)	New	11	Conventional	Paper only
C4	New	11	Conventional	Computer and paper
C5_V(C)	New	5	Computer-enhanced	Computer only
C5_V(P)	New	11	Conventional	Paper only
C6	New	11	Conventional	Computer and paper
C7_V(C)	New	5	Computer-enhanced	Computer only
C7_V(P)	New	11	Conventional	Paper only
C8	New	11	Conventional	Computer and paper
C9_V(C)	New	6	Computer-enhanced	Computer only
C9_V(P)	New	11	Conventional	Paper only
C10	New	11	Conventional	Computer and paper
C11	Trend	11	Conventional	Computer and paper
C12	New	11	Conventional	Computer and paper
C13_V(C)	New	5	Computer-enhanced	Computer only
C13_V(P)	New	11	Conventional	Paper only
C14	New	11	Conventional	Computer and paper

Allocation of clusters to test forms (computer-based and paper-based delivery)

The computer-based and paper-based delivery modes used the same test design comprising 14 test forms (Table 2.6). The content differed between the two modes only where the alternative clusters (shown as _V in Table 2.6) would be either a computer-enhanced cluster in the computer-based delivery mode or an alternative conventional item cluster in the paper-based mode.

ICCS 2022, as in previous cycles of ICCS, used a partially balanced incomplete block design. Each cluster (including the alternative paper or computer clusters) appeared once in each of the first, second, and third positions across the test forms relevant to their delivery mode. While each cluster was presented once in each position across the test forms (relevant to their delivery mode), each cluster did not appear an equal number of times with each other cluster.

Table 2.6: Field trial test form design (computer-based and paper-based delivery)

Test form	Position		
	1	2	3
1	C1	C2	C4
2	C2	C3_V	C5_V
3	C3_V	C4	C6
4	C4	C5_V	C7_V
5	C5_V	C6	C8
6	C6	C7_V	C9_V
7	C7_V	C8	C10
8	C8	C9_V	C11
9	C9_V	C10	C12
10	C10	C11	C13_V
11	C11	C12	C14
12	C12	C13_V	C1
13	C13_V	C14	C2
14	C14	C1	C3_V

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

Allocation of clusters to test forms (mode-effect study)

The mode-effect study was conducted in selected countries using the computer-based delivery mode. The purpose of the mode-effect study was to collect data using paper-based test forms that could then be used to provide information about the relative impact of the delivery mode (paper or computer) on estimates of students' civic knowledge.

The mode-effect study test forms were administered to an additional sample of students in participating countries. To minimize the additional burden of this work for countries, we used a reduced sample of field trial test items in the mode-effect study forms, as this would keep to a minimum the number of additional students needed for the study.

The mode-effect study forms were created using a subset of the conventional item clusters used in the paper-based delivery mode of the field trial (Table 2.7). Each cluster appeared once in each of the first, second, and third positions across the test forms.

Table 2.7: Field trial test form design (mode-effect study)

Test form	Position		
	1	2	3
1	C2	C11	C8
2	C4	C2	C1
3	C6	C4	C11
4	C1	C6	C2
5	C8	C1	C4
6	C11	C8	C6

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

Scoring guides for human scoring (field trial)

All CR items were scored by humans only. The scoring guides for the items are a critical component of the item, as interpretation of the guides by trained scorers generate the data for each item. Preliminary scoring guides were created and developed and reviewed for all CR items developed for ICCS 2022.

The scoring guides for the CR items selected for inclusion in the field trial were further reviewed and revised.

Draft detailed scoring guides were used as the basis for training human scorers in the field trial. This was a *train-the-trainer* model in which lead scorers from each country were trained to score the items according to the draft scoring guides. These lead scorers were then responsible for training scorers and supervising scoring operations in their own countries.

The scoring guide for each CR item consisted of:

- descriptions of the necessary components, content, and characteristics of the score codes;
- descriptions of the conceptual categories that can be considered within each level of credit;
- sample student responses that were indicative of each conceptual category for that item;
- some annotations to example responses to explain why certain responses were allocated to a specific category; and
- some additional notes to assist trainers and scorers to interpret the intention and content of aspects of the scoring guide.

It was essential that the scoring guide was used consistently across countries. It was emphasized that under no circumstances should the agreed scores be altered during the scorer training.

All sample responses selected for scorer training were thoroughly vetted, discussed, and cross-marked by the lead scorers.

The selection and discussion of sample responses was a pivotal part of scorer training. This exercise helped scorers become more familiar with the scoring criteria being applied to the responses and helped clarify any areas of misunderstanding or concern. Sample responses are important in helping scorers calibrate and refine their judgments and gain confidence in their ability to mark effectively.

In summary, for scorer training, sample responses were used to:

- teach scorers how to apply the scoring guide correctly;
- help scorers calibrate their marking in preparation for 'live' responses; and
- give scorers confidence in their judgments so that they can mark efficiently.

Sample responses were intended to give scorers information about the sorts of responses they can expect to encounter during the scoring operation. Consequently, it was expected that the sets of sample responses selected would showcase variety and breadth in:

- scores assigned;
- content;
- styles of response; and
- interpretations of items (and stimulus).

The draft scoring guides were reviewed and refined following the field trial scorer training meeting. This was done to remove any ambiguities or unclear information in the guides that was revealed in discussions at the training.

Lead scorers were encouraged to contact the ISC with any questions they had during the field trial scoring. Relevant questions and ISC responses were published online for the information of lead scorers across all countries.

2.5 Main survey test design and content

Selection of items for inclusion in the main survey

Selecting items for possible inclusion in the main survey test instrument

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 the items' 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.

In total, 25 items were deemed to be unsuitable for inclusion in the main survey instrument. These included one entire computer-enhanced cluster.

The remaining items were considered as viable for inclusion in the main survey instrument. These findings were further discussed with the ICCS PAC, and subsequently at the fourth meeting of NRCs.

Following these meetings and in response to further review comments, a final pool of items was established from which the content for the ICCS 2022 main survey test instrument could be selected. The final pool included three of the five computer-enhanced clusters that had been included in the field trial, and 137 new conventional items.

Selection of item content and allocation of item content to clusters

The initial ICCS 2022 test development specifications (Table 2.2) indicated that the main survey test instrument would comprise 12–13 clusters, including 3–4 computer enhanced clusters and up to 55 trend items. Following the review of the field trial items, we decided to include 55 secure trend items (equivalent to five 11-item clusters), the three available computer-enhanced clusters (comprising 20 items in total) and a further 66 newly developed test items (selected from the 137 available). This corresponded to a total of 11 clusters of conventional test items to be delivered across 11 test forms in countries completing the tests on paper. In countries completing the tests on computer, the three computer-enhanced clusters would be delivered together with the 11 clusters of conventional test items across 14 test forms.

The decision to include 55 trend items was made to maximize the availability of trend item content to support equating the ICCS 2022 test data to the ICCS civic knowledge scale. The decision to include 66 new conventional items was made to keep the overall test design as close as possible to that established in the test development specifications and planned for by countries. Having a pool of 137 items from which 66 were to be selected offered considerable flexibility when selecting items to ensure that the clusters were well balanced while still offering a broad range of content coverage.

The contents of the 11 conventional item clusters were selected using similar principles that applied for the field trial. The items were selected such that the clusters were balanced with respect to: item format (MCQ and CR), reading load (in English), and CCE content and cognitive processes (as specified by the ICCS 2022 assessment framework). Furthermore, the items were selected to ensure that a range of contexts was presented across the instrument. Unlike the field trial, the secure trend items were distributed relatively evenly across the 11 clusters rather being included as two discrete trend item clusters. Further checks were also completed to ensure that there was no interference of item content across items within test forms.

The contents of the three computer-enhanced clusters were predetermined within each cluster to ensure a good balance across cognitive and content dimensions.

Main survey instrument composition and framework mapping

The main survey test instrument consisted of 141 items. There were 121 conventional items (55 secure trend and 66 new) and 20 items from the computer-enhanced clusters.

The relative proportions of item content coverage by content sub-domains (Table 2.9) was similar to but varied slightly from those reported for ICCS 2016. In comparison with ICCS 2016, the ICCS 2022 test instrument contained a lower proportion of items associated with content domain 1 and a higher proportion of items associated with content domains 2 and 3. These differences can be attributed to:

1. The reallocation of selected content from content domain 1 to content domain 4 in the ICCS 2022 assessment framework;
2. The addition of the sustainability sub-domain under content domain 2, and a consequent increase in the number of items associated with content domain 2 as a result of the increased number of items addressing ESD in 2022; and
3. An increase in the number of items associated with content domain 3 in 2022 resulting from the inclusion of the three computer-based delivery only clusters with themes that focused on civic participation.

Across both previous ICCS cycles, civic knowledge, as described across the four content domains in the relevant assessment frameworks, was reported as a single measurement dimension. The slight variation in the distribution of test content across the four domains in ICCS 2022 was not anticipated to prevent ICCS 2022 test data from being equated to and reported on the single ICCS civic knowledge scale. However, in the ICCS 2022 test instrument, there were sufficient GCED and ESD items (Table 2.9) to evaluate empirically the justification or need to report GCED and or ESD as separate measurement dimensions. Based on evaluation of the dimensionality within the ICCS 2022 test data, the decision was made to report civic knowledge as a single scale in ICCS 2022 (see Chapter 11 for full details of these analyses).

Table 2.8: Composition of main survey test instrument by item format and origin

Item format	New items (n)	Trend items (n)	Total (n)
MC	62	49	111
CR	16	6	22
Drag and drop	2		2
Dynamic interactive	6		6
Total	86	55	141

Table 2.9: Main survey item mapping to assessment framework

Domain type	Domain	New items (n)	Trend items (n)	Total items (n)	All items (%)	GCED (n)	ESD (n)
Content	1. Civic institutions and systems	15	14	29	21%	5	5
	2. Civic principles	33	18	51	36%	13	18
	3. Civic participation	30	18	48	34%	1	4
	4. Civic roles and identities	8	5	13	9%	8	1
	Total	86	55	141	100%	27	28
Cognitive	1. Knowing	28	14	42	30%	9	14
	2. Reasoning and applying	58	41	99	70%	18	14
	Total	86	55	141	100%	27	28

Test design

Item clusters

The main survey test instrument comprised 11 conventional item clusters (121 items) and three computer-enhanced clusters (20 items).

For the purposes of creating and documenting the test design, the 11 conventional item clusters were labelled C1 to C11. The three computer-enhanced clusters were labelled C12, C13, and C14.

Table 2.10: Main survey cluster composition

Cluster	MC items (n)	CR items (n)	Dynamic interactive and drag and drop items (n)	Trend items (n)	New items (n)	Total items (n)
C1	10	1	N/A	5	6	11
C2	10	1	N/A	5	6	11
C3	10	1	N/A	5	6	11
C4	10	1	N/A	5	6	11
C5	10	1	N/A	5	6	11
C6	10	1	N/A	5	6	11
C7	10	1	N/A	5	6	11
C8	9	2	N/A	5	6	11
C9	9	2	N/A	5	6	11
C10	10	1	N/A	5	6	11
C11	10	1	N/A	5	6	11
C12	1	3	3	N/A	7	7
C13	1	3	1	N/A	5	5
C14	1	3	4	N/A	8	8

Allocation of clusters to test forms (computer-based delivery)

All 11 conventional item clusters and all three computer-enhanced clusters were included in the test design used for computer-based delivery (C1 to C14). These 14 clusters were allocated across 14 test forms with each cluster appearing once in each of the first, second and third positions across the test forms. Furthermore, no test form contained more than one computer-enhanced cluster (Table 2.11).

Table 2.11: Main survey test form design (computer-based delivery)

Test form	Position		
	1	2	3
1	C1	C12	C9
2	C2	C13	C10
3	C3	C14	C11
4	C4	C1	C12
5	C5	C2	C13
6	C6	C3	C14
7	C7	C4	C1
8	C8	C5	C2
9	C9	C6	C3
10	C10	C7	C4
11	C11	C8	C5
12	C12	C9	C6
13	C13	C10	C7
14	C14	C11	C8

Allocation of clusters to test forms (paper-based delivery)

All 11 conventional item clusters were included in the test design used for paper-based delivery (C1 to C11). These 11 clusters were allocated across 11 test forms with each cluster appearing once in each of the first, second and third positions across the test forms (Table 2.12).

Table 2.12: Main survey test form design (paper-based delivery)

Test form	Position		
	1	2	3
1	C1	C2	C6
2	C2	C3	C7
3	C3	C4	C8
4	C4	C5	C9
5	C5	C6	C10
6	C6	C7	C11
7	C7	C8	C1
8	C8	C9	C2
9	C9	C10	C3
10	C10	C11	C4
11	C11	C1	C5

Allocation of clusters to test forms (bridging study)

The bridging study was conducted in selected countries using the computer-based delivery mode. The purpose of the bridging study was to collect data using paper forms that could then be used to provide information about the relative impact of the delivery mode (paper or computer) on estimates of students' civic knowledge.

The bridging study test forms were administered to an additional sample of students in participating countries. To minimize the additional burden of this work for countries, we used a reduced sample of main survey test items in the bridging study forms, as this would keep to a minimum the number of additional students needed for the study.

The bridging study test forms were created using eight of the 11 main survey conventional item clusters (C1 to C8). These eight clusters were allocated across eight test forms with each cluster appearing once in each of the first, second, and third positions across the test forms (Table 2.13).

Table 2.13: Main survey test form design (bridging study)

Test form	Position		
	1	2	3
1	C1	C2	C4
2	C2	C3	C5
3	C3	C4	C6
4	C4	C5	C7
5	C5	C6	C8
6	C6	C7	C1
7	C7	C8	C2
8	C8	C1	C3

Scoring guides for human scoring (main survey)

The scoring guides for the CR items selected for inclusion in the main survey were further reviewed and revised.

The initial review and revision were completed with reference to any questions of clarification raised by national centers as they completed their field trial scoring. The example student responses in the guides were updated to include relevant and useful examples that were uncovered during the field trial.

These revised draft main survey guides were used as the basis for training human scorers for the main survey. As for the field trial, this was a *train-the-trainer* model in which lead scorers from each country were trained to score the items according to the draft scoring guides. These lead scorers were then responsible for training scorers and supervising scoring operations in their own countries.

The draft main survey scoring guides were reviewed and refined following the main survey scorer training meeting.

Lead scorers were encouraged to contact the ISC with any questions they had during the main survey scoring. Relevant questions and ISC responses were published online for the information of lead scorers across all countries.

2.6 Released test items

The ICCS 2022 international report includes 15 sample items taken from the ICCS 2022 main survey test instrument. Nine of these items are conventional item format, delivered both on paper and on computer. The remaining six items comprise one computer-enhanced test cluster, delivered on computer only. Full details of these items are available in the ICCS 2022 international report and the ICCS 2022 user guide. In the ICCS 2022 user guide, details of a further 13 items are provided. Full texts for the 13 items in the restricted use item set that are not sample items from the ICCS 2022 international report are available only on request from IEA.³

Table 2.14: Characteristics and mapping of ICCS 2022 released test items

	Characteristics	Sample items in ICCS 2022 IR	Additional restricted use items	Total items (n)
Item format	MC	9	12	21
	CR	4	1	5
	Drag and drop	1		1
	Dynamic interactive	1		1
	Total	15	13	28
Trend status	Trend	3	8	11
	New	12	5	17
	Total	15	13	28
GCED/ESD status	GCED/ESD	3	9	12
	Not GCED/ESD	12	4	16
	Total	15	13	28
Content domain	1. Civic institutions and systems	3	2	5
	2. Civic principles	2	9	11
	3. Civic participation	9		9
	4. Civic roles and identities	1	2	3
	Total	15	13	28
Cognitive domain	1. Knowing	6	6	12
	2. Reasoning and applying	9	7	16
	Total	15	13	28

³ Please complete and submit IEA's permission request form, which is available on the IEA website.

References

Schulz, W., Fraillon, J., Losito, B., Agrusti, G., Ainley, J., Damiani, V., & Friedman, T. (2023). *IEA International Civic and Citizenship Education Study 2022 assessment framework*. Springer. <https://doi.org/10.1007/978-3-031-20113-4>

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. Damiani, V., & Friedman, T. (2024). *Education for citizenship in times of global challenge. IEA International Civic and Citizenship Education Study 2022 international report*. Springer.

Chapter 3:

Questionnaire Development

Gabriella Agrusti, Valeria Damiani, Tim Friedman, and Wolfram Schulz

3.1 Introduction

This chapter describes the development of the questionnaires for students, teachers, schools, and national research centers used in ICCS 2022. While there has been a growing recognition of the importance of questionnaires as crucial parts across international studies of education, for studies of civic and citizenship education these instruments have always played a particularly important role in the measurement of affective-behavioral data as learning outcomes and the factors related to the varying contexts for civic and citizenship education across participating countries. The development followed elaborate procedures to appropriately consider the diversity of education systems involved in a study of this type (see Schulz & Carstens, 2020).

The student questionnaire was designed to measure contextual information (i.e., student characteristics such as age, gender and expected education, information about the students' home and their parents or guardians), 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.

As in the previous cycle, ICCS 2022 also administered regional questionnaires for countries from the European and Latin American regions. These questionnaires were developed in close cooperation with participating countries from each region and designed to address region-specific topics not covered in the international instruments. While at the beginning of the cycle, five or more countries from the Latin American region were expected to join the study, only two of these countries (Brazil and Colombia) implemented the final main survey data collection. Therefore, the Latin American student questionnaire used in the main survey only included item material from the previous cycle except for one question about students' expectations for their individual future adapted from the European regional questionnaire in ICCS 2016.

3.2 Conceptual framework for questionnaire development

The assessment framework provided a conceptual underpinning for the development of the international instrumentation for ICCS 2022 (Schulz et al., 2023). It included the *civic attitudes and engagement framework* as well as the *contextual framework* and both served as reference documents that guided the development of ICCS 2022 instruments (international and European student questionnaires, teacher questionnaire, school questionnaire, national contexts survey), which we will describe in this chapter.

The *civic attitudes and engagement framework* outlines affective-behavioral aspects that were measured with student questionnaire items and included in international and regional instruments. These are regarded as important learning outcomes and had a similar standing in the process of development, analysis, and reporting as cognitive measures of students' civic knowledge. Student attitudes, perceptions, and behaviors relevant to civic and citizenship issues are organized according to the following two affective-behavioral areas:

- *Attitudes* (e.g., judgements in relation to ideas, people, objects, events, or situations) related to the

following three subareas: (i) Civic principles, (ii) civic issues and institutions, and (iii) civic roles and identities.

- *Engagement* (e.g., interest in, and expectations of, civic engagement through civic action and future political participation) related to the following three subareas: (i) Experiences with engagement, (ii) dispositions toward engagement, and (iii) expected future engagement.

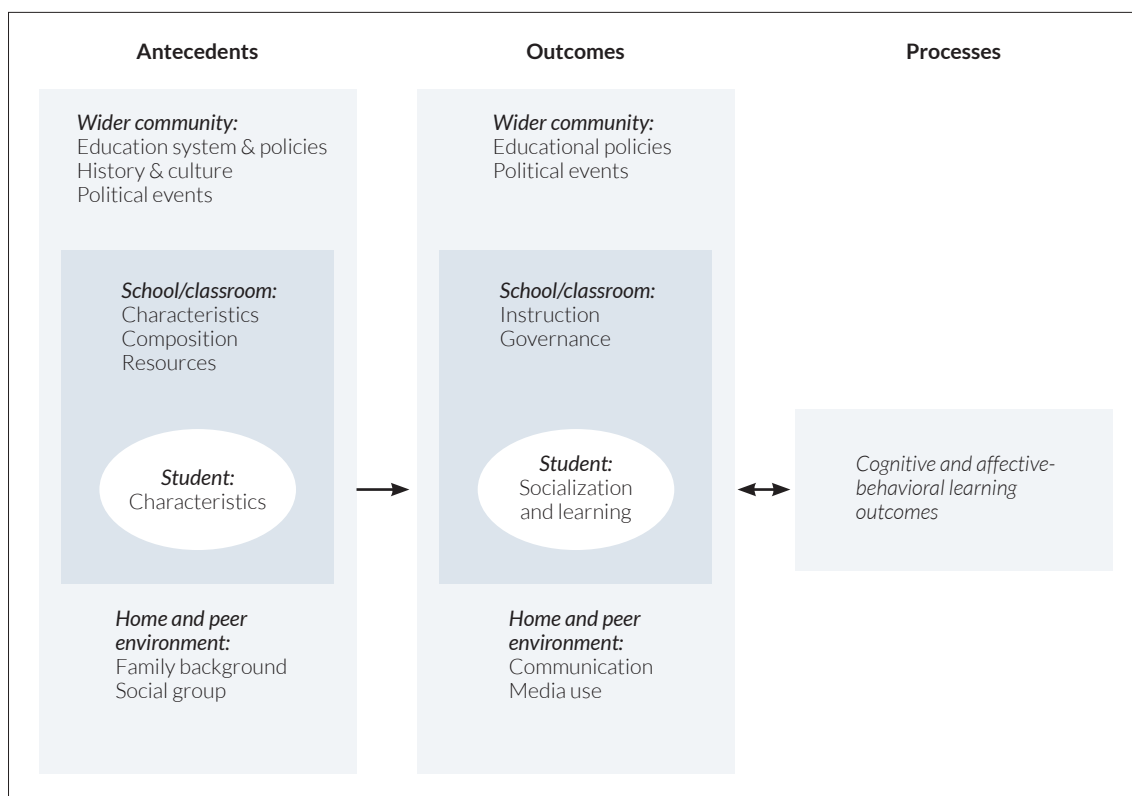
As in previous ICCS frameworks, the contextual framework for ICCS 2022 distinguishes the following four levels at which factors influencing civic learning outcomes may be located:

- The *context of the wider community* is multi-layered and comprises the broader context within which schools and home environments operate (ranging from local to global levels) with relevant factors found at local, regional, and national levels, also considering the importance of the supra-national level (e.g., for data collected within the European Union).
- The context of *schools and classrooms* comprises factors related to teaching and learning, the school culture, and the general school environment.¹
- The context of *home and peer environments* comprises factors related to the home background and the immediate social out-of-school environment of the student (e.g., peer-group activities).
- The context of the individual refers to the individual characteristics of the *student* (e.g., their gender or educational aspirations).

Further, the ICCS 2022 contextual framework distinguishes the following types of variables:

- *Antecedents* are defined as pre-existing variables that shape how student learning and acquisition of civic-related understandings and perceptions takes place.
- *Processes* are defined as those variables that are related to civic-related learning and they are constrained or enabled by antecedents.

Figure 3.1: Contexts for the development of learning outcomes related to civic and citizenship education in ICCS 2022



¹ Because of the sampling design for ICCS, school level and classroom level cannot be disentangled. Generally, only one classroom will be selected within each sampled school.

Figure 3.1 illustrates how these two types of contextual variables and civic learning outcomes are interrelated at each of the four levels. It is important to acknowledge that there may also be reciprocal associations between processes and outcomes variables, for example, civic learning processes may be positively influenced by already higher levels of civic knowledge among students in a given classroom. While the model assumes unidirectional relationships between antecedents and process-related variables, it is also possible that, from a long-term perspective, there could be effects from outcomes on antecedents (for example, through the provision of increased educational resources to schools with lower learning progress).

Different instruments administered as part of ICCS 2022 provided data on different types of variables located at varying levels. Table 3.1 shows examples of variables collected by ICCS 2022 instruments that were administered to students, teachers, schools, and national experts. The national contexts survey and data from published sources provided information at the level of the wider community, while also teacher and school data contributed to the description of national contexts. Teacher and school questionnaires are the primary sources for collecting data about the school and classroom contexts, but the student questionnaire also provided relevant information from the perspective of learners (e.g., regarding the school climate). The latter instrument, apart from measuring civic attitudes and engagement, was the main information source about student background, home, and peer contexts.

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, StQ: Civic knowledge, attitudes, and engagement
School/classroom	ScQ and TQ: School characteristics Resources	StQ, ScQ and TQ: Implemented curriculum Policies and practices School climate	
Student	StQ: Gender Age	StQ: Civic learning Practiced engagement	
Home and peer environment	StQ: Parental socio-economic status Language used at home Country of birth	StQ: Family communication Communication with peers Media information	

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

3.3 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 2016 questionnaire items (including items that had already been administered in ICCS 2009) and reviews of first new draft material by national centers and experts, accompanied by piloting of a draft questionnaire;
- In the second phase, first draft material was finalized following an expert and national center review and tested in an international field trial undertaken across participating countries;
- During the final phase, the results from the field trial were discussed with experts and national centers, followed by a final selection of main survey items.

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

- Variables measured with the instrument needed to be relevant with regard to the ICCS 2022 assessment framework;
- Question wording and content needed to be deemed appropriate across national contexts in participating countries;
- Item material should have high psychometric quality in terms of measuring latent traits postulated in the initial formulation and confirmed during piloting and field trialing.

A pilot version of international student questionnaire material was carried out in five countries (Chinese Taipei, Colombia, Norway, Spain, and Sweden). Three forms each of about 10-15 minutes duration were administered to convenience samples of target grade students to pilot a total of 137 questionnaire items. For parts of the sample, national centers also carried out qualitative group discussions following the administration of the instruments. 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 2022 field trial student questionnaire included a total of 273 items (including 34 items that were optional for countries) and was administered to samples in 20 participating countries during an extended international field trial period between late 2020 and mid-2021.² Due to the impact of the COVID-19 pandemic on education, most countries had to implement their field trial in the first half of 2021 and not in the originally planned time window at the end of 2020. 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 combinations of item sets and scales. In each participating country, data were typically collected from about 1,000 students (on average). As part of the field trial, in countries that used computer-based delivery and had participated in the previous cycle, student questionnaires were administered both on paper and on computer (using smaller extra samples) to allow an analysis of differences across the two modes, which showed only very small effects on questionnaire item responses.

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 2022 questionnaires (see an overview 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 selecting items for the main survey.

The final international student questionnaire consisted of 218 items (including 34 items that were optional for countries). Thirty-two of these items (14 of which were optional to countries) were designed to capture student background information, while 186 were designed to measure the affective-behavioral domains specified in the ICCS 2022 civic attitudes and engagement framework, including 20 items that were optional for countries. The main survey student questionnaire was designed to take about 40 minutes to complete and consisted of the following sections:

- *About you* included questions about student characteristics such as age, gender and expected education;
- *Your home and your family* comprised questions about characteristics of the household and the students' parents;
- *Your activities outside school* consisted of questions about activities reported by students at their home, peer, and community environment;
- *Your school* included questions about students' reports on different aspects of their schools, as well as their participation in civic activities at school;
- *Citizens and society* contained questions about the political system, democracy, and the importance of different behaviors for good citizenship;
- *Rights and responsibilities* included questions designed to measure student support for equal rights for gender groups, all ethnic groups in society, and immigrants, as well as attitudes toward environmental protection;
- *Institutions and society* contained questions regarding student trust in civic institutions, and their level of concern regarding different potential threats to the world's future;
- *Participating in society* consisted of 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* included optional questions pertaining to the international option about religious background and practices, and about their attitudes toward the influence of religion on society.

² Due to long-term school closures during the COVID-19 pandemic, a few countries administered their field trial (often using smaller samples) only later in 2021 and their results could not be included in the field trial analyses.

Most of the items were developed by the international study center, but national centers also proposed new items or modifications to student questionnaire material that were included in the final survey instrument. In light of the impact of the COVID-19 pandemic, an additional item set measuring students' support for restrictions during national emergencies was developed in close liaison with national centers.

Optional items were designed to capture variables that were perceived as either not relevant or inappropriate in some of the participating countries, but that were 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 (e.g., trust in European institutions) within the context of the international student questionnaire, and a number of optional Europe-related items were added to some of international item sets.

The following international options related to students' background were offered to countries participating in ICCS 2022:

- Students' ethnicity (administered in nine education systems);
- Household composition (administered in 18 education systems);
- Number of languages spoken by students (administered in 21 education systems);
- Use of ICT at home (administered in 20 education systems);
- Religious affiliation (administered in 20 education systems); and
- Attendance of religious services (administered in 19 education systems).

The ICCS 2022 student questionnaire also included questions related to attitudes that were optional for national centers:

- Endorsement of equal rights for all ethnic groups in society (administered in 20 education systems); and
- Students' positive attitudes toward the influence of religion in society (administered in 21 education systems).

3.4 Teacher questionnaire development

The teacher questionnaire was developed to collect information about school and classroom contexts, the links between schools and local communities, the perceived objectives of civic and citizenship education, and the delivery of this learning area at school. In detail, the instrument aimed to gather data on characteristics of the school context, including school and classroom climate, teachers' and students' participation in school governance, teaching practices, students' behavior at school and classroom levels, activities in the classroom to address diversity, and issues related to digital technologies, sustainability, and global citizenship. It also captured the relationships between schools and local communities through the measurement of 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 double perspective of teachers and school principals. Specifically, we included some questions about school climate, social relations and participation at school, sustainability, global citizenship education, activities related to digital technologies and diversity, the link between the school and the local community, and civic and citizenship education objectives in both teacher and school questionnaires.

Under the assumption that teaching staff constitute an important factor in determining school and classroom climate, and in shaping 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 that were 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' activities related to global issues, teachers' preparedness and training in civic and citizenship education, and their perceptions of students' opportunities to learn about civic topics and skills.

The questionnaire development process took place in four phases:

- The first phase included development and review of the draft material by the international study center and NRCs. The first questionnaire draft was piloted in a qualitative small-scale pilot study implemented in seven countries. The results from the pilot were used to refine the item material.
- In the second phase, material from the pilot was revised, and the international study center developed a first draft field trial questionnaire, which was subsequently reviewed by national centers and by experts from the Project Advisory Committee (PAC).
- During the third phase, the revised item material was administered to samples of teachers as part of the international field trial in all 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 PAC and the national centers.

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 suggestions from the PAC and the results of a review of the appropriateness of the questionnaire material by national centers, which aimed at highlighting existing differences between participating education systems, as well as between schools within each system.

The field trial teacher questionnaire included 28 questions, with a total of 188 items (3 items were optional) and one filter question introducing the international option for teachers of civic-related subjects at the target grade. It was administered to target grade teachers in the schools selected for the field trial. On average, field trial teacher samples consisted of about 400 teachers in each participating country. The final main survey teacher questionnaire included 26 questions (179 items, with 3 optional items) and was divided into the following four sections:

- *General* included questions about teacher background characteristics;
- *The school* comprised questions about the school environment and issues related to participation in teaching and learning activities at school and in the community;
- *Civic and citizenship education* at school consisted of questions about the aims of civic and citizenship education and about teachers' perception of the importance of citizenship behaviors; and
- *Teaching of civic and citizenship education* contained questions for teachers of subjects directly related to civic and citizenship education and offered as an international option to countries.

3.5 School questionnaire development

The school questionnaire was designed to collect information about school characteristics and school principals' perceptions of school processes that are expected to have an influence on students' civic and citizenship education, for example, school contexts for civic and citizenship education, the context of the local community where the school is located, and the opportunities offered to students to participate in the school life and in civic-related activities in the wider community. The variables included in this questionnaire are related to school size and resources; the school environment; school collaboration with the local community; school autonomy in planning civic and citizenship education; student, teacher, and parent participation in the running of the school; initiatives related to sustainability and global citizenship, and to the use of digital technologies for civic and citizenship education; activities to deal with diversity at school; approaches to civic and citizenship education. In addition to this, the school questionnaire 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 assessed through the school questionnaire were also measured 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 with an estimated completion time of up to 30 minutes.

The questionnaire development process took place in four phases:

- The first phase included development and review of the draft material by the international study center and NRCs. The first questionnaire draft was piloted in a qualitative small-scale pilot study implemented in seven countries.
- In the second phase, material from the pilot was reviewed and a subsequent first draft of the field trial school questionnaire was revised by national centers and by the experts on the PAC.
- 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, that is, 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 25 questions with a total of 140 items (4 items were optional) and was administered to the principals of schools that were selected across ICCS 2022 countries for inclusion in the field trial. In most countries, principals from about 30 schools per country 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 final school questionnaire consisted of 24 questions containing a total of 129 items (with 3 optional items), and was divided into the following five sections:

- *General* included only one question about the principals' career;
- *The school environment* included questions about school governance and students', teachers', and parents' participation at school; students' activities in the community, and initiatives related to diversity, sustainability, use of digital technologies for civic and citizenship education and global citizenship;
- *The local community* contained questions about the resources available to students in the local area and about issues of social tension within the local community;
- *Civic and citizenship education at school* included questions about how civic and citizenship education is implemented at school, schools' autonomy in civic-related activities, and aims of civic and citizenship education; and
- *School size and resources* comprised questions about basic school characteristics like school size, numbers of students, or school location.

3.6 European student questionnaire development

The European questionnaire measures issues that were viewed as of particular relevance for the region. The European-related topics to be included in the instrument were selected after a mapping and reviewing process of potential regional issues against the broader ICCS 2022 framework (Schulz et al., 2023). These issues cover specific attitudes in addition to those already measured within the international student questionnaire.

The development of the European questionnaire was a collaborative effort undertaken in close cooperation between the international study center, the Project Advisory Group, and national centers. In particular, the selection of the region-specific issues to be measured through the instrument was made in close cooperation between the international project team and European national center representatives. These issues, deemed to be of high importance within the regional context, were related to the European identity, opportunities for learning about Europe at school, freedom and restriction of movement within Europe, cooperation among European countries on a different range of issues, including environmental sustainability, sustainable behaviors, the European Union, and the future of Europe.

The development of the European questionnaire consisted of three interrelated phases:

- The international study center made recommendations for the initial instrument development and discussed them with country representatives. NRCs provided suggestions for a refinement of the draft proposal. Some countries also submitted materials to be included in the questionnaire. A first version of the questionnaire was drafted by the international study center based on feedback received from country representatives.
- The draft version of the European questionnaire was tested in an international field trial. All European countries participating in ICCS 2022 administered the field trial version of the European student questionnaire.
- The international study center 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 the material. The version of the European questionnaire administered in the field trial contained 13 questions (including one optional question) comprised of 103 items (4 optional items). The final European student questionnaire had an estimated completion time of up to 20 minutes and consisted of 12 questions comprising a total of 92 items (with 4 optional items) related to:

- Students' sense of European identity;
- Students' reports on opportunities for learning about Europe at school;
- Students' attitudes toward freedom of movement within Europe;
- Students' attitudes toward cooperation among European countries;
- Students' endorsement of environmental cooperation in Europe;
- Students' perceptions of discrimination in their own countries;
- Students' expectations for the future of Europe;
- Students' expectations for their own individual future;
- Students' perceptions of the importance of some aspects of their life in the future;
- Students' reports on political consumerism behaviors;
- Students' reports on their sustainable behaviors; and
- Students' attitudes toward the EU.

3.7 National context survey development

The ways in which students acquire civic knowledge and understanding, and how they develop dispositions toward civic society and civic engagement are expected to be strongly influenced by factors at the national context level. Factors of interest include, among others, the historical background, the nature of the political system, the structure of the education system, and the curricular framework for civic and

citizenship education. The national contexts survey was designed to systematically collect 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 the ICCS consortium and in cooperation with NRCs from participating countries. Staff at the international study center 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.

The national contexts survey for ICCS 2022 was largely sourced from the version used in ICCS 2016, which had undergone major revisions since the inception of this instrument in ICCS 2009 (see Agrusti et al., 2018). The development of the survey consisted of the following process and implementation steps:

- An initial development phase was undertaken between the first and second meetings of NRCs, including initial discussion about the scope and a revision of the questions used in ICCS 2016.
- A second phase consisted of reviewing a draft questionnaire by national coordinators and experts from the PAC, with the aim of further revision and refinement.
- A third phase included gathering NRC feedback, which assisted the finalization of content by international study center staff.
- 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 collections from students, teachers, and schools in their countries.

Using the ICCS 2016 version of the national contexts survey as a starting point, the following aspects were considered when revising the instrument:

- The intended structure of national contextual chapters of international/regional reports;
- New content reflecting new and expanded focus areas;
- The reduction of content that had been included but not been used previously and was unlikely to become relevant for the current cycle;
- Recent changes in education systems and policies related to civic and citizenship education;
- Impacts of COVID-19 pandemic on education systems;
- Impacts of the war in Ukraine on student learning and education systems;
- The reduction of subjectivity of national center responses, where possible; and
- The reduction of response burden on national centers.

During the process of the development of the content, we applied the following criteria when considering 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.

The final national contexts survey included greater inclusion of content related to the new focus areas, alignment with changes to the other questionnaires, and inclusion of content related to the COVID-19 crisis and the Ukraine conflict. The final survey contained 37 questions organized into the following content areas:

- Education system
 - o Background
 - o Structure of the education system
 - o COVID-19 disruptions
- Civic and citizenship education in the curriculum
 - o Education policies related to civic and citizenship education
 - o Civic and citizenship education at school
 - o Civic and citizenship education at the target grade
 - o Current reforms and debates

- Teachers and teacher education
 - o Teacher education in general
 - o Teacher education for civic and citizenship education
 - o In-service teacher education for civic and citizenship education
- Assessments and quality assurance
- The 2022 Ukraine crisis

The online system 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 the required information became available). The online data received from each of the 24 participating education systems were checked for consistency and plausibility by project staff at the international study center. National centers were invited to review the draft tables that were to be used for the international report and would be included in the database. 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 2022 reporting.

References

Agrusti, G., Ainley, J., Losito, B., & Schulz, W. (2018). ICCS questionnaire development. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.). *ICCS 2016 technical report* (pp. 21–32). International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study. *IERI Monograph Series*, 2, 113–135. Retrieved from: https://ierinstitute.org/fileadmin/Documents/IERI_Monograph/Volume_2/IERI_Monograph_Volume_02_Chapter_05.pdf

Schulz, W. & Carstens, R. (2020). Questionnaire development in international large-scale assessment studies. In H. Wagemaker (Ed.). *Validity and Reliability: Quality Control Strategies in IEA's Large Scale Assessments* (pp. 61–83). Cham: Springer.

Schulz, W., Fraillon, J., Losito, B., Agrusti, G., Ainley, J., Damiani, V., & Friedman, T. (2023). *IEA International Civic and Citizenship Education Study 2022 assessment framework*. Springer. <https://doi.org/10.1007/978-3-031-20113-4>

Chapter 4:

ICCS computer-based systems

Tim Friedman and Julian Fraillon

4.1 Overview

In the first two cycles of the International Civic and Citizenship Education Study (ICCS), all student response data were collected on paper. As part of a planned transition to computer-based data collection, ICCS 2022 offered countries the option to collect student response data on computer rather than on paper. In ICCS 2022, 16 countries and the two German benchmarking participants chose to participate in the computer-based data collection while six countries collected all student data on paper.

This chapter describes the key aspects of the computer-based delivery system used in ICCS 2022. It also details some of the challenges of using computer-based systems to collect computerized data in an international large-scale assessment (ILSA) such as ICCS.

The focus of the chapter is on the overall approach, architecture, and design of the computer-based assessment (CBA) system, as well as the relationship of the components of the system to other technical systems used for the survey operations in ICCS. Details and procedural aspects related to the operational use of the components of the CBA system as well as reviews of mode effects are described in chapters associated with their use in ICCS 2022: assessment and test design in Chapter 2, translation and adaptation in Chapter 5, sampling design and implementation in Chapter 6, field operations in Chapter 7, data flow and integration in Chapter 9, scaling and analysis of the test (Chapter 11), and questionnaire data (Chapter 12).

4.2 ICCS 2022 computer-based components and architecture

Preexisting components

Before the inception of computer-based test delivery, IEA was already using computer-based applications to organize and support countries to coordinate field operations and collect questionnaire data from teachers and schools. These applications, developed by IEA, include the following:

- The IEA Windows Within-School Sampling Software (IEA WinW3S), which supports countries to manage within-school sampling and test administration procedures;
- The IEA Data Management Expert (IEA DME), which is used to capture data from paper-based questionnaire material and to integrate and verify national databases;
- The IEA Coding Expert, which is used to code the student responses with respect to parental occupation; and
- The IEA Translation System, which is used to translate and review the school, ICT coordinator, and teacher questionnaires online.

Versions of each of these software applications were configured for and used in ICCS 2022.

RM Results™ (previously SoNET Systems) provided the AssessmentMaster (AM) software suite, used to prepare, deliver, and score the computer-based test and questionnaires to students, and the questionnaires to teachers and school principals. The software components specifically relating to the delivery of the ICCS 2022 materials include: an administration module (AM Manager); an item authoring and translation module (AM Designer); a delivery engine module (AM Examiner), and a scoring module (AM Marker).

4.3 Procedures and software user manuals

In all IEA studies, the Survey Operations Procedures (SOP) manuals guide the work of national center staff. IEA, in close cooperation with the ICCS consortium, has developed and refined instructions and guidelines for IEA WinW3S and IEA DME, and included these in the SOP manuals (see Chapter 7 for details).

During ICCS 2022, separate user manuals were developed for the AM software components. These manuals were further refined between the field trial and the main survey. Use of the relevant features in AM Examiner were incorporated in the manuals for school coordinators and test administrators.

4.4 ICCS 2022 system architecture

ICCS 2022 used a mixture of pre-existing and newly developed and/or customized components and tools to support the preparation, administration, and post-processing of the study. Table 4.1 provides an overview of the ICCS 2022 computer-based systems. The table also shows how these systems interacted. Table 4.1 is organized by the major components of the systems and references the ICCS 2022 respondent populations and response modes.

Table 4.1: ICCS computer-based or computer-supported systems and operations

Process	Student test and questionnaires (computer-based)	Student test and questionnaires (paper-based)	Principal and teacher questionnaires (online)	Principal and teacher questionnaires (paper)
Authoring/master instruments	AM Designer			
Translation	AM Designer			
Sample selection and ID provisioning	IEA WinW3S			
Delivery systems	AM Examiner	Paper-based instruments exported and adapted from AM Designer	AM Examiner	Paper-based instruments exported and adapted from AM Designer
Initialization	Student ID, password, and language information from IEA WinW3S	Labels with student ID from IEA WinW3S	Respondent ID and password from IEA WinW3S	Labels with respondent ID from IEA WinW3S
Primary delivery mode(s)	One school computer per student <i>Online</i> through web-browser <i>Offline</i> through USB sticks: MS Windows-based on local school computers	Paper, administered by TA	Any internet browser: self-administered	Paper: self-administered
Alternative delivery mode(s)	Laptop server mode or carry-in laptop sets (where school infrastructure is insufficient)	N/A	Paper questionnaires (where online delivery is not feasible)	N/A
Data capture	<i>Online</i> : Directly into AM database <i>Offline</i> : Saved to local device (USB or laptop server), must then be uploaded to AM database	Paper-based responses coded and entered into IEA DME software after administration	<i>Online</i> : Directly into AM database	Paper-based responses coded and entered into IEA DME software after administration
Data merging (across respondents)	Upload to AM Manager	Merging from multiple IEA DME databases (where used)	N/A	Merging from multiple IEA DME databases (where used)
Data scoring	AM Marker	Paper-based responses scored and entered into IEA DME software	N/A	N/A
Data management	IEA WinW3S; crosscheck of expected versus available data			

4.5 Functions of the computer-based delivery platform for ICCS 2022

Instrument Preparation using AM Designer

Authoring

All instruments were authored within AM Designer. The authoring functions allowed users to create and edit items (together with any stimulus materials) as well as any ancillary instrument content, such as instructions or information messages. In addition, item metadata reflecting fixed characteristics of items (such as the item format) were recorded automatically within AM Designer and configurable metadata

(such as framework references) could also be stored. All instrument content developed within AM Designer was consequently linked to the adaptation and translation facilities within the system.

Simple item content, based on existing item entry templates, was authored by staff in the International Study Center (ISC). Bespoke authoring, such as for the computer-enhanced dynamic content, was authored by RM developers in consultation with ISC staff.

AM Designer was also used to allocate test and questionnaire content to their respective instruments. This was managed within the system by ISC staff.

Adaptation

Two types of adaptation were possible for the ICCS instruments.

Structural adaptations were possible for the questionnaires but not for the student test. Structural adaptations include some adaptation of the structure of a question for use in a specific country. These could be, for example, the addition of some national items within an international question, the addition of an entirely new question, or creation of additional response categories (that are sub-categories of the original international response category). Structural adaptations needed to be bespoke coded into the instruments for each country by the RM development team. Once structural adaptations requested by a country were approved by the ISC, a request was made to the RM team to implement the agreed adaptations in AM Designer. These adaptations were then checked by the ISC and national centers.

Language adaptations involved the replacement of specific words or phrases in an instrument with locally equivalent terms. This differs from translation in that, in some instances, a direct translation of a word or phrase in the source language may not convey the intended meaning in the target language. Terms or phrases requiring language adaptation were shown in the instruments in square brackets []. Once language adaptations requested by a country were approved by the ISC, they were included in the translations entered into AM Designer by the national center team (see below).

Translation

In ICCS 2022, translations for all instruments were completed in AM Designer. To access the system, users needed computers with the following requirements:

- A computer with a high-speed internet connection;
- Any compatible operating platform (e.g., Windows, Mac OS, or Linux) and;
- A recent version (i.e., less than 12 months old) of a major web-browser (e.g., Google Chrome, Internet Explorer, Microsoft Edge, Mozilla Firefox, Opera, Safari).

Development of the translation systems provided the following features:

- Some selective functionality relating to the role of the user (e.g., administrator, translator, translation reviewer, translation verifier, verification reviewer);
- Opportunity to enter translated text for all screen elements;
- Ability to view the translation history of a text element;
- Ability to view the source text, the translated text, and any previous revisions of the translated text, and to compare different versions of translations;
- A comments interface to allow translators, reviewers, and verifiers to enter comments linked to elements of text;
- Opportunity to view “live” translated versions of the tasks at any point during the translation and to simultaneously view (in a separate window) a live version of the source question or task;
- Ability to enter text as plain text, including editable HTML tags or to enter and use formatting tools to format text;
- Ability to view translated text elements as plain text or as rendered HTML;
- Ability to search for and/or selectively bulk-replace text within and across tasks;

- Ability to bulk-import the field trial translations so that they could be used as a starting point for main survey translations; and
- Opportunity for translators, reviewers, and verifiers to monitor the progress of translation for each item throughout the translation workflow.

Instrument Delivery using AM Examiner

Delivery of the student instruments

In order to maximize participation, the consortium worked with RM results to ensure that the delivery could be as flexible as possible, allowing for two main forms of delivery of the student instruments:

- i. Online (internet-based) delivery; and
- ii. USB-based delivery for use with school computers running Windows operating systems.

The test delivery engine itself was web-based. Offline solutions made use of a locally installed web server (on USB or on the local server) through which the application could be accessed. Although the USB-based delivery engine was Windows-based, it could be run using some Windows emulation software on Mac OS X or Linux.

All offline solutions used a portable version of the Mozilla Firefox browser. All ICCS 2022 materials were developed to be delivered using recent (not more than one year old) versions of Mozilla Firefox, Google Chrome, MS Internet Explorer 11, MS Edge, Apple Safari, and Opera. AM Designer, AM Marker, and AM Manager functioned on recent versions of all major browsers. As part of the field-operation procedures, data collected using offline solutions (USB or local server) needed to be uploaded to a central database as soon as practicable after each test session. Web-based delivery did not require any additional steps, as the data were immediately written to the AM database during each test session.

Two backup delivery alternatives were also available to countries:

- i. Taking laptop computers with AM Examiner already installed into schools; or
- ii. Connecting a single laptop computer into the school network to act as a web server that hosts the student instruments (this was recommended for use with schools using networked computers that did not use Windows operating systems such as Mac OS or Linux). This method is hereafter referred to as the “server method.”

Minimum specifications for each delivery mode of the student instruments

For online delivery, the following minimum specifications were required:

- Internet access with SSL security;
- Screen resolution 1280x800;
- Screen size 9.7 inches;
- Any OS that can run a modern browser (IE 11, Edge, Opera, Firefox, Chrome, Safari);
- Sufficient bandwidth (assessed by the School Coordinator using a practice test);
- Power supply, keyboard, mouse.

For USB-based delivery, the following minimum specifications were required:

- Screen Resolution of at least 1280x800;
- Windows 8 operating system (preferably with patches) or later;
- Display (fonts) set to the optimal size (minimum: 100%; the system font size can be set from the computer’s control panel/system settings);
- A USB Port 2.0 or higher; and
- A quality well-known brand of USB stick with a transfer Rate of at least 3 MB/s.

For local server delivery,¹ the following minimum specifications were required:

Requirements for the **server computer**:

- Windows 8 operating system (up-to date with patches) or later;
- 2.2 GHz Processor;
- 8GB Memory;
- Available Storage Space: 10GB;
- Admin rights;
- LAN or Wi-Fi connectivity.

Requirements for the **client computers**:

- Screen Resolution of at least 1280x800;
- Connectivity to the same LAN or Wi-Fi network as the server computer;
- Recent version (less than 12 months old) of Mozilla Firefox, Google Chrome, MS Internet Explorer 11, MS Edge, Apple Safari or Opera;
- Display (fonts) set to the optimal size (minimum: 100%; the system font size can be set from the computer's Control Panel/Appearance and Personalization/Display).

Checking the suitability of school resources to deliver the ICCS 2022 student instruments

School coordinators were provided with a compatibility check program that was conducted to assess the computers for suitability for administration of the ICCS 2022 main survey student instruments. National Research Coordinators (NRCs) were advised to provide the compatibility check program to school coordinators long enough before the scheduled administration dates to enable time to deal with any issues, but not so long that large changes to schools IT infrastructure were likely to have occurred in the interim. In most cases, the compatibility checks were conducted approximately two months before the dates of testing. The compatibility check program contained three components:

1. The ICCS 2022 technical readiness test: This involved accessing a URL that assessed whether the browser of the computer, the JavaScript version, and the screen resolution met the minimum requirements for delivery of the ICCS student instruments using [online delivery](#);
2. The ICCS 2022 bandwidth test: This involved running a simulated test session using four screens, so that the school coordinator could make a subjective decision on whether the internet bandwidth was sufficient for [online delivery](#); and
3. The ICCS Compatibility Check program: This was an executable file run via USB and was used to assess whether the computer meets the minimum requirements for delivery of the student instruments using [USB-based delivery](#).

Delivery of the ICCS 2022 school questionnaires

In ICCS 2022, the teacher and principal questionnaires were authored and translated in AM Designer, and subsequently delivered using the AM Examiner. These instruments were delivered in the online mode only. Respondents received unique user credentials and national URL to access the instruments online. As with the online delivery of the student instruments, responses were written to the AM online database as they were completed. The online questionnaires were available to respondents to access for as long as staff in national centers deemed appropriate. This was typically between six weeks and two months. Respondents could complete the questionnaires in more than one session if they wished, and they could return to their questionnaire to review and edit responses as often as required until they 'submitted' their responses. Respondents submitted their questionnaire responses by clicking on a button at the end of the questionnaire. Respondents could not log back in after they had submitted their responses.

¹ The server method could only be used in schools with an existing Local Area Network (LAN).

Scoring using AM Marker

AM Marker was used for the scoring of those ICCS 2022 test responses that were collected on computer.

AM Marker was developed to include the following features:

- Selective functionality relating to the role of the user (e.g., administrator, scoring trainer, team leader, scorer);
- Facility to specify a proportion (in ICCS 2022, 20%) of student responses to be blind double-scored for the purpose of monitoring inter-rater agreement; and
- Capacity to begin scoring before all student responses had been uploaded to the system (i.e., before completion of data collection in schools) without compromising the double-scoring procedure.

Scorers were able to:

- View tasks (as they appeared to students in the test) along with student responses on screen;
- Enter a score for each student for each task;
- Flag responses for follow-up with a more senior staff member;
- Navigate back to previously scored responses and amend scores;
- View computer-enhanced items with full functionality; and
- Enter a “training” mode to score pre-scored responses and receive feedback on scoring accuracy.

In addition, team leaders could:

- Review (check-score and amend) scorers' scores; and
- Monitor and respond to flagged responses.

In addition, scoring trainers could:

- Select, score, order, and annotate responses for use in scorer training.

In addition, scoring administrators could:

- Allocate scorers to teams; and
- View reports of interscorer reliability.

Managing the online systems and data collection using AM Manager

The ICCS 2022 administration module, AM Manager, was configured to (i) support national center staff monitor the progress of test sessions (i.e., data upload), create user accounts, and define roles for users of the other modules (scoring and translation), and (ii) export test-session data for importing into IEA WinW3S.

The administration module enabled users to:

- Create user accounts and allocate roles;
- View national test session details;
- Monitor national test session status; and
- Export test session data for use in IEA WinW3S.

4.6 Monitoring data completeness from computer-based delivery in ICCS 2022

As ICCS transitioned to computer-based administration of student instruments for the first time, there were some challenges. Firstly, there was a challenge to ensure students who were administered the instruments on computer were not unfairly advantaged or disadvantaged. A bridging study was included in ICCS 2022 within selected countries using computer-based delivery. The purpose of the bridging study was to investigate any relative impact of the delivery mode (computer or paper) on student test scores and questionnaire responses (see Chapters 11 and 12 for further details).

Successful collection of data in large-scale computer-based surveys relies on individual participants being able to provide responses on computers in controlled uniform environments from which data can be recorded, organized, and stored for later use.

Paper-based assessment and CBA present different challenges regarding the uniformity of test presentation and administration. As the previous cycles of ICCS were conducted on paper, ICCS had well established protocols regarding the production and distribution of paper-based instruments and their administration within schools. The introduction of CBA into ICCS brought with it a new set of administrative challenges to maximize the uniformity of the test-taking experience as well as technical challenges associated with ensuring that the visual on-screen display and user experience was consistent across students. Differences in the user-experience may affect students' performance, and any differences may be more pronounced in complex tasks, such as those involving interactive stimuli. These tasks place greater demands on delivery methods than standard, flat, stimulus material, and multiple-choice response options.

The ICCS 2022 test-delivery platform, including both USB and online-delivery options, was designed to minimize the potential for variation in students' test-taking experience. The layout of the ICCS 2022 instruments could be checked (as a feature of the translation module) and corrected throughout the translation and verification processes and was formally verified during the layout verification process. Chapter 5 of this report describes these processes in detail.

The data collected using AM Examiner during ICCS 2022 needed to be extracted, transformed, and then loaded into IEA systems for processing, weighting, and analysis. These tasks required the interaction of a complex set of computer-based system components (as shown in Table 4.1). As is the case with any system comprising interconnected components with a range of specific functions, the interfaces between each component are potential points of failure. To ensure that the system worked successfully, national research center staff, in collaboration with ISC staff needed to monitor both the integrity of the functioning of the individual components of the system (paying special attention to ensuring results data were being stored) and the interfaces between the components.

It is not possible in this section of the report to look at the full range of sources and types of information described above. Instead, the focus is on key findings relating to one particular type of information, that is, the proportion and typology of cases that school coordinators, test administrators or national center staff classified as technical problems during administration. Nonresponse at the unit or item level, whether due to a lack of cooperation, technical issues, or flawed administration, is one of the more serious factors influencing the robustness and stability of inferences from sample surveys. Therefore, those responsible for implementing a survey nationally regularly monitor response and participation rates, especially as these are used as key indicators of success, quality, and fitness for use.

During ICCS 2022, the examination of unweighted but cleaned data from all 18 participating education systems that administered the ICCS student instruments using CBA yielded some interesting insights. Of the 62,055 sampled students with response data, just over one-third (36%) used USB delivery, and just under two-thirds (63%) used online delivery. Only a small proportion (less than 0.5% of all students), used the server method. CBA was exclusively used in 19 educational systems for the teacher questionnaire, and in 18 educational systems for the school questionnaire. Approximately 80 percent of all teachers completed the teacher questionnaire online, and approximately 74 percent of all principals completed the school questionnaire online.

The online teacher and school questionnaires were available to be completed within a period of several weeks and in multiple sessions at the discretion of the respondent. As such, point-in-time technical issues in delivery of the teacher and school questionnaires were not able to be recorded systematically by a single test administrator.

A review of the survey activities questionnaire (SAQ) completed by NRCs (see Chapter 8 for further details) revealed that national centers had many difficulties associated with the computer-based delivery in ICCS 2022. Difficulties encountered within countries included:

- Various problems with the translation system;
- Problems with production of paper-based materials using the AM suite (extended issues and delays required ISC staff to manually produce these for countries) (see Chapter 5 for further details);
- Issues with the web application, half of the respondents to the SAQ indicated there was a problem;

- Instances of technical problems in the field where the instruments failed to load properly;
- Instances of source material showing for respondents, instead of nationally translated text;
- Instances of unavailability of participation monitoring reports;
- Issues with the marking system for particular items; and
- Instances of data captured using USB delivery not being uploaded to the server properly. Five of the nine respondents to the SAQ indicated a problem here.

Chapter 5:

Preparation and verification of the ICCS 2022 study instruments

Lauren Musu, Kateřina Hartmanová, Jan-Philipp Wagner, Tim Friedman, and Laila Helou

5.1 Introduction

This chapter describes the preparation procedures for national versions of the International Civic and Citizenship Education Study (ICCS) 2022 study instruments. The instrument preparation process aims to maximize comparability of the national instruments to the international versions, maintain equivalence across countries and languages, and ensure appropriateness for each country and participant's national context and education system. To achieve the overarching goals of fidelity to the international source instruments and cross-national comparability, all national ICCS 2022 study instruments underwent a thorough and highly standardized preparation process described in detail in this chapter.

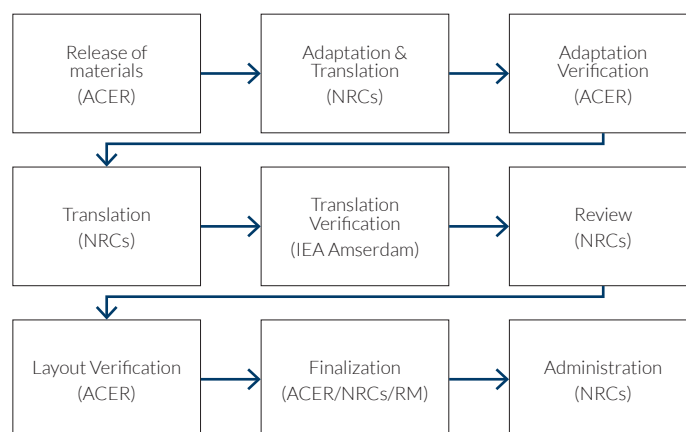
The chapter focuses on the following major topics:

- Adaptation and translation of the international source version of the ICCS 2022 instruments into national languages;
- International verification of the national adaptations and translations; and
- International layout verification and quality control of the final national instruments.

The international versions of the ICCS instruments (see Chapters 2 and 3 respectively for details on the development of these instruments) were produced in English by the ICCS 2022 consortium, in close collaboration with ICCS 2022 national centers. After the release of the international source version of the instruments, all national centers were required to adapt and translate the international version into their language(s) of instruction.

The International Study Center (ISC) at Australian Council for Educational Research (ACER) and International Association for the Evaluation of Educational Achievement (IEA) Amsterdam oversaw the instrument preparation procedures and were responsible for different stages of the whole process. Figure 5.1 shows the overall workflow for preparing the national ICCS 2022 study instruments. The overall process was conducted once before the field trial and then a second time before the main survey. To the extent possible, national centers were encouraged to use the field trial adaptations and translations as a basis for their main survey instruments.

Figure 5.1: Instrument preparation workflow for the ICCS 2022 Main Survey



Notes: NRCs = national research coordinators, RM = RM Results (provider of the software systems underpinning the computer-based student assessment instruments for the main survey – see Chapter 4)

5.2 Adaptation and translation of the ICCS 2022 study instruments

ICCS 2022 instruments requiring adaptation and translation

The ISC provided national centers with the source versions of all materials to be adapted and/or translated. These included:

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

In addition to conventional cognitive items, this cycle included six computer-enhanced items (CEIs) for computer-based administration only (see the Layout preparation section for more details). The CEIs, included in ICCS for the first time, were designed to assess civic knowledge using more interactive functionality made possible by the computer-based delivery mode (see Chapter 2 – Test Development for further details).

All instruments were released together with multiple supporting resources set up to describe and document each step of the adaptation, translation, and verification processes. As written guidance, the Survey Operation Procedures (SOP) Unit 3 described in detail the procedures and all necessary steps of the translation and verification process. Training webinars were also provided to complement the written guidance.

In addition to the ICCS instruments, the procedural manuals and scoring guides for constructed-response items were translated into the national languages. These materials were not subject to the international verification procedures. However, the translations of Test Administrator Manual and School Coordinator Manual underwent a final quality control spot check by an independent quality observer (for further details please see Chapter 8).

Languages used in ICCS 2022

For most countries, the language of the assessment (the target language) is 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 education system and in such cases, on the basis of consultation with national centers, the student instruments could be translated into multiple languages to ensure that the assessment was appropriate for the defined population of students in the country. The decision on which language(s) were used in each country depended on multiple factors such as the population definition of students, the inclusion or exclusion of given students in that definition and consequently whether schools that provide instruction in a minority language were included in the sample of students in a given country.

For ICCS 2022, a total of 22 countries and 2 benchmarking participants prepared 36 different sets of materials in 31 different languages (see Table 5.1). Of the participating entities, eight administered the instruments in more than one language. For the convenience of countries that participated in the Latin American Module (LAM) and had Spanish as their assessment language, a source version of the LAM questionnaire was also released in Castilian. Table 5.1 also shows the delivery mode for each set of instruments. The majority of countries administered the student instruments using one of the computer-based assessment (CBA) modes (online, USB or both) (see Chapter 4 for further details). Those countries that participated in ICCS 2016 and delivered the student instruments using CBA in 2022, also had to complete the ICCS bridging study (see Chapters 2 and 11 for details). Countries completed the teacher or school questionnaires using paper or online delivery, but not both.

Table 5.1: Languages and delivery modes used for the ICCS 2022 survey instruments

Country	Language	Instruments					Delivery mode student instruments	Delivery bridging study instruments	Delivery mode teacher and school questionnaires
		Student Test	International student questionnaire	Regional student questionnaire	Teacher questionnaire	School questionnaire			
Brazil	Portuguese	•	•	StL	•	•	USB	N/A	Online
Bulgaria	Bulgarian	•	•	StE	•	•	Online + USB	Paper	Online
Chinese Taipei	Chinese	•	•	o	•	•	Online + USB	Paper	Paper
Colombia	Spanish (Castilian)	•	•	StL	•	•	Online + USB	Paper	Online
Croatia	Croatia	•	•	StE	•	•	Paper	N/A	Online
Cyprus	English	•	•	StE	•	•	Online	Paper	Online
	Greek	•	•	StE	•	•	Online	Paper	Online
Denmark	Danish	•	•	StE	•	•	Paper	N/A	Paper
Estonia	Estonian	•	•	StE	•	•	Online	Paper	Online
	Russian	•	•	StE	•	•	Online	Paper	Online
France	French	•	•	StE	•	•	Online	N/A	Online
Germany (North Rhine – Westphalia and Schleswig-Holstein)	German	•	•	StE	•	•	USB	N/A	Online
Italy	Italian	•	•	StE	•	•	Online + USB	Paper	Online
Latvia	Latvian	•	•	StE	•	•	Online	Paper	Online
	Russian	•	•	StE	o	o	Online	Paper	Online
Lithuania	Lithuanian	•	•	StE	•	•	Online + USB	Paper	Online
	Polish	•	•	StE	o	o	Online + USB	Paper	Online
	Russian	•	•	StE	o	o	Online + USB	Paper	Online
Malta	English	•	•	StE	•	•	Online + USB	Paper	Online
	Maltese	•	•	StE	•	•	Online + USB	Paper	Online
Netherlands	Dutch	•	•	StE	•	•	Paper	N/A	Paper
Norway	Bokmål	•	•	StE	•	•	USB	Paper	Online
	Nynorsk	•	•	StE	•	•	USB	Paper	Online
Poland	Polish	•	•	StE	•	•	Paper	N/A	Online
Romania	Romanian	•	•	StE	•	•	Paper	N/A	Paper
Serbia	Bosnian	•	•	o	•	•	Paper	N/A	Paper
	Serbian	•	•	o	•	•	Paper	N/A	Paper
Slovak Republic	Hungarian	•	•	StE	o	o	Online + USB	N/A	Online
	Slovak	•	•	StE	•	•	Online + USB	N/A	Online
Slovenia	Slovenian	•	•	StE	•	•	USB	Paper	Online
Spain	Basque	•	•	StE	•	•	USB	N/A	Online
	Spanish (Castilian)	•	•	StE	•	•	USB	N/A	Online
	Catalan	•	•	StE	•	•	USB	N/A	Online
	Galician	•	•	StE	•	•	USB	N/A	Online
	Valencian	•	•	StE	•	•	USB	N/A	Online
Sweden	Swedish	•	•	StE	•	•	Online + USB	Paper	Online

Notes: • = administered, o = not administered, StE = European Student Questionnaire, StL = Latin American Student Questionnaire

Administration modes and instrument formats

Sixteen countries and the two benchmarking participants administered ICCS as CBA and the remaining six countries as a paper-based assessment (PBA). In addition, the 11 CBA countries who participated in ICCS 2016 also administered paper booklets to an additional sample of schools, to provide a link between the two administration modes (see Chapters 2 and 11 for further details about the bridging study).

To maintain comparability across delivery modes, the ICCS 2022 items common to CBA and PBA were designed to have similar visual structural features to students. The six CEIs were administered using CBA only and did not have PBA equivalents.

All instrument content was available for translation within the RM Assessment Master (AM) system (see Chapter 4). The procedures for translation and verification took place in the online translation system for both administration modes.

In the CBA mode, the RM AM system assembled and delivered the translated test and questionnaire content based on the prescribed instrument design that was configured within the system. In for the PBA mode, countries were provided additional support documents showing the test design so that they could manually assemble and check the configuration of the PBA test content. This approach allowed countries/participants to translate each component once only and assemble materials at the end into different booklets to be administered.

The translation system

To prepare the ICCS 2022 study instruments for the main survey, including translation, the national centers used the Designer module of the Assessment Master (AM) online system from RM Results (see Chapter 4 for more details about the system). The system allowed users to apply and edit their translations, keep a full record of the editing history for the text, and to retrace changes made by other users ensuring transparency of the translation and verification process. In addition, the system displayed the differences between a previously saved translation and the current translation in the form of tracked changes, as well as allowed for exports of the translations in the form of an Excel or PDF file. These functions supported national research coordinators (NRCs) and translation verifiers with their reviewing process.

Documentation supporting the instrument preparation process

Any changes, selections, and adaptations to the survey instruments were done with the goal of attaining comparable data from all participating education systems and were documented in the national adaptation form (NAF) for each language and set of instruments. In addition, changes to trend translations (where applicable) were documented in additional trend verification forms (TVFs) for each language.

The national adaptation form (NAF)

The NAF was an Excel document containing a comprehensive adaptation and verification history of each set of national instruments. It consisted of individual worksheets for each separate instrument (the cognitive test, international student questionnaire, teacher questionnaire, school questionnaire, and regional student questionnaire where applicable), as well as separate worksheets for completion instructions, additional country documentation (e.g., language information, version number, inclusion status of international optional questions) and a style guide. The style guide included information indicating the language variant, form of address, and national conventions for numbers, time, grammar, and measurement used in the materials. During various stages of the instrument preparation process, sections of the form were entered and reviewed.

The first version of the NAF was to be completed within national centers. This was done, usually as national center staff began the processes of adapting and translating the instruments. Adaptations recorded in the NAF then required approval by the ISC and their implementation (based on the agreed and approved adaptations) was checked by international verifiers during translation verification. International verifiers and NRCs updated and revised the NAF accordingly 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 further alternations. For the ease of use and documentation of the different stages of verification, the NAF included designated areas for each item, user, and instrument. For countries measuring comparisons in achievement over time, adaptations for existing ICCS 2016 items had to be documented in the NAF, in the same way as the other national adaptations, and the adaptations needed to be kept unchanged as much as possible.

The NAF was an important record of each country's final instrument versions, 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 8 for further details).

The NAF was referenced when adding national data to the international database and during data analysis.

Trend verification forms (TVFs)

TVFs were Excel documents with a structure similar to the NAF, created to keep a record of changes made to trend translations. There were two separate TVFs, one for the cognitive test items and one for the questionnaires.

Documenting a change to a trend translation in the TVF required entering the original (2016) translation, the proposed new (2022) translation, selecting the category of the proposed change (major linguistic error, minor linguistic error, necessary update to terminology, proposal for alternative wording) and providing a justification for the change to the trend translation, including an English back translation. During translation verification, international verifiers reviewed the TVF, commented on the degree of agreement between the original translation and the new translation, and provided linguistic insight on whether the proposed change to translation was necessary. The final decision to keep or reject the change to trend translation was made by NRCs, in consideration of the comments provided during translation verification.

The TVFs were important to keep record of each country's trend translations (where applicable) and additionally used to gain an overview of changes made to these translations. Since such changes in translation could lead to changes in item performance, the overview of changes was taken into consideration during the review of main survey data.

Adaptation process and overview

Together with the instrument release, the ICCS 2022 Consortium provided national centers with written guidance on the procedures required to adapt and translate the cognitive items and questionnaires appropriately, within the SOP Unit 3.

Adapting the instruments to fit the country/participant's context and education system was the first of the key steps in the national instrument preparation process. There were two different types of adaptations that could be applied to the instruments: structural and non-structural adaptations. Structural adaptations refer to alterations that change the structure of the study instruments. This includes adding national questions; adding, splitting, or omitting answer options; and removing questions. National centers were allowed to apply this type of adaptation solely to the background questionnaires and not to the cognitive test items.

Non-structural adaptations were defined as any adaptation that did not alter the structure of the study instruments but could change the meaning of text to fit the local context. The international version of the study instruments indicated where such adaptations were necessary using square brackets to note terms to be adapted (i.e., required adaptations). For example, required adaptations were the [target grade], personal names (e.g., [Male name 1]), places (e.g., [M-town]), or International Standard Classification of Education (ISCED) level equivalents (see <http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf>).

NRCs could also add non-structural adaptations to their instruments wherever they considered these necessary to ensure that terms or phrases fit the national context and were appropriate for the target audience.

To help countries/participants adapt their materials to the national context, the ISC provided them with notes that included brief descriptions of each required adaptation 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). In addition, general guidelines for translation and adaptation were provided (listed in the *Translation and adaptation of cognitive items* section below).

Although NRCs were strongly encouraged to keep adaptations to a minimum, some (required) adaptations were necessary in order to prevent students from facing unfamiliar contexts or vocabulary that could hinder their ability to read and understand an item. These included changes to follow national conventions, punctuation, and expressions of date and time. Where the names of fictional cities or towns were adapted, the use of real place names was not advised to prevent students' responses from being influenced by their perception and knowledge of the names. Instead, names in the target language of similar length, familiarity, and complexity were recommended, 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, direct translations of some terms from English may not result in the term maintaining its meaning across countries. Terms relating to 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 "Eerste 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.

Translation process overview

Along with adaptation, another key part of the national instrument preparation process was the translation of the materials into the target language(s).

Countries were strongly advised to hire highly qualified translators and reviewers who were well suited to the task of working with the ICCS 2022 materials. The translators had primary responsibility for translation of materials. Reviewers were primarily responsible for assessing the readability and accuracy of the translations for the target population. Essential *qualifications for translators and reviewers* included:

- Excellent knowledge of English;
- The target language as mother tongue; and
- Some experience in the country's present cultural and educational context.

IEA also advised the countries that the translators and/or reviewers should have experience with translating texts on social or political issues, translating achievement tests, and working with students in the target grade.

Each national center was responsible for ensuring the suitability of their translators and reviewers. All countries were required to include at least one translator and one reviewer in the instrument preparation process. Some countries employed multiple translators and reviewers, either working together to complete the tasks, or working independently to provide multiple 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 the involvement of professionals familiar with the various languages to ensure equivalent translations across all the national versions.

Translation and adaptation of cognitive items

One of the main challenges when translating the ICCS 2022 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.

In the **translation guidelines**, translators and reviewers were asked to ensure that:

- The translated text did not explain 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.

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 (or omitted) in the translated national version. As previously mentioned, no additions were permitted for the assessment.

Cognitive items for trend comparison over time

Fifty-five test items were carried over from the previous cycle (ICCS 2016) in order to measure changes in student achievement over time. These items (trend items) provide the basis for comparisons of changes in achievement from the earlier assessment for those countries that participated in both study cycles with the same target population. To ensure the quality of comparisons over time, these items needed to be administered in the same way in each cycle.

For countries that previously participated in ICCS 2016, the translations of the items used in the previous assessment (trend translations) were compared against the ICCS 2022 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 in the TVF and referenced during the data analysis. Some items with changes were not included in the comparison analyses over time for that country. For countries that did not participate in ICCS 2016, the preparation of these specific items followed the same general procedures as the newly developed test items for the current cycle.

Translation and adaptation of the questionnaires

National centers were also required to adapt some terms in the questionnaires and 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 adapted.

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 Bulgaria, these terms would be replaced by equivalents “Bulgarian” 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 the translation would be equivalent to “students’ representative body” in English.

Countries were permitted to add a small number of national questions or categories within the existing questionnaires (these are examples of structural adaptations). NRCs were instructed to place all substantive national items at the end of the international questionnaires. All national questions needed to be documented and approved by the ICCS 2022 Consortium before they could be included in the questionnaires.

Layout preparation

The final step in the country's preparation of national instruments was layout preparation. NRCs were asked to review each question in the proposed delivery mode (paper-based and/or computer-based) to ensure that the content was accurate, and that the layout and functionality were working as intended.

For all instruments (irrespective of delivery mode), NRCs were asked to check for the following:

- Content is accurate and matches translations entered;
- Source version content is not displayed at all (unless this is intended);
- No adaptation brackets or html tags are displaying;
- National extensions are displaying appropriately; and
- All content is displayed appropriately (e.g. text not spilling outside of boxes/screen or stimulus material not spanning across pages in a printed document).

5.3 International verification processes

Adaptation verification

Following completion of adaptation and translation at the national centers, NRCs were required to consult with ISC staff for a final review of all proposed national adaptations (see Figure 5.1). 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 for consultation with the ISC. Following its review, the ISC provided the national centers with feedback on their adaptations and, where appropriate, made suggestions for better alignment to the source version. Some of the common issues that were identified during adaptation review were:

- Incorrect completion of the NAF;
- Inconsistent adaptations used within or across test booklets or questionnaires;
- Inconsistent adaptations used in comparison to previous cycle;
- Difficulties in establishing country appropriate adaptations for ISCED levels;
- Difficulties identifying an appropriate set of subjects that are 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.

Translation verification

The national translations of the instruments underwent international translation verification. IEA oversaw the international translation verification process in coordination with an external translation verification company, cApStAn (based in Brussels, Belgium). 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;
- Being a professional translator/interpreter/translation verifier;
- A good level of computer literacy; and
- Residency in the target country, or close contact with the country and its culture.

Verifiers were trained through an online seminar, where they received detailed instructions for reviewing the survey instruments and documenting deviations from the international source version. They also received general information about the study and design of the instruments, together with a brief description of the translation procedures used by the national centers.

Verifiers were provided with comprehensive materials to support their work. Each verifier received access to the relevant manuals and instruments in the online system, as well as a comprehensive set of directions, instructions and relevant examples for reviewing the national instruments and documenting deviations from the international source version. IEA and cApStAn also ensured all international translation verifiers received continuous support and provided the verifiers with constructive ongoing feedback as needed.

There were three modes of translation verification in the ICCS 2022 main survey: full verification of new translations, focused verification of translations which were already verified for the field trial and trend verification of trend translations. 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, checking correct implementation of national adaptations, consistent use of words/terms listed in the style guide, and identicalness of translation with previously used translations (where applicable);
- Documenting any deviations (between the national and international source versions, or between new and previously used translations), including additions, deletions, and mistranslations; and
- Suggesting an alternative translation/adaptation to improve the accuracy and comparability of the national instruments.

Materials undergoing focused verification for the main survey had already completed a full translation verification during the field trial. Therefore, this process was only applied to items where a translation export provided to verifiers by IEA (via cApStAn) indicated a change in the translation between the field trial and the main survey.

Verifiers provided feedback from translation verification in both the AM online system and in the associated NAF. Verifiers were asked to correct the text of the cognitive test items and questionnaires and/or to add comments specifying errors using the edit and comment functions in AM Designer. Corrections or suggestions could be applied directly to the text elements in the online environment. Any changes made by the verifier were displayed as track changes in the text element. Using the comments, verifiers were required to document their corrections and suggestions, describe the issue, and provide an English back translation for every affected text element.

For materials undergoing trend verification, a key part of the verifiers' feedback was focused on the changes in the translation of the items that were meant to be kept identical for the comparison over time. 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;
- Checking proposed changes to trend translations documented in the TVF, their adequacy and correct implementation; and
- Documenting any undocumented differences in trend translations in the TVF.

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") and the ISCED levels were a particular challenge to adapt for some countries. Countries were informed of this and asked to address all discovered issues. Any adaptations reported in the NAF were also reviewed by the verifiers, who were asked to comment on their adequacy.

With the documented comments and suggestions from the verifiers, NRCs were able to revise and improve their national versions. Each deviation of the translated text from the international source was allocated one of five codes that indicated to the NRCs the severity and type of deviation:

Code 1: Major Change or Error. Examples include 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 link items to compare results across cycles (trend items); and agreed national adaptations not being implemented in the files.

Code 1? was assigned where the verifier was in doubt over the severity of an error or where the verifier was unsure how to correct a possible error. 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 OK: Acceptable Adaptation. The adaptation (either required, where <text is in angle brackets>, or nonrequired yet necessary) is acceptable and appropriate.

All issues that the verifiers deemed to be severe deviations from the source (i.e., Code 1), as well as all issues related to required adaptations, were documented in the NAF where they could be addressed by the country and reviewed by the ISC.

After translation verification, IEA reviewed the verifiers' feedback and returned the materials to the national centers. NRCs were responsible for finalizing their instruments and thus had to review the feedback carefully and consider the corrections and suggestions made by the verifier.

The procedure required NRCs to react to every comment the verifier made in the AM Designer, the NAF, and TVFs (where applicable). In principle, NRCs could accept, modify, or reject the verifier's suggestion. For the latter two options, NRCs had to give an explanation for why they changed or disagreed with the suggestion.

For issues involving adaptation implementation and trend translations, NRCs were further requested to discuss any undocumented or unclear changes with the international project team. The project team additionally assessed the appropriateness of each trend change (using the verifiers' linguistic expertise and input) and advised the NRCs on the consequences of implementing the changes. This information supported NRCs in making their final decision to keep or reject the trend change. Information about trend changes across items and languages was later used as a reference during the analysis of main survey data. Findings from an analysis of the 121 documented trend translation changes showed no significant associations between those changes and DIF (differential item functioning) for the trend items.

Layout verification

Once translation verification had been completed, the ISC asked national centers to make any changes resulting from translation verification and notify the ISC that their student materials were ready for layout verification. Layout verification is the final external review of the assessment instruments and questionnaires, its main goal is to ensure comparability of assessment instruments across all countries. During this stage, the ISC checked each participating country's materials to ensure they were appearing correctly, had the correct structure and function and importantly that they were comparable to the international source.

For those CBA countries and benchmarking participants, layout verification also included reviewing their instruments in their final test forms to verify that all elements of the instruments including system login, orientation, section logins, scrolling reminders, navigation, progress bars, and filter rules all functioned and appeared correctly.

Countries using paper instruments had their instruments reviewed in AM. These were then exported from the platform into PDF format where they were reviewed for pagination and alignment with the international source and as well for formatting issues. Where countries were using both paper and computer-based instruments the final versions of both were compared to ensure consistency.

The most common layout issues that were identified across the instruments were:

- Formatting issues (e.g., spacing, font size, ensuring consistency with international source version);
- Issues with structural changes in the questionnaires; and
- Technical issues related to the translation system.

Some of the interactive units presented issues that required correction by the software developer team, especially for languages where the translated segments were much longer than the source.

All inconsistencies were listed against the relevant section in the NAF and reported back to the national centers for their review and revision. Countries using the computer-based assessment were given test form logins to review their draft final forms and indicate any unresolved issues. While countries using paper forms were provided with their final forms in PDF format for final review.

This final review by the national center sometimes resulted in the identification of issues that required correction. These were mostly spelling corrections that were corrected in the platform by the national center and then the final instruments were reproduced and reviewed by the ISC before sending them back to the national center for final review.

The production of instruments and layout verification were impacted to some degree by issues within the online system (see Chapter 4 for further details). Consequently, some layout issues that were intended to be fixed by national center staff, had to be fixed by staff from the ISC or from the software development team. The overall quality of the instruments was not impacted, but it did result in delays in finalizing instruments for some countries.

Quality observers' review

As a final step, international quality control observers reviewed the implementation of verification feedback after the data collection period and documented irregularities, if any. Furthermore, they reviewed the translations of a few crucial parts of the Test Administrator and School Coordinator manual and their alignment with the international source (for more details see Chapter 8 on quality assurance procedures).

Chapter 6:

Sampling design and implementation

Diego Cortes and Umut Atasever

6.1 Overview

To achieve accurate, precise, and internationally comparable estimates of the population parameter of interest, the International Civic and Citizenship Education Study (ICCS) aims to select high-quality random samples of students and teachers. This chapter describes the random sampling procedures used to select samples of teachers and students to participate in ICCS 2022. These were developed in accordance with the requirements for sampling quality outlined in the *Technical Standards for IEA Studies* (Martin et al., 1999).

The approach to sampling for ICCS 2022 followed a stratified two-stage random sampling design. During the first sampling stage, schools were sampled with probability proportional to their size, where size was defined as the number of students enrolled at school that belong to ICCS target population. Across sampled and participating schools, during the second stage, one or more intact class(es) of target-grade students were randomly selected for inclusion in the student survey. All classes within the same school had the same probability of being selected to participate. Further, a fixed number of target-grade teachers was randomly selected in each sampled and participating school for the teacher survey. This chapter provides a description of this sampling design, with special attention to the following topics:

- The definition of all relevant populations that were targeted in ICCS 2022;
- The definition of the parts of the population not covered by or excluded from ICCS 2022;
- International sample design with special attention to;
 - o Sample sizes,
 - o Sampling procedures (school and within-school selection).

An important feature of ICCS 2022 was the introduction of computer-based assessment (CBA) in this study (see Chapter 4). Countries were given the choice to administer ICCS using either computer-based or (traditional) paper-based delivery, with 18 out of the 24 participating countries or benchmarking entities opting for the former. Both versions of the assessment were developed to measure the same latent constructs and manifest indicators. Yet, the dual assessment mode presented us with two important challenges. First, it was important to measure the risk of, and control for, any systematic difference in student responses across the two administration modes. Second, to protect the measurement of trends from previous cycles, where appropriate, it was essential to adjust for potential mode effects (for more detail about this, please see Chapters 11 and 12). For these reasons, countries that participated in ICCS 2016 and also opted for the CBA in 2022 were asked to assess a separate sample of schools and/or students to provide ‘bridging’ data, which contributed to the ICCS 2022 *Bridging Study*. That is, the bridging sample represented the same student target population as the sample participating in the CBA but was administered a paper version including a reduced version of test items (in a format that could be delivered in both modes) as well as the international and (where relevant) regional student questionnaire material. Since the bridging study was developed to inform about mode effects for student test and questionnaire data, it did not include the teacher survey.

Overall, 11 out of the 18 countries administering the computer-based version of the survey participated in the bridging study. These countries were: Bulgaria, Colombia, Estonia, Italy, Lithuania, Latvia, Malta, Norway, Slovenia, Sweden, and Chinese Taipei. This chapter also provides information about how schools and students were selected to participate in the bridging study.

6.2 Target population definitions

When undertaking a quantitative study, it is essential to clearly define the target population that is under study. Survey estimates derived from a random sample allow drawing conclusions about the population

from which the sample was selected. As ICCS 2022 was designed as both a student survey and a teacher survey, two distinct target populations had to be defined.

Student target population

As in the two previous cycles, ICCS 2022 defined the target population of students as follows (see, for example, Weber, 2018):

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 1,¹ 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. As in the previous cycle, Norway assessed Grade 9 instead of Grade 8 and its results were annotated in the ICCS 2022 reporting tables.

Students who were not covered by the definition above were regarded as “out of scope” (i.e., students in a different grade than the target grade). In the following sections, the term “students” is used to describe “students in the ICCS 2022 target population.”

Teacher target population

As in previous cycles, ICCS 2022 defined the target population of teachers as follows (see, for example, Weber, 2018):

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

This definition included all teachers teaching regular school subjects to students of the target grade (regardless of the subject or the number of hours taught) during the ICCS 2022 testing period.

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

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

In the following sections, the term “teachers” will be used to describe “teachers of students in the ICCS 2022 target population.”

6.3 Coverage and exclusions

Population coverage

The ICCS 2022 international sampling team encouraged all participating countries and benchmarking entities to include in the study all students and teachers that were covered by the target population definition. However, countries could choose to remove larger groups of schools, students, or teachers from the target population for political, operational, or administrative reasons. This is referred to as reduced population coverage. Note that in ICCS 2022 all countries and benchmarks were considered to have full coverage.

¹ ISCED stands for International Standard Classification of Education (UNESCO, 2011).

Student exclusions

In most ICCS 2022 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 was estimated as the sum of the school-level exclusion rate and the within-school-level exclusion rate. The former was calculated based on information provided by the NRCs prior to sampling. Whereas the within-school-level exclusion rate was estimated using the data collected for ICCS 2022 across participating schools. Each country was required to keep the overall rate of excluded students below five percent of the target population. Four participating countries exceeded this limit, and their results were annotated in the ICCS 2022 reporting tables.

Prior to the selection of the school sample, national centers had to define the group of schools that had to be excluded in their respective national contexts from the ICCS 2022 student survey. However, the within-sample exclusions could only 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). NRCs were instructed on how to define and identify potential groups of students to be excluded. The following definitions were shared with all countries and benchmark entities:

- Intellectually disabled students—These are students who are considered in the professional opinion of the school principal, or by other qualified staff members, to be intellectually disabled or who have been tested psychologically as such. This includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students should not be excluded solely because of poor academic performance or normal disciplinary problems.
- Functionally disabled students—These are students who are permanently physically disabled in such a way that they cannot perform in the ICCS testing situation. Functionally disabled students who are able to respond should be included in the testing.
- Non-native language speakers—These are students who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier of the test. Typically, a student who has received less than one year of instruction in the language(s) of the test should be excluded, but this definition may need to be adapted in different countries.

Any other types of within-school 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.

Across the 11 countries that assessed a bridging sample, the same set of school- and student-level exclusion categories used were used as for the bridging study.

Teacher exclusions

Unlike the context described above about the student survey, there was no intention to exclude teachers from the ICCS 2022 teacher survey. If a teacher was part of the teacher target population, they were eligible to participate in the study. Therefore, no minimum exclusion rates for teachers were specified. However, exclusions of schools from the survey also had an effect on the coverage of the teacher population. 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, only rough estimates, or no estimates at all. Therefore, it was not possible to systematically collect information about exclusion rates for the ICCS 2022 teacher survey.

Overall exclusions

Table 6.1 shows the population coverage and the exclusion rates for the student survey across all countries participating in ICCS 2022. Since the set of countries assessing a bridging sample used the same exclusion criteria in the core survey as in the bridging study, its exclusion rates were estimated using the same population parameter. Therefore, Table 6.1 does not distinguish between exclusion rates for ICCS 2022 core survey and the bridging study.

Table 6.1: Population coverage and exclusion rates

Country	Student Survey			
	Population coverage (%)	School-level exclusions (%)	Within-sample exclusions (%)	Overall exclusions (%)
Bulgaria	100	0.2	2.0	2.2
Brazil	100	2.3	0.9	3.1
Colombia	100	0.2	1.9	2.2
Chinese Taipei	100	0.8	1.8	2.6
Croatia	100	2.5	5.2	7.6
Cyprus	100	1.2	2.2	3.4
Denmark	100	2.9	2.5	5.4
Estonia	100	2.8	2.1	4.9
France	100	2.9	1.4	4.3
Italy	100	0.8	3.7	4.5
Lithuania	100	2.9	1.1	4.0
Latvia	100	5.9	1.8	7.7
Malta	100	1.4	2.4	3.8
Netherlands	100	3.8	1.5	5.3
Norway (9)	100	3.3	4.2	7.4
Poland	100	2.0	1.9	4.0
Romania	100	3.2	1.1	4.3
Serbia	100	1.0	1.9	2.8
Slovak Republic	100	0.7	0.5	1.2
Slovenia	100	2.9	1.1	3.9
Spain	100	1.0	3.3	4.3
Sweden	100	2.1	4.3	6.4
Benchmarking participant				
North Rhine-Westphalia	100	2.5	1	3.4
Schleswig-Holstein	100	1.5	0.6	2.1

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

6.4 Sample size requirements

ICCS has strict standards about sample sizes, defined as the number or expected number of participating units. Where units can be schools, as well as students or teachers within participating schools. Importantly, the same standards, which are explained below, applied for the bridging study.

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 survey outcome. This means that the complex sample design used in ICCS 2022 should achieve the same sampling precision as a hypothetical simple random sample of 400 students. The loss in precision associated to a complex sampling design is known as *design effect*. Importantly, one of the main drivers of the design effect in ICCS is a result of the two-stage sampling design, as students from the same schools tend to be more alike than students from different schools. Therefore, to have robust estimates it was necessary to survey a far larger number of actual students to achieve this goal.

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

The ICCS sampling team asked, with respect to the ICCS 2022 student survey, each participating country to have a minimum 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 considered, these requirements were expected to result in about 3,000 surveyed students.

Countries or benchmarking entities 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 were implemented for different reasons:

- As shown in previous student surveys, variation in student achievement across schools in a country can be large depending on the structure of the education system (e.g., due to segregation). Evidence from previous studies in ICCS 2022 countries was used to determine whether 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 achieve the student sample size requirement of 3,000 students based on a sample of 150 schools;
- The NRC requested a sample-size increase in order to increase the amount of data available for analysis (e.g., through oversampling of certain sectors in the education system).

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

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

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

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

Countries participating in the bridging study required a sample size of 1,500 students. Two different strategies were offered to coordinate the selection of students participating in the core survey and the bridging study. The first one was to include a set of schools to participate in the core survey and a distinct set of schools to participate in the bridging study. This meant there should not be any overlap between the two school samples. The second option was to include schools either in the core survey or bridging study, with the possibility of assigning one classroom to the bridging and one or more to the core survey in larger schools. This second option meant a partial overlap between both school samples. It is important to note that there were no classrooms or students that were selected to participate in both the core survey and bridging study. Chinese Taipei, Colombia, and Slovenia opted for the design with a partial overlap between the two school samples, while all other countries participating in the bridging study using separate school sample.

Sample size in the teacher survey

The school sample size requirements for the ICCS 2022 teacher survey were the same as those for the student survey. Within each selected school, a minimum 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 a few teachers were not included in the survey. Teacher samples sizes were determined similarly as for students; that is, the effect of the sample design was considered, and the sample size determined to achieve a similar precision level as a hypothetical simple random sample of 400 teachers.

Some NRCs requested that all teachers in sampled schools who were teaching civic-related subjects be part of the national teacher sample. Other countries expressed the wish to select all home-class teachers in a sampled school. In both cases, teachers from these groups were sampled with certainty.

Overall sample sizes

Table 6.2 lists school sample sizes, as well as the number of schools, students, and teachers participating in ICCS 2022 core survey.

Table 6.2: School, student, and teacher sample sizes in ICCS

Country	Originally sampled schools	Student Survey		Teacher Survey	
		Participating schools	Participating students	Participating schools	Participating teachers
Bulgaria	150	144	3113	140	1806
Brazil	222	190	5997	184	1716
Colombia	150	148	4193	120	1202
Chinese Taipei	150	148	3852	149	2307
Croatia	150	141	2766	148	2290
Cyprus	98	95	3136	85	1147
Denmark	216	134	4769	70	237
Estonia	202	166	2685	145	1721
France	150	147	3533	97	1091
Italy	160	158	2945	157	2121
Lithuania	184	177	3560	176	2611
Latvia	152	147	2876	118	1638
Malta	29	29	2720	26	415
Netherlands	150	124	2609	92	1062
Norway (9)	149	142	5687	123	1355
Poland	170	169	4437	167	2259
Romania	158	154	2768	153	2242
Serbia	150	138	2659	150	2257
Slovak Republic	160	154	3202	153	1907
Slovenia	174	168	3466	168	2461
Spain	160	157	3487	157	1954
Sweden	154	149	3263	131	1559
Benchmarking participant					
North Rhine-Westphalia	152	145	3267	129	2916
Schleswig-Holstein	121	84	1487	13	173

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

Similarly, Table 6.3 lists school sample sizes, as well as the number of schools and students participating in the bridging study.

Table 6.3: School and student sample sizes in the ICCS bridging study

Country	"Originally sampled schools"	Participating schools	Participating students
Bulgaria	64	55	1202
Colombia	48	47	1356
Chinese Taipei	55	55	1375
Estonia	97	86	1276
Italy	72	72	1402
Lithuania	70	69	1365
Latvia	52	48	943
Malta	13	13	1180
Norway (9)	50	49	1921
Slovenia	78	73	1478
Sweden	69	66	1429

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

6.5 School sampling design

Stratification of schools

Prior to sampling, the population of schools was *stratified*. That is, schools sharing some observable characteristics were put into mutually exclusive groups. Generally, ICCS 2022 used stratification for the following two 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 ICCS 2022 outcomes;
- To apply different sample designs, such as disproportionate sample allocations, to specific groups of schools (e.g., states or provinces).

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 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 IEA sampling team members. Appendix B of this report provides details about the stratification variables for each participant.

Countries transitioning to CBA implemented the same stratification design in the bridging study as in the core survey.

School sampling frame

To prepare the selection of schools that should participate in ICCS 2022, the IEA sampling team asked national centers to provide a list of all schools with students enrolled in the target grade. This comprehensive national list of all eligible schools is called a *school sampling frame*. The team carefully verified that all school-sampling frames 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 corroborated 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;
- Information about the intended stratification variables.

Schools sample selection

To select the school samples for the ICCS 2022, the sampling team used *stratified systematic sampling*, where the probability of being selected for each school was proportional to its size. As noted earlier, this method is customary in most large-scale social surveys, and notably in most IEA surveys.

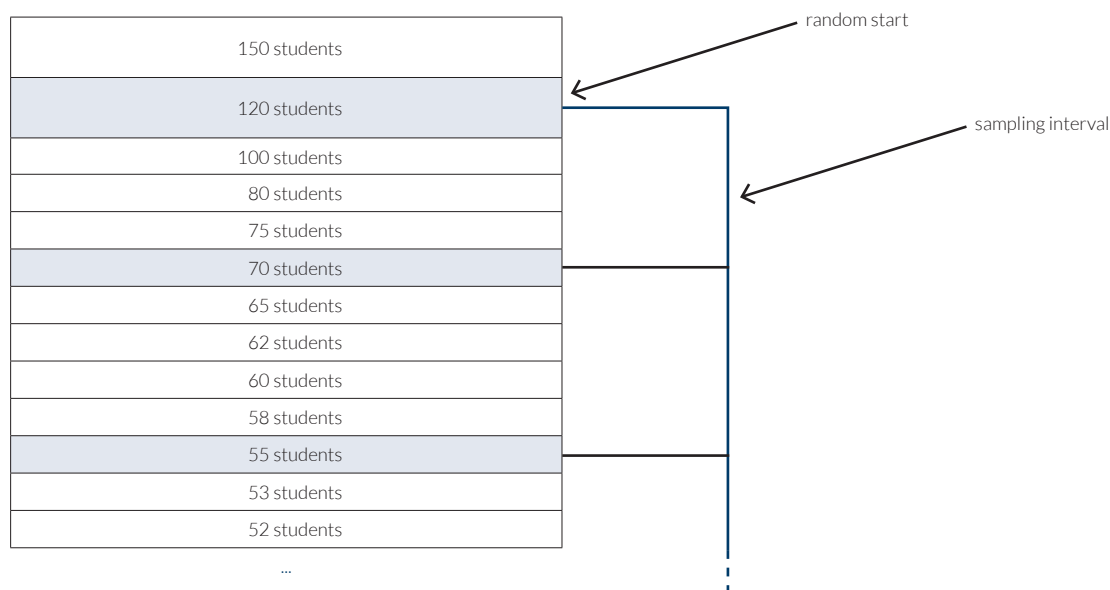
The process of selecting the school samples of each country started with the sorting of the school sampling frame. Within each explicit stratum, schools were sorted by implicit strata, and finally within each implicit stratum by MOS (alternately sorted in increasing and decreasing order).

The team next selected a sample from the sorted school sampling frame with the following steps:

- Calculating a sampling interval in each explicit stratum, this interval was defined by dividing the total MOS in this stratum by the number of schools to be sampled from that stratum;
- Determining a random starting point in each explicit stratum, defined as the product of the sampling interval times the value of a random draw from a standard uniform distribution. The school with the cumulative MOS that was equal or larger to the corresponding product was the first sampled school;
- Selecting schools 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 was equal or larger the corresponding value, the team selected the corresponding unit.

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

Figure 6.1: Systematic selection of schools with probability proportional to their size



In certain cases, the team systematically deviated from this general procedure. If very small schools are selected with probability proportional to size (PPS), there is a risk of obtaining extremely large sampling weights for students from those schools and therefore substantially increasing the variability of the weights. To prevent this, it was 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.

Most ICCS 2022 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, this overlap control involved selecting the main survey sample and the field trial sample simultaneously.

The sampling team selected a sample of replacement schools while selecting the primary sample of schools. 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. That is, replacement schools were similar to originally sampled schools in terms of size and stratum affiliation. 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. Replacement schools were used to ensure that enough information about the target population was collected for ICCS 2022. However, it was emphasized that collecting data within the set of originally sampled schools was a priority, as replacing these with (similar) replacement schools required making very strong assumptions about distributions.

For countries using CBA that had participated in the previous cycle, the samples of schools for the core and the bridging study were drawn simultaneously. In a first step, a relatively large sample of schools was drawn. Then, sampled schools were randomly allocated to participate either in the core survey, the bridging study, or both (if applicable). For example, a country may have required 150 schools to be sampled for the core survey, 75 schools to be sampled for the bridging study and required no school overlap between both studies. Moreover, let us assume an average class size of 20 students. Under this scenario, 225 schools were sampled in a first step, while in a second step, 75 schools were randomly allocated to participate in the bridging study, and the remaining to participate in the ICCS 2022 core survey.

6.6 Within school sampling design

Within-school sampling constituted the second step 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

One or more classes from each participating school were included in the ICCS 2022 assessments. All participating schools were asked to list all their target-grade classes and to provide this list to their ICCS 2022 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 2022.

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.

An additional within-school sampling step was required for countries transitioning to CBA and that chose to have an overlap between the core survey and the bridging study school samples. In such cases, some schools were selected to participate in both studies. Countries were asked to use WinW3S to randomly allocate sampled classes to participate in the core survey or in the bridging study.

Teacher sampling

WinW3S software employed systematic sampling with equal selection probabilities to select teachers from lists provided by the participating schools. To ensure a proportional allocation of teachers by gender, 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*. International Association for the Evaluation of Educational Achievement (IEA).

UNESCO (2011). *International Standard Classification of Education: ISCED 2011*. UNESCO Institute for Statistics.

Weber, S. (2018). Sampling design and implementation. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.), *ICCS 2016 technical report*. International Association for the Evaluation of Educational Achievement (IEA).

Chapter 7:

Field operations procedures

Hannah Kowolik and Falk Brese

7.1 Overview

Successful administration of the International Civic and Citizenship Education Study (ICCS) 2022 depended heavily on the contributions of the study's national research coordinators (NRCs) and national center staff. To assist the national centers and to promote consistency in the study administration across countries, the ICCS international study center (ISC) at the Australian Council for Educational Research (ACER), as well as the associate research center at Lumsa University worked together with the International Association for the Evaluation of Educational Achievement (IEA) to develop internationally standardized field operations procedures. The ICCS operations procedures were designed to be flexible enough to meet the needs of individual participants whilst maintaining IEA's high quality survey standards.

The ICCS operations procedures were developed with reference to those used in ICCS 2016 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 tailored to suit the specific requirements of ICCS 2022. The 2022 cycle of ICCS marked the transition to computer-based assessment, with participating countries having the option of using paper-based administration (PBA) or the new computer-based administration (CBA) of the student instruments. As occurred in previous cycles of ICCS, the teacher and school questionnaires could be delivered on computer or on paper.

Adapting operational procedures from the existing ICCS operations and integrating the CBA operations to create a workflow for this new assessment mode was a significant undertaking. In order to control for any assessment mode effects, countries transitioning to CBA were required to administer "bridge" paper instruments to an equivalent sample of students, in addition to their nationally representative sample of students. This bridge data collection also required integrating operations and procedures into the overall ICCS 2022 assessment administration.

All national centers received guidelines on survey operations procedures for each stage of the study. These guidelines provided advice on contacting schools, listing and sampling students, preparing materials for data collection, administering the assessment, scoring and coding 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.

7.2 Field operations personnel

National research coordinators

One of the first steps in establishing the ICCS study was the appointment of an NRC (or NRCs) by each participating country or education system.¹ 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 within their countries, 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.

¹ The majority of the entities that participated in ICCS were countries. Further, two subunits of countries featuring a distinct education system also participated in ICCS: The German states North Rhine-Westphalia and Schleswig-Holstein. For reasons of simplicity, the text will refer to both participating countries and education systems as "countries."

School coordinators and test administrators

School coordinators and test administrators were key to successful administration of ICCS. 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 a school. The school coordinator could also be the test administrator at a school, but was not to be a teacher of any of the sampled students. In some cases, national centers appointed external school coordinators, for example from their own staff. 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 school coordinators' responsibilities included the following major tasks:

- Identifying eligible students and teachers belonging to the target populations to allow the national center to perform within-school sampling;
- Arranging the date(s) and modalities of the test administration with the national center;
- Distribute parental permission forms including data protection declarations (if necessary), and ensure that they are signed and returned;
- 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;
- Coordinating the completion of the student tracking forms and teacher tracking forms; and
- Ensuring that the test administrators return all testing materials after the testing session.

For CBA countries, the following additional steps had to be undertaken by the school coordinators:

- Running a hardware and software system check on computers intended to be used for the testing, using an ICCS Compatibility Check Program (on USB) and/or a check to confirm whether or not the test delivery system could be accessed online (for internet-based delivery). School coordinators were provided with the "Preparing computers for ICCS: Instructions for School Coordinators" manual to support with that task;
- Completing the CBA school resources survey (contained in the "Preparing Computers for ICCS" manual); and
- Arranging for any technical configuration and setup of computers in the school to run the test.

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, applying a test administration protocol standardized across countries (see section *Administering the ICCS assessment* for more details on the administration procedures).

7.3 Field operations resources

Manuals and documentation

The ICCS survey operations procedures manuals (SOP) 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 National Study Instruments** 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, administration of the instruments in schools, and the quality monitoring procedures.
- **Unit 5: Preparing for Scoring, Data Capture, and Parental Occupation Coding** described the processes around scoring procedures including recruiting and training scorers, manual data capture, parental occupations coding procedures, data processing and verification, and final data submission to IEA Hamburg.

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. In ICCS 2022 five different possible configurations of test administration modes in schools were possible:
 - o Paper-based: In case of solely paper-based administration in the country.
 - o Computer-based: In case of the computer-based administration in the country, without administering the bridging study.
 - o Computer-based AND paper-based (bridging study): In case of the computer-based administration in the country, including administering the bridging study. The respective school has been selected for the computer-based and paper-based administration.
 - o Computer-based only (bridging study): In case of the computer-based administration in the country, including administering the bridging study. The respective school has been selected for the computer-based administration but NOT for the paper-based.
 - o Paper-based only (bridging study): In case of the computer-based administration in the country, including administering the bridging study. The respective school has been selected ONLY for paper-based administration.

For each scenario a specific manual was provided that included guidelines and information for the respective administration mode (for the *computer-based* and the *computer-based only (bridging study)* configuration the same manuals were used).

- **Preparing computers for ICCS: Instructions for School Coordinators** described the tasks for testing and preparing computers for the computer-based ICCS assessment administration, including some technical troubleshooting scenarios.
- **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. In ICCS 2022 five different possible configurations of test administration modes in schools were possible (see the description of *The School Coordinator Manual* above). For each scenario a specific manual was provided that included guidelines and information for the respective administration mode.
- **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.
- **Supplement 1 to SOP Unit 5- Assessment Master™: Monitoring Participation and Exporting Data Availability Status for IEA WinW3S** provided instructions for national monitoring of the computer-

based data collection. It described (i) how to export information from the RM Assessment Master™ for import into the IEA WinW3S and (ii) how to monitor computer-based data collection within the RM Assessment Master™.

- **Assessment Master™ Manual** provided technical information on how to use the RM Australia Assessment Master™ throughout the process of preparing national instruments (translation and translation verification, prepare instruments for paper- and computer-based administration and administer the instruments in the CBA platform).
- **Supplement 1 to Assessment Master™ Manual – Exporting PDFs** described how to generate and edit the instruments and export them as PDFs.
- **Main Survey Scoring Software Manual** provided information about how to conduct the online scoring of constructed-response items using the AM Marker component of the Assessment Master™ system.
- **Scoring Guides for Constructed-response Items** provided detailed and explicit guidance on scoring each constructed-response item.

Software

To digitally deliver the student test, as well as the student, teacher, and school questionnaires, ICCS 2022 made use of the Assessment Master™, an assessment platform owned by RM Results (see Chapter 4 for further details). National center staff could also use Assessment Master™ to monitor the upload of data. Countries also were provided with compatibility check software for checking the suitability of computer devices for ICCS 2022 for the intended delivery mode.

In addition to the computer-based delivery software and consistent with the goal of automating and streamlining procedures wherever possible, IEA Hamburg provided NRCs with custom-built software packages to assist with sampling and data collection for PBA:

- **The Windows® Within-school Sampling Software (WinW3S)** 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 paper-based test booklets and questionnaires.
- **The IEA Data Management Expert (DME)** 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 2022 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 questionnaires.

In addition to the software and manuals, IEA conducted a field operations training seminar in advance of both the field trial and main survey. These seminars were designed to train national center staff on all procedures and familiarize them with the supporting software, namely the IEA WinW3S, and IEA Data Management Expert software.

7.4 Field operations 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 processes, and provided the information needed to compute sampling weights.

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

- **School sampling frame:** The IEA Sampling Team provided national centers with an Excel file, listing the sampled schools and their replacements, the school identification code, a measure of school size, and stratification information. The file was used to track school participation.

- **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 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 (an option allowing more precise estimates for 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 7.1), 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 teacher within each sampled 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 7.1: Hierarchical identification (ID) system

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

Contacting schools and sampling classes

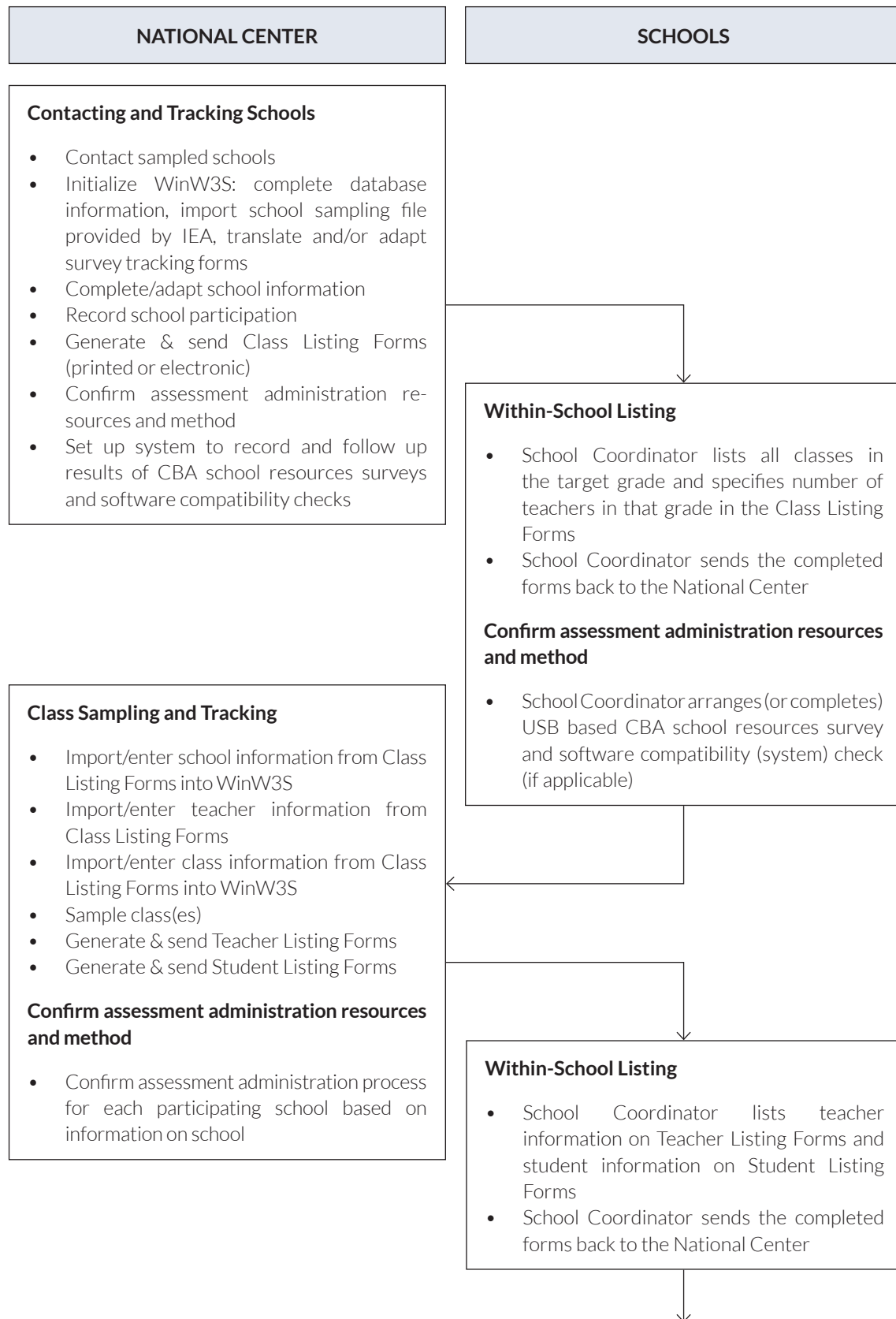
Once NRCs had obtained a list of the schools that had been sampled for ICCS, (for more information on all sampling procedures, please refer to Chapter 6), it was important for the ongoing success of the study that good working relationships with these schools were established. NRCs were responsible for 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 school coordinators for all participating schools (see section Field operations personnel of this Chapter for details).

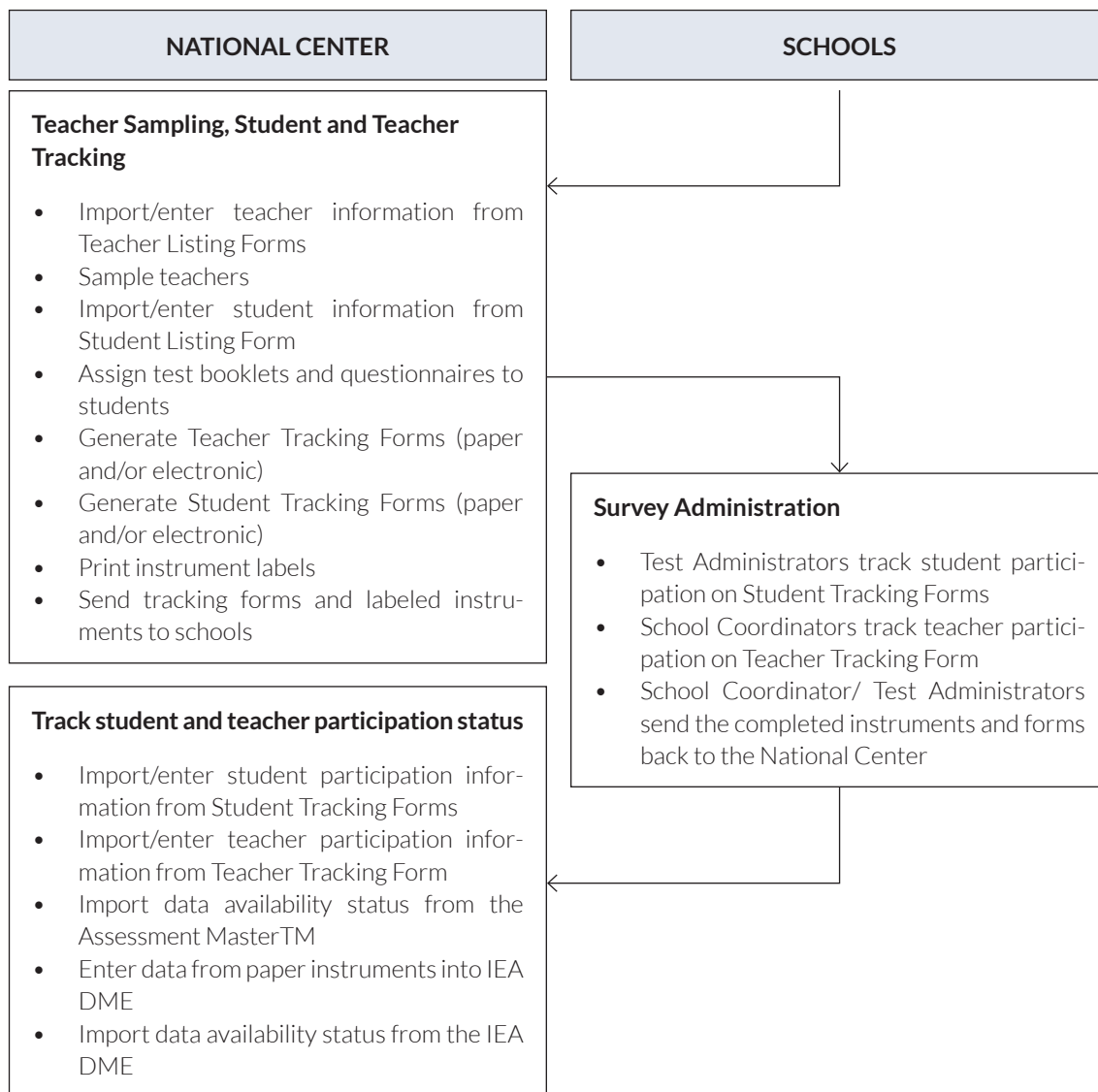
National centers sent a class listing form to each school coordinator and asked them to provide information on all the eligible target grade classes in the school. Using this information, the national centers sampled classes within the schools. Intact classes had to be sampled, ensuring that classes are mutually exclusive and exhaustive groups of students, that is, every eligible student in the school was allocated to exactly one, and only one class (course). This requirement formed the basis 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.²

The national centers worked with schools to sample classes, track schools, teachers, and students, and prepare for test administration (Figure 7.1).

² 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 2022 school coordinator manual).

Figure 7.1: Procedures for working with schools to prepare for test administration





Notes: WinW3S = Windows® Within-school Sampling Software, DME = Data Management Expert

Although all students enrolled in the sampled classes were part of the target population, ICCS recognized that students with a functional or intellectual disability, or students with little experience of using the target language may be excluded from participation. Accordingly, the sampling guidelines allowed the exclusion of students (i) with severe disabilities, unable to follow even the general instructions of the test or making it impossible to perform the ICCS testing situation, and (ii) non-native language speakers who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier of the test. National centers were asked to translate these guidelines to their contexts, in order to enable school coordinators to apply the definitions in a standardized way in all countries. 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

ICCS 2022 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. These included adapting the instruments to the national context and translating all instruments to the language(s) used for administration.

As outlined in Chapter 5, 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.

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 5 for details). Following the translation verification and revision of the instruments, the instruments were finalized for administration.

For the PBA 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.

The CBA option allowed countries to administer all ICCS instruments in digital format, entailing the creation of so-called “players” as applications to administer the ICCS student instruments. The computer-based instruments for students could be administered in three different ways: (i) on USB sticks; (ii) online using a common internet browser; and (iii) via a local network server. For the school and teacher questionnaires, participating countries could choose to administer these online instead of using paper-based questionnaires.

To ensure the comparability of the data from the computer-based and online mode, only those countries that had field trialed it were allowed to use the computer-based/online option during the main survey. Eighteen countries and benchmark participants administered the computer-based student instruments for the ICCS main survey: Bulgaria, Brazil, Chinese Taipei, Colombia, Cyprus, Estonia, France, North Rhine-Westphalia (Germany), Schleswig-Holstein (Germany), Italy, Lithuania, Latvia, Malta, Norway, Slovak Republic, Slovenia, Spain, and Sweden. Twenty countries administered school and teacher questionnaires online during the ICCS main survey: Bulgaria, Brazil, Chinese Taipei, Colombia, Croatia, Cyprus, Estonia, France, North Rhine-Westphalia (Germany), Schleswig-Holstein (Germany), Italy, Lithuania, Latvia, Malta, Norway, Poland, Slovak Republic, Slovenia, Spain, and Sweden.

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.

After the instruments had gone through translation and layout verification processes, they were prepared for delivery online using the RM Results Assessment Master™ software, as described in more detail in Chapter 5.

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.

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 operations procedures. NRCs were asked to complete these questionnaires with the assistance of their national data managers and other 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. Chapter 8 presents insights derived from responses to the SAQ.

The country-specific login information was sent to the NRCs via e-mail. Respondents could log in and out as many times as needed during the administration period 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 screen, and respondents could change any answer at any time before submitting the responses to IEA.

Administering the ICCS instruments

Preparing assessment materials and distributing them to the participating schools required careful organization and planning by national centers. The assessment materials were provided to the school coordinators prior to testing, giving ample time for the school coordinators to confirm the receipt and correctness of the materials.

For the PBA, school coordinators then distributed school questionnaire and teacher questionnaires to the designated individuals. Teacher questionnaires were assigned and sent to each sampled teacher, and a school questionnaire was sent for completion to the school principal. Paper instruments for students were kept in a secure room until the assessment date.

For the online and CBA, individual login information was provided to ensure confidentiality and separation. For the school and teacher questionnaires, the national centers sent the login information, along with general information on how to access the online questionnaire, to the school coordinator in the form of “cover letters.” In line with the procedures used during distribution of the paper questionnaires, the school coordinator delivered the school and teacher login information to the designated individuals. Student login information was provided to the school coordinator with the student tracking forms and kept in a secure room before they were provided to the Test Administrators.

For the school and teacher online questionnaires, respondents could log in and out as many times as needed during the administration period 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 screen, and respondents could change any answer at any time before submitting the questionnaire.

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 civic knowledge test along with the international student questionnaires and a regional instrument (where applicable) to students. This person was chosen and trained by the national center. In some cases, the school coordinator also undertook the test administrator role. The test administrator was responsible for logging in students to the test system or distributing materials to the appropriate students, reading the instructions provided in the Test Administrator manual to the students, leading students through the assessment, and assuring 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 student ICCS instruments consisted of either two or three parts. The first part concerned the civic knowledge test booklets. 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 (see Chapter 2). Each student also was asked to complete a questionnaire. If a country opted for administering one of the two additional regional student questionnaires, each student was also assigned an additional regional instrument. In the case of PBA, questionnaires were labeled in a way that linked them to the corresponding test booklet. In the case of CBA, students were automatically directed to the corresponding questionnaire(s).

The time allotted for each of these sections was standardized across countries. To complete all parts of the civic knowledge test, target grade students were allowed 45 minutes in total. If students completed the test before the allotted time was over, they were allowed to review their answers or read quietly, but were 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 a 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 7.2).

Table 7.2: Time allowed for administering the ICCS student instruments

Activities	Length
Preparing students, reading instructions and distribution of Student Test Booklets (PBA) / logging students in to the test system (CBA)	10 minutes (approx.) (PBA) / 20 minutes (approx.) (CBA)
Administering the Student Test Booklets	45 minutes (exact)
Short break	5 - 10 minutes
Administering the Student Questionnaire	40+ minutes
Break	10 - 15 minutes
Administering the European Student Questionnaire*	20+ minutes*
Administering the Latin American Student Questionnaire*	10+ minutes*
Collecting the assessment materials and ending the session	Approx. 5 minutes
Total	Approx. 2 - 2.5 hours

* Only countries participating in the European or Latin American Student Questionnaire option.

The test administrator used the student tracking form to distribute the booklets to the correct students or to log in students to the test system in case of CBA, and to document student participation. During the CBA, technical support was given by the national center, which, in turn, could contact the ISC at ACER if they were unable to solve the problem themselves.

Because the national centers were able to monitor the responses to the computer-based and online questionnaires in real-time, they could send reminders to those schools where people had not responded within 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. The school coordinator used the information on the participation status to calculate the participation rate. If participation was below 90 percent in any class, it was the school coordinator's responsibility to conduct 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.

Countries delivering the instruments using computer-based delivery in ICCS had the benefit of a reduction in workload after data collection. Answers to instruments were already in electronic format, and responses were stored on central servers hosted by RM Results, thus there was no need for any data entry of the survey responses.

7.5 Scoring the ICCS assessment and checking scorer reliability

Scoring the assessment

The success of assessments containing constructed-response items depends on the degree to which student responses are scored reliably. The number of items to be scored within a country depended on the delivery mode (CBA or PBA) and, for CBA countries, whether or not the country participated in the bridging study. As 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.

Scoring responses on the constructed-response items collected online was done online via the RM Assessment Master™ Marker application, which was configured to address IEA standards and reliability control procedures. Scoring the paper-based constructed-response items was completed within the test booklets, with scorers writing their scores in a designated box in each booklet. These scores were later manually entered into the ICCS database.

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 participants with their first opportunity to give extensive feedback on the scoring guides for the new ICCS 2022 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 new ICCS 2022 items in light of their experiences of scoring the items in the field trial, and the guides were revised accordingly (see Chapter 2 for more details).

The scoring guides for the trend items stemming from the 2009 and 2016 cycles of ICCS were not revised and were identical to those used in ICCS 2009 and ICCS 2016 respectively. Further detail of the development and revision of the ICCS Main Survey Scoring Guide for Open-ended Response Items is provided in Chapter 2.

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. In countries participating in the bridging study, it was strongly recommended that scorers score both computer-based and paper-based responses. The ISC also provided guidelines on how to train scorers to accurately and reliably score the constructed-response civic knowledge items.

Assuring and documenting scoring reliability

Assuring and 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 different scorers who independently scored a random sample of approximately 200 responses per item.

For the CBA mode, double scoring was managed automatically by the RMAM Marker system. Approximately 200 responses to each item (this was set as a percentage of the total number of responses per item) were provided to two different scorers. Scorers were unaware of whether any given response they scored had been allocated for double scoring.

For the PBA mode, booklets to be double-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 first scorer designated to a booklet would enter their scores on a separate sheet (not in the booklet) and the second scorer would enter their scores in the booklet.

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 11 outlines the adjudication process relating to inter-rater reliability of items.

7.6 Field trial procedures

The ICCS field trial was a smaller administration of the ICCS assessment; on average, approximately 1,000 students were tested in each participating country. 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 was conducted.

Data collected in the field trial were used to evaluate the psychometric properties of the test and questionnaire items to inform revision and item selection for the main survey data collection. The field trial also served the purpose of testing the ICCS survey operations procedures to identify and resolve any issues prior to the ICCS 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 survey data collection.

The field trial resulted in some small modifications to survey operations procedures and contributed significantly to the successful execution of ICCS. The field trial was planned to be conducted during October 2020 to April 2021, but due to the COVID-19 pandemic not all countries managed the administration in the planned administration period. The COVID-19 pandemic greatly disrupted the field trial data collection administration period, requiring that some countries postponed the administration period up to six months. These were unprecedented circumstances for the ICCS study.

Chapter 8:

Quality Assurance in ICCS 2022

Lauren Musu and Kateřina Hartmanová

To meet the highest standards for the International Civic and Citizenship Education Study (ICCS) 2022 data, considerable effort has been put into ensuring standardized materials and survey operations procedures. Quality assurance for ICCS 2022 encompassed internal mechanisms built into each stage of the data collection process.

The International Association for the Evaluation of Educational Achievement (IEA) administered a quality assurance program for ICCS 2022 to document data collection activities and verify that the standardized procedures were followed. The program consisted of two independent components: An international quality assurance program, managed by IEA, and a national quality assurance program, managed by each national center and based on the international guidelines and manuals provided.

8.1 International Quality Assurance Program

Overview

This section provides an overview of the international quality assurance program and the data collected for the program.

The international quality assurance program was implemented by independent international quality observers (IQOs) appointed and trained by 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 testing sessions at participating schools. A total of 15 school visits were required for countries using paper-based administration (PBA) or those countries who were new to the study and using computer-based administration (CBA). In countries who were transitioning from PBA in 2016 to CBA in 2022, IQOs were required to visit three additional schools (18 in total) to observe the paper bridge booklet administration.

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 (SCs) and Test Administrators (TAs) about their experiences coordinating/administering ICCS 2022. Due to the COVID-19 pandemic, there was one country where outside observers were not allowed to enter schools to physically observe the ICCS 2022 main survey administration. Instead, the IQO completed extended (online or phone) interviews with the SCs and TAs to obtain as much information as possible about the administration, including the structure of the testing session and any issues that may have occurred.

Prior to beginning their assignments, the IQOs were required to attend a training session conducted by IEA. The online training session was held over two days in January 2022. During the training, IQOs were introduced to the ICCS operation procedures, study instruments, and tasks associated with their role. IQOs were also provided with a manual detailing their responsibilities and the necessary materials for completing their tasks.

In most participating countries and benchmarking entities, the national research coordinators (NRCs) assisted IEA in finding qualified individuals to serve as an IQO by nominating two candidates for the role. There were also cases where the IEA recruited IQOs who had served the same role in previous IEA studies. The independence of the IQOs from national centers is a crucial aspect, and as such the nominated individuals could not be a member of the national center, a family member, or personal friend of the NRC. Often, the nominees were school inspectors, retired school teachers, or researchers. The IQO was required to be fluent in both English and the main language(s) spoken in the country. When necessary, the IQOs recruited assistants in order to effectively cover the geographic territory or the schedule of ICCS administration. For ICCS 2022, a total of 25 IQOs were trained across the 22 participating countries and two benchmarking

participants. In turn, the IQOs trained 32 assistants. Altogether, IQOs and their assistants observed 378 (258 CBA and 120 PBA) administration sessions.

The IQOs were required to:

- Collect the complete set of ICCS 2022 national instruments and manuals, the national adaptation form(s), and tracking forms for the selected schools from the national center;
- Review selected sections of the School Coordinator (SC) and Test Administrator (TA) Manuals to check consistency with the international version;
- Review the national assessment instruments and the international translation and layout verification feedback to document whether the verifiers' comments were addressed appropriately in the final national instruments;
- Select the schools and classrooms to be observed in consultation with the NRC and according to specified guidelines;
- Contact the SC of each selected school to organize the visit and arrange the interviews;
- 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 SC and TA and record the responses on the classroom observation record; and
- Submit all collected national materials and completed observation records to IEA.

Observing the ICCS Administration Sessions

IQOs were required to document their observations of testing sessions using standardized Classroom Observation Records. 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 and the Test Administrator.

In Section A of the classroom observation record, IQOs registered their observations of the condition of the survey materials and (for CBA) testing equipment. Section B addressed activities that took place during the actual administration sessions of the student cognitive test, international questionnaire, and regional questionnaire (if applicable). Regardless of administration mode, 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 prior to administration of the regional questionnaire.

Throughout, IQOs were asked to observe the activities of the TA, specifically the following:

- Distributing, collecting, and securing the test booklets and questionnaires (PBA);
- Setting up workstations and securing the computers (CBA);
- Following the ICCS 2022 administration script; and
- Making time announcements during the administration sessions.

IEA received documentation of the international quality assurance program from all 24 countries and benchmarking participants in ICCS 2022. The results of the IQO observations are reported in the following sections. For each section, combined results from any common questions between CBA and PBA are presented first, followed by the results from questions specific to each of the two administration modes.

8.2 Findings from the ICCS 2022 administration

Preliminary activities and condition suitability for the assessment

As mentioned previously, ICCS 2022 marked the transition from PBA to CBA, with most participating countries choosing to administer the student materials on computers and a small number of countries administering only on paper. In terms of quality assurance, this meant slightly different procedures and

preparations took place for each of the administration modes, as reflected in the questions IQOs received as part of the observation records. For example, while paper booklets were expected to be distributed at the beginning of each part of the administration, workstations for CBA were required to be prepared in advance with a single workstation to be used by each student throughout the whole administration.

For each of the administration modes, IQOs noted their observations of the condition of the survey materials, the testing equipment, the TA's level of preparation, and the suitability of the room where the administration took place.

In line with Table 8.1, the survey materials were usually stored in a secure location at the school, online (for CBA), or kept with the TA. There was only one incident reported by IQOs where the appropriate security measures were not followed and although stored securely, the materials were not sealed (PBA).

Table 8.1: Quality of ICCS 2022 administration preparation—Across all 378 observed ICCS administration sessions (percent of IQO responses, rounded to a whole number)

	Yes (%)	No (%)	Not Applicable ^a (%)	Not Answered (%)
Were the assessment materials safely stored and securely sealed? (safely and securely stored)	91	0	9	-
In your opinion, did the Test Administrator familiarize himself or herself with the script prior to the administration of the assessment instruments?	95	3	-	2
In your opinion, were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	96	4	-	-
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	95	5	-	-

^a As selected by IQOs, no further explanation given

For PBA, any discrepancies between the student identification information on the instruments and the student tracking form (as reported in Table 8.2) were attributed to typographical errors in the lists that were then corrected. In two of the visited classrooms, it was not possible to thoroughly check the correct paper material distribution, due to the privacy measures of the country, preventing the use of names to identify students.

Table 8.2: Quality of ICCS 2022 administration preparation—across PBA administration sessions (percent of IQO responses)

	Yes (%)	No (%)	Not Answered/Don't know (%)
Did the Test Administrator verify adequate supplies of the test booklets and questionnaires prior to the students' arrival?	96	3	1
Does the student identification information on the test booklets and student questionnaires correspond with the student tracking form?	97	3	1

^a Because of rounding some results may appear inconsistent. Applicable to all tables in this section.

In terms of equipment for the CBA administration, computers with screen sizes of at least 15 inches were used in most sessions (72%). These were most frequently desktop computers (55%), but portable computers with (30%) or without (15%) a power plug were also used. In the majority of the cases (82%), no other person besides the Test Administrator, School Coordinator, or IT support person had access to the assessment room prior to the testing (see Table 8.3). The people who were reported to have had access to the room in the other cases were mostly the class teacher or the school principal.

The workstations were specifically prepared for each student in most cases (72%) and in the instances they were not, the students picked their workstations and were given their assigned login details later. In one observed classroom, students used their own laptops to take the assessment using login details provided by the TA. Most of the already prepared login screens (86%) displayed the student ID and the remaining 14 percent displayed the student's name.

Table 8.3: Quality of ICCS 2022 administration preparation—across CBA administration sessions (percent of IQO responses).

	Yes (%)	No (%)	Not Answered (%)
Did the Test Administrator set up all workstations (either desktop or laptops) with each of them displaying the welcome screen prior to the students' arrival?	72	27	1
Is the Test Administrator ensuring that each student is sitting in front of the computer specially prepared for him or her?	91	8	0
Did anyone else, besides the Test Administrator, have access to the assessment room prior to the testing session? ^a			
School Coordinator	75		25
School computer network administrator or IT support person	62		38
Other	18		82

^a For the last question, the No and Not Answered categories are merged, indicating the percentage of respondents who did not tick a checkbox for the given option.

In general, the reported information confirmed that the preparations for ICCS administration were of a high standard (Tables 8.1, 8.2, and 8.3). As reported in Table 8.1, IQOs judged most TAs (95%) as being familiar with the test administration script. In general, the observers observed no procedural deviations in the test preparations that they deemed severe enough to jeopardize the integrity of the test administration. In addition, nearly all of the observed administration sessions took place under favorable room conditions that were suitable for students to work without distraction (96%). The most commonly reported inconveniences were the classroom being cold due to open windows (2%, often following COVID-19 measures), not enough space between students (1%), and external noise from other students (3%, especially during break time).

Administration of the Cognitive Test

Table 8.4 shows that IQOs reported that the assessments were conducted in accordance with the international procedures, particularly in regard to following the script and adherence to time limits. In very rare instances, the IQOs observed and carefully documented significant deviations from the administration procedures which will be noted throughout this section.

In the majority (96%) of cases, paper test booklets were distributed by the TA according to the booklet assignment on the Student Tracking Form and booklets label. In one case, the IQO reported that some booklets were distributed to the wrong students, but this was corrected before the assessment began as the students' names were on the booklets. In another case, students' names were called out and booklets handed to them, and in one observed classroom, two students distributed the test booklets based on the student names written on them. Additionally, in one country the privacy measures prevented the TA from writing the names of the students in the Student Tracking Form, resulting in them relying fully on the students to make sure to take booklets with the same number for each part of the assessment. Finally, in one school the test had been administered a week prior to the IQO's scheduled visit due to a miscommunication, so they completed an extended interview with the SC and TA since it was impossible for the IQO to observe this part of the assessment.

The average time spent on test instructions was 13 minutes across all of the observed sessions, ranging from 2 to 51 minutes (with the vast majority (92%) falling between 6–30 minutes and only 2% above). In about 60 percent of the sessions, the actual test administration duration was exactly equal to the allocated time (45 mins). In about 83 percent of the remaining sessions (about a quarter of all observed sessions), all students completed the test before the allotted time has elapsed. Consequently, the average test time was 43 minutes (two minutes under the time allocated). It is also important to note that for CBA the timer was built into the assessment making it impossible for students to exceed the allocated time. IQOs' records showed that in most of the observed sessions the break between test and questionnaire administration sessions did not exceed 10 minutes. In only 4 percent of the cases was there no break between sessions, usually at the request of students who preferred to continue.

Table 8.4: Test administration sessions –378 ICCS administration sessions (percent of IQO responses)

	Yes (%)	No (%)	Not Answered/ Not applicable (%)
Does the Test Administrator announce, “you have 10 minutes left” prior to the end of testing session?	89	11	0
Are there any other “time remaining” announcements made during the testing session?	23	76	1
Did the Test Administrator follow the test administration script in the ICCS 2022 Test Administrator Manual?	75	24	1
Did the Test Administrator address students’ questions appropriately?	94	1	5
Are the computers/test booklets (collected and) secured at the end of testing session?	89	7	4

In accordance with the ICCS procedures, TAs were asked to (collect and) secure the test booklets or computers at the end of the testing session. The IQOs reported that this occurred in 89 percent of the testing sessions. The majority of the reported exceptions were regarding desktop computers, as students either left the room during breaks or after the test was finished (so computers did not have to be secured), or students stayed by their computers under the TA’s supervision even during the break. There was only one reported case of the students taking the device with them on a break and one case of students being asked to leave the room which was then closed but not locked/supervised.

One of the most important methods of standardizing the assessment administration was to have all Test Administrators follow the script in the Test Administrator Manual. IQOs reported that in about three-quarters of the observations, the TAs exactly followed the script when preparing the students, distributing the materials, and giving directions and examples. In the cases where the TAs deviated from the script (92 total), nearly all modifications were reported as “minor” and characterized most frequently as additions (53 cases), followed by omissions (32 cases), and revisions (31 cases). These categories were not mutually exclusive and selecting multiple of them to characterize the changes (additions, revisions, omissions) was possible.

Table 8.5 presents observations on students’ compliance with instructions and overall cooperation during the assessment administration. According to the IQOs’ observations, in almost all the sessions, students complied very well or well with the instruction to stop work. In addition, IQOs described the students as extremely orderly and cooperative for more than two thirds of the testing sessions.

Table 8.5: Student cooperation during test administration—378 ICCS administration sessions (percent of IQO responses)

	Very Well (%)	Well (%)	Not well at all (%)	Not Answered (%)	
When the Test Administrator ends testing session, how well do the students comply with the instruction to stop work?	96	4	0	0	
	Extremely (%)	Moderately (%)	Somewhat (%)	Uncooperative (%)	Not Answered (%)
To what extent would you describe the students as orderly and cooperative?	68	29	3	1	-

Administration of the Student Questionnaires

Table 8.6 shows selected questions from IQOs’ observations of the Student Questionnaire administration. IQOs reported that the Student Questionnaires were distributed according to the Student Tracking Forms and questionnaire labels. A deviation from the procedure occurred in one country because the questionnaires were distributed together with the tests at the beginning of the administration and in another case, students distributed the assigned booklets.

In most cases (85%), TAs followed the Student Questionnaire administration script exactly. Additions (24 cases), revisions (18 cases), or deletions (27 cases) to the script were reported almost equally in those instances where changes occurred (53 cases in total, with multiple categories of changes to one script possible).

Table 8.6: Student Questionnaire administration—378 ICCS administration sessions (percent of IQO responses)

	Yes (%)	No (%)	Not Answered/ Not applicable (%)
Did the Test Administrator follow the questionnaire administration script in the ICCS 2022 Test Administrator Manual?	85	14	1

The Student Questionnaire was administered in exactly 40 minutes (full allocated time) in 31 percent of the sessions. As instructed in the TA Manual, students were able to ask for extra time to complete the questionnaire, and this took place in approximately one quarter of sessions (27%). The allotted extra time varied from a few minutes to 56 additional minutes. Administration of the questionnaire took 42 minutes on average (ranging between 15–96 minutes).

The IQOs also provided observations on the administration of the two regional instruments for ICCS 2022 (Table 8.7) as follows:

- European Student Questionnaire (18 participating countries and the 2 benchmarking participants);
- Latin American Student Questionnaire (2 participating countries).

Paper-based Regional Student Questionnaires were distributed according to the Student Tracking Forms and questionnaire labels in 80 percent of cases. Deviations from the procedure were reported in two countries: in one country, both questionnaires were distributed at the beginning of the ICCS administration and in another, students handed out the assigned questionnaires.

IQOs judged the Test Administrators to have followed the administration script exactly in 78 percent of the sessions. Changes were generally considered minor (in 20 out of 34 total cases) and were most commonly characterized as omissions (21 cases).

Table 8.7: Regional student Questionnaire Administration—378 ICCS administration sessions (percent of IQO responses)

	Yes (%)	No (%)	Not applicable/not answered (%)
Did the Test Administrator follow the regional questionnaire administration script in the ICCS 2022 Test Administrator Manual?	78	9	14

Summary Observations of the ICCS 2022 Administration

Section C of the classroom observation record included questions about general impressions of how the administration was conducted, how well the TA monitored students, and any unusual circumstances that arose during the session. Table 8.8 reports on these.

Overall, IQOs reported that the quality of ICCS administration was very good or excellent (90%). Instruments of high quality were delivered to schools, only one case of defective booklets was reported, and these were detected and properly replaced after the administration session began.

Table 8.8: General impressions on the quality of administration—378 ICCS administration sessions (percent of IQO responses)

	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered (%)
In general, how would you describe the overall quality of the administration sessions of ICCS 2022 assessments?	58	32	7	3	0	-

Student refusal to take the assessment was rare and only occurred in 4 percent of observed assessment sessions. There were no problems observed in relation to test refusals and no late students in the majority (88%) of observed sessions (see Table 8.9). Out of the students who did arrive late, most were still admitted into the room, either before (7%) or after (4%) the beginning of the administration. There were only four reported cases (1%) of late students not admitted into the room. In 19 percent of the observed administration

sessions, students left the room for an “emergency” (usually a bathroom visit) during the test administration. In such cases, TAs were instructed to secure the students’ computers or collect their test booklets until the students reentered the assessment room. In some sessions, the booklet was left closed on the student’s desk until the student returned to the room.

In 10 percent of the sessions there were students requiring special accommodation. Most commonly, students with dyslexia/reading difficulties had the questions read out to them in a separate room or used Optical Character Recognition (OCR) files. In addition, a few students with autism or other conditions had an assistant available to help them take the test.

While in most testing sessions students complied with the instructions to store everything away, sometimes cell phones or tablets were kept on the tables for listening to music after test completion and until the time for completion of the test or questionnaire(s) expired. In some other instances, books were left on the tables for students to read after finishing the test. In a few classrooms students also kept their computers on the tables, but they were turned off during the test (with the exception of a few students with dyslexia using them for the OCR files).

TAs were not walking around the room in a few of the observed sessions, but IQOs noted they were nonetheless able to monitor students from the front of the room. In 5 percent of the cases, IQOs reported evidence of students attempting to cheat. However, in some of these instances, the IQOs observed that students only tried to communicate between themselves until they found out that the tests were different. IQOs reported that TAs intervened when necessary in appropriate ways.

Table 8.9: General observations of the administration—378 ICCS administration sessions (percent of IQO responses)

	Yes (%)	No (%)	Not Applicable/ Not Answered (%)
Did any students refuse to take the test either prior to the testing or during the testing?	4	96	-
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student?	4	5	91
Were any late students admitted to the room?	11	89	-
Did any students leave the room for an “emergency” during the testing session?	19	80	1
If a student left the room for an “emergency,” did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	18	7	75
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	10	90	-
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?	90	10	-
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?	95	5	-
Did you see any evidence of students attempting to cheat on the test (e.g., by copying from a neighbor)?	5	95	-

Interview with the School Coordinator and Test Administrator

The IQOs conducted interviews with the SCs and TAs according to the guidelines included in Section D of the Classroom Observation Record. The purpose of this activity was to solicit an evaluation of 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;
- Organization of classes in the school (to validate within-school sampling procedure).

School Coordinators

A large majority of SCs (82%) expressed a favorable impression of the ICCS survey, reporting that the ICCS 2022 administration in their school went very well overall (Table 8.10). When SCs reported issues, they were primarily related to students' absenteeism, lack of time to prepare for the assessment, waiting time for those students who finished earlier, and technical issues (difficulty setting up USB sticks, spotty internet connection, test login details, or the link not working). Staff members' attitude toward the administration was reported as positive in 63 percent of the observed sessions. Although the majority of staff could see the importance of the survey, there were complaints in some schools; teachers considered the timing unfortunate (due to COVID-19 and other assessments happening at the same time), saw the assessment as an additional burden after the difficult period of the pandemic, and expressed concerns that the administration time took time away from lessons.

SCs were also asked about the quality of the *School Coordinator Manual*. Less than six percent said that the manual needed improvements and provided reasons such as the length, repetition, and lack of visual appeal. Some SCs also asked for a digital version to be provided and mentioned that additional visuals to show the login process and details on handling technical difficulties would have been helpful. It should also be noted that not all SCs had received or read the manual by the time of the administration.

Table 8.10: Interview with the School Coordinator, overview— (percent of SC responses)

	Very well, no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered (%)
Overall, how would you say the assessment administration went?	82	15	1	2
	Positive (%)	Neutral (%)	Negative (%)	Not Answered (%)
Overall, how would you rate the attitude of the other school staff members towards the ICCS 2022 assessment?	63	32	3	1
	Worked well (%)		Needs improvement (%)	Not Answered (%)
Overall, do you feel the ICCS 2022 School Coordinator Manual worked well or does it need improvement?	90		6	5

There were only a small number of cases where components were missing from the shipments of test materials as Tables 8.11 shows. The time available for checking the materials ranged from weeks to days. In most of the cases where materials were not checked by the SC, this happened because external TAs brought the ICCS materials to the school on the administration day. In two countries, the School Coordinator Manuals were reported as missing. In such cases, they were either provided later or replaced with training by the respective national centers.

Table 8.11: Interview with the School Coordinator, details (percent of SC responses)

	Yes (%)	No (%)	Not Answered / Not Applicable (%)
Prior to the testing day, did you check your shipment of materials from the National Study Center?	91	2	7
Did you receive the correct shipment of the materials as listed in your School Coordinator Manual and according to the tracking forms?	91	2	7
<i>If no, did the national center provide the missing materials in time for the testing?</i>	4	5	91
Was the national center responsive to your questions or concerns?	94	0	6
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the ICCS 2022 assessment?	60	38	2
Was the Teacher Questionnaire administered online?	67	21	9
<i>If the Teacher Questionnaire was administered online, did the teachers encounter any problems?</i>	9	53	37
Was the School Questionnaire administered online?	68	25	7
<i>If the School Questionnaire was administered online, did the person completing it encounter any problems?</i>	2	59	39
Is this a complete list of the classes in this grade in this school?	94	4	2
To the best of your knowledge, are there any students in this grade level who are NOT in any of these classes?	4	95	1
To the best of your knowledge, are there any students in this grade level in more than one of these classes?	-	98	2
If there was another international assessment, would you be willing to serve as a School Coordinator again?	88	10	2

In more than half of the cases (60%), SCs indicated that the principal, the civic and citizenship education teacher, or the students' classroom teacher gave students special instructions, motivational talks or incentives prior to administration. In a few countries, the students received leaflets or were shown a short video. In one country there was an article about ICCS published in the school newspaper.

Because the sampling of classes required a complete list of all classes in the school at the target grade, IQOs were also asked to verify that the class list really included all classes. The majority of the SCs (94%) confirmed that the complete list of classes had been documented. There were some schools where IQOs reported one or multiple missing classes on the list. In one country, the SC explained that the missing class was in line with the national center's instructions to exclude certain types of classes in line with the sampling frame. In another country, the class listing form had not been used for class selection but rather an official school registry provided by the Ministry of Education was used in its place.

The presence or absence of students from the initial student listing forms sent to the national centers could be explained due to recent changes in the school enrolment during the school year or students with severe learning disabilities who were excluded from the assessment. There were no reported cases of students enrolled in more than one of the classes.

As a reflection of the planning and implementation of ICCS 2022, 88 percent of interviewed SCs said that they would be willing to serve in the same role in future international assessments. Several remarked that the experience was enjoyable and positive overall. The main reasons mentioned for the lack of willingness to serve again were the timing of the test (during a busy end of the school year with other assessments happening in the same period) and that it required a lot of extra time and effort. Finally, it is notable that the response rate for the Classroom Observation Records, Section D, was extremely high on all questions. The only exceptions recorded were three SCs who were ill or unavailable for the interviews.

Test Administrators

In the 2022 cycle, IQOs also interviewed the Test Administrators for the first time. About 27 percent of TAs were teachers from a different class, 8 percent teachers from the same class, 3 percent school counsellors and 1 percent were principals at the school or school district. The rest (48%) fell into the “other” category and were mostly external staff trained by the national center or an external agency, administrative assistants (secretaries), IT personnel, or teachers from a different school district.

Similar to SCs, a large majority of TAs (82%) also expressed a favorable impression of the ICCS survey, reporting that the administration in their school went very well overall (Table 8.12). In addition, nearly all (98%) TAs reported they were satisfied with the set up for the administration of the assessment and questionnaire(s). As for the quality of the Test Administrator Manual, about 15 percent said that the manual needed improvements. Most commonly they mentioned that the manual was too long and repetitive (especially the instructions for students on how to behave and answer questions included in the script). TAs also suggested some structural changes such as making the necessary codes easily accessible in one place (on one page) or dividing the manual into sections (preparations before, on the day at the school, the administration itself, after the administration). They also suggested to emphasize in the script that all answers are anonymous, that the assessment is not graded and is only an investigation, and the implications of pressing the “Finish” button (no possibility to go back to review previous questions). In addition, the TAs would have appreciated if the manual contained more guidance on how to login and to deal with (regular) school breaks and FAQs from students.

Table 8.12: Interview with the Test Administrator, overview— (percent of TA responses)

	Very well no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered (%)
Overall, how would you say the assessment administration went?	82	18	0	-
	Worked well (%)		Needs improvement (%)	Not Answered (%)
Overall, do you feel the ICCS 2022 Test Administrator Manual worked well or does it need improvement?	83		15	2

As a reflection on the administration and related tasks, 94 percent of interviewed TAs said that they would be willing to serve again as a TA in future international assessments (Table 8.13). They often remarked that the manuals were easy to follow and things well organized, however they also mentioned that the role is time-consuming, and a different time of the school year would be better to administer the survey. Although some questions were left unanswered, there were no TAs refusing to take part in the interview.

Table 8.13: Interview with the Test Administrator, details (percent of TA responses)

	Yes (%)	No (%)	Not Answered/ Not applicable (%)
Were you satisfied with the set up for the administration of the assessment and questionnaire(s)?	98	1	0
Did you receive the Test Administrator Manual?	99	1	0
Did you receive the Test Administrator summary information sheet?	79	12	9
Did you receive the Test administration form?	98	1	1
Did you receive the Student tracking form?	99	1	0
Did you receive the USB sticks (if applicable)?	35	9	56
Did you list any problems encountered during the ICCS administration on the Test Administration Form?	19	81	0
If there was another international assessment, would you be willing to serve as a Test Administrator again?	94	6	0

8.3 Findings from the national materials review

As part of their tasks, IQOs reviewed selected sections of the national versions of the School Coordinator and Test Administrator Manuals, to ensure the content was comparable to the international versions. In general, any modifications reported by the IQOs were minor. For example, a sentence specifying that no exclusion code should be assigned to students who lack parental permission to participate was omitted in the School Coordinator Manual in multiple countries. While the text specifying different exclusion codes for students was often rephrased, the overall meaning was retained, and the information remained comparable to the source. Similarly, there were no significant deviations from the international version found in the national Test Administrator Manuals. In particular, IQOs reviewed the administration script provided in the Test Administrator Manual. In a few cases, the instruction to put away any schoolbooks or papers on the desk was omitted or adjusted to include other objects, such as mobile phones. Additionally, minor adjustments were made to other instructions in some cases to reflect a specific response mode or classroom reality.

In addition, IQOs also reviewed the assessment instruments to ensure that the results of the international translation and layout verifications documented in the national adaptation forms had been properly addressed and integrated into the final assessment instruments. There were no significant issues reported in this review.

8.4 Findings from the survey activities questionnaire

As another valuable source of information on the overall quality of the ICCS administration, the survey activities questionnaire (SAQ) gathered information from the NRCs about whether the implementation of the ICCS procedures accorded with the standards outlined in the survey operations procedures manuals. The SAQ also collected feedback on the strengths and weaknesses of the approaches and materials used. It covered various 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;
- Entering and coding occupation data;
- Entering and submitting data;
- Other experiences.

One major purpose of the SAQ is 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 20 of the participating countries (all of those who filled the survey out within the specified deadline).

Contacting Schools and Recruiting School Coordinators

The first part of the SAQ collected information regarding school recruitment. Table 8.14 provides the NRCs' responses to questions about school participation. Generally, countries used a variety of different materials to request school participation in the main survey, including sample letters provided in the *Survey Operations Procedures Manual* and letters from relevant ministries.

Table 8.14: Contacting schools and recruiting School Coordinators (Number of NRC responses, max 20)

	Yes
How did you contact the sampled schools?	
By phone	17
By regular mail	7
By email	20
Other	2
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)	5
Letters based on examples from other national projects	7
Other	8
Is school participation for ICCS 2022 in your country or benchmarking region compulsory?	
No, for none of the schools	15
Yes, for all schools	3
Yes, for some of the schools	2
Did you use incentives to convince schools to participate (for example, vouchers, money, books, lotteries, school-specific feedback)?	8

Participation was voluntary for all schools in 15 of the participating countries/benchmarking entities and for all private schools in one additional country. A number of NRCs reported difficulties in convincing the selected schools to participate (as shown in Table 8.15). Common reasons stated were logistical issues (COVID-19 related problems, timing of the study, available time of students and staff), concerns about overburdening students and teachers (including "survey fatigue" as Program for International Student Assessment (PISA) took place around the same time) and the lack of school level feedback after participation. Eight NRCs reported using incentives such as participation certificates, small monetary rewards or gifts, and school reports to further encourage participation.

Table 8.15: School recruitment challenges (Number of NRC responses, max 20)

Of the schools you found challenging to convince or that rejected participation, please specify how frequently each of following PRIMARY reasons was stated.	Always or almost always	Often	Rarely	Never	Not answered
Workload for the school staff is too high	6	6	3	2	3
Too time consuming for students	3	3	6	5	3
Too many surveys ("survey fatigue")	3	10	2	2	3
Reluctance to assess this learning area	0	0	4	13	3
Problems related to the COVID-19 pandemic	3	5	4	5	3

Nearly all SCs for ICCS were appointed by the school principals or selected within the participating schools and trained by the National Study Center (Table 8.16). They were trained mostly through a combination of telephone calls, email communication, online meetings, and written instruction (such as manuals and letters) designed to support them in their roles. Other means mentioned were a formal training session held and recorded to be later shared with those who could not attend in one country, and a special training video prepared for SCs in one other country.

Table 8.16: Recruiting School Coordinators (number of NRC responses, max 20)

	Yes	No/Not answered
How did you train the School Coordinators?		
Through formal face-to-face training sessions	1	19
Through telephone calls	14	6
Through email communication	16	4
Through online meetings	11	9
Through written instructions	16	4
Through other means (please specify below)	2	18
Did the School Coordinators report any difficulties with understanding any of the following aspects of the ICCS administration procedures?		
The student population definition	2	18
The teacher population definition	9	11
The necessity for listing all target grade classes	1	19
The necessity for listing all students of the sampled class	2	18
The necessity for listing all target grade teachers	3	17
Preparing for computer-based delivery, if applicable	0	20
The survey administration procedures	5	15
Other issues	1	19

Adapting and Translating the ICCS Assessment Materials

NRCs provided information about the process of adapting and translating the ICCS student cognitive test and questionnaires into national languages. Most NRCs identified national center staff as the individuals responsible for adapting and/or translating the survey instruments, as well as their review (Table 8.17). In about half of all participating countries, an outside specialist was involved in the translation in addition to the own staff.

When asked about their experiences with the external adaptation and translation verification processes organized by the International Study Center (ISC) and IEA (respectively), a small number of NRCs indicated that they had had problems, mainly with changes being made to the materials within the survey system without proper documentation. Fifteen of the countries started the adaptation and translation process by using the international version as the source and the vast majority (19) used their field trial translations as a starting point. Countries participating in the 2016 cycle also used the 2016 national versions of the instruments for items administered during that cycle. One country also made use of their 2009 translations where applicable.

Table 8.17: Adapting and translating the ICCS assessment materials (Number of NRC responses, max 20)

	Adaptation	Translation
Who adapted and translated the international version of the student test booklets?		
Own staff	20	16
Outside translator(s)	4	13
Who adapted and translated the international version of the student, teacher and school questionnaires?		
Own staff	20	17
Outside translator(s)	4	10
Who reviewed adaptations and translations of the assessment instruments?		
Own staff	20	20
Outside translator(s)	3	4
Did you experience difficulties with the adaptation and/or translation of the international version of the following instruments?		
Student test	5	6
Student questionnaire	5	6
Teacher questionnaire	4	5
School questionnaire	4	4
Regional student questionnaire (where applicable)	5	5
Did you have major problems regarding the processes of external verification for any of the instruments (test booklets and questionnaires)?	Yes	No
Adaptation verification	3	17
Translation verification	3	17
Did you have any problems with translating the School Coordinator Manual, the Test Administrator Manual, and/or the Scoring Guides?		
School Coordinator Manual	1	19
Test Administrator Manual	1	19
Scoring Guides	1	19
How did you adapt and translate the international version of the instruments (test booklets and questionnaires)?		
Directly from English source version in Assessment Master	15	
Directly from English source version in a Word document	10	
Directly from the translation provided by another national center	0	
Through revision of the 2022 field trial translations	19	
Using the 2016 national version for trend items	13	
Using the 2016 national version for content other than trend items	8	
Other (please specify below)	1	

Finalizing the ICCS Assessment Materials

Table 8.18 shows that countries completed the final step of the verification process—layout verification conducted by the ISC—with some difficulties, stemming from the system used for the study and unrequested changes introduced to the materials by the delivery system provider.

In general, NRCs did not find it difficult to assemble the various survey instruments. Few of them detected printing errors (missing pages and wrong page order). In all cases these errors were minor and were resolved in a straightforward manner before the distribution of the materials.

All external providers were asked to sign confidentiality agreements where applicable. Potential security breaches were reported in two cases, both of which involved materials getting lost before reaching the National Study Center after being (supposedly) shipped back from a participating school.

Table 8.18: Assembling and printing the ICCS assessment materials (Number of NRC responses, max 20)

	Not at all	A little	Somewhat	Very
How difficult was it to finalize the production of the test booklets and questionnaires?				
Student test	5	7	4	4
Student questionnaire	5	7	4	4
Teacher questionnaire	9	4	3	4
School questionnaire	9	4	3	4
Regional student questionnaire (where applicable)	6	5	3	4
	Yes	No	Does Not apply	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?	4	16	0	-
Did you detect any of the following problems during the printing process: low print quality, missing pages, page disorder, not intended upside down pages?	3	12	4	1
Did you discover any potential breaches of security (for example with printing, shipping, administration, submission to the national center)?	2	18	0	-
Did you print the survey materials in-house, or did you use an external provider?	In-House	External	Combination	Does not apply
Test Booklets	4	10	1	5
Questionnaires	4	10	1	5
Test Administrator and School Coordinator Manuals	11	5	2	2
Scoring Guides and Scoring Training Materials	14	3	0	3
National Quality Observer Manual and Classroom Observation Record forms	17	0	0	3
	Yes	No	Does not apply	Not Answered
If you used any external provider, did you ask them to sign any confidentiality agreement(s)?	11	0	9	-

Sampling

The summary information provided in Table 8.19 indicates that the sampling process worked well overall. In most countries, the NRCs reported no difficulties adapting the international sampling design to national specifications or compiling a sampling frame (a list of eligible schools). Among those national centers that experienced some level of difficulty, all felt well supported by IEA's sampling team. The Within-School Sampling Software (WinW3S) provided by IEA to select classes and teachers was used by all countries participating in the Main Study without difficulty.

Three NRCs reported encountering special conditions or organizational constraints that required deviations from the standard ICCS within-school sampling design. In one country, all classes of the target grade at a selected school were already sampled in other large scale international studies and thus could not be included in the ICCS sample. Other deviations stemmed from specific stream/track or special class characteristics of the educational systems making it necessary to adjust the design to ensure reasonable class selection. In each of these cases, a sampling expert was consulted to ensure that the altered design met all sampling requirements. In about two thirds (13) of the countries, numbers or other codes were used on student and/or teacher lists instead of or in addition to names.

Table 8.19: Sampling (Number of NRC responses, max 20)

	Not difficult at all	Somewhat difficult	Very difficult
How difficult was it to...			
adapt the international sampling design to your national specifications?	18	2	0
compile a list of all eligible schools?	19	1	0
select the samples with the IEA WinW3S software (if applicable)?	7	0	0
	Yes		No
Did you use the Within-School Sampling Software (WinW3S), provided by IEA, to sample classes and teachers?	20		0
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		17
Did you use other means instead of names to identify students and/or teachers on the forms and labels?	13		7

Administering the Main Survey

The section of the SAQ inquiring about main survey administration addressed some important aspects of the data collection; selected results are presented in Table 8.20. Some countries reported that at least some of their TAs were drawn from national center staff or contracted externally, but it was also common to recruit TAs from among the SCs or teachers, either of the sampled schools or from other schools.

As with the training of SCs, most countries used more than one approach to train their TAs, relying most heavily on written instructions (12 countries) supplemented with online meetings (12 countries) and email/telephone correspondence (12/11 countries, respectively). Formal face-to-face training sessions were conducted in only three of the participating countries. This large shift towards online trainings compared to the previous cycle (where 18 countries conducted formal in person sessions) is likely at least partially related to the COVID-19 pandemic.

Almost two thirds (9) of the schools administering paper-based instruments encountered problems with student participation. NRCs reported absences due to sickness, COVID-19 outbreaks, and lack of parental consent as the most common reasons for students not to participate. Make-up sessions were conducted when possible to reach high participation rates. Issues with teacher and principal participation were also reported in some circumstances.

Table 8.20: Administering the booklets and questionnaires (Number of NRC responses, max 20)

	Yes		No/ Not Answered		
Who were the Test Administrators in the main survey? (Check all that apply)					
National center staff	6		14		
Regional or district government staff	1		19		
External contractor staff	9		11		
Teachers from other schools	1		19		
Teachers from the sampled schools but not of the sampled students	11		9		
Teachers of the sampled students	6		14		
School Coordinators	11		9		
Other (please specify below)	4		16		
How did you train the Test Administrators? (Check all that apply)					
Through formal face-to-face training sessions	3		17		
Through telephone calls	11		9		
Through email communication	12		8		
Through online meetings	12		8		
Through written instructions	12		8		
Through other means (please specify below)	3		17		
	Less than 25 %	25-49 %	50-75 %	76-100 %	Does not apply
For the administration of (paper-based) instruments, approximately what proportion of schools returned the materials on time?	0	0	0	12	8
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?	Yes		No		Does not apply
Students	9		6		5
Teachers	2		3		15
School Principals	1		7		12

Preparing and Administering Online Questionnaires

As shown in Table 8.21, more than two thirds of countries elected to administer teacher and school questionnaires online. Half of the respondents reported experiencing major problems with the online platform (Assessment Master) during the administration and five reported issues with the online data collection of these questionnaires in general. The most common issues mentioned were downtime of the system resulting in users unable to access or being logged out of the system, changes not being saved properly, and difficulty monitoring the collected data in the system through reports. More than half (nine) of these respondents also reported problems with the login procedures, especially the URLs, checksums and IDs being confusing, mistyped or lost by the participants, or not working properly.

NRCs also reported difficulties reaching high participation rates for online questionnaires from school staff, mainly teachers. The explanations provided included lack of interest, time and motivation, high workload, or getting discouraged due to technical issues. NRCs also mentioned that there were notable differences between individual schools and that SCs can have a large impact on participation rates since they serve as the only channel to remind and encourage the respondents to participate. Relatedly, the most common means to organize follow-up procedures for teachers and principals that did not complete their questionnaire by the given deadline were contacting the SC through an email or a phone call, reminding the school principals, and sometimes extending the deadline.

Table 8.21: Preparing and administering Online Questionnaire (number of NRC responses, max 20)

	Yes	No
Did you use the online data collection mode for administering the School and the Teacher Questionnaires?	16	4
During the administration, were there any major problems concerning the Assessment Master web application? (where applicable)	8	8
During the administration, were there any problems concerning the login procedures (e.g., respondents who confused ID and checksum, problems entering the URL)? (where applicable)	9	7
Did you experience any difficulties with the online data collection for the teacher and/or principal survey? (where applicable)	5	11
For the Online administration, did you experience any difficulties to reach a high participation of teachers or school principals within the school? (Check one box in each row, where applicable)		
Teachers	8	8
School Principals	4	12

National Quality Assurance Program

IEA provided materials to assist national centers in conducting their own national quality assurance program. The ICCS 2022 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 2022 *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 SCs. Seventeen NRCs confirmed that they had implemented a national quality assurance program, and that NQOs and their assistants had visited 281 schools in the participating countries (Table 8.22). In three countries, NRCs did not conduct national programs because there were COVID-19 restrictions in place preventing physical school visits, they considered that information collected through the international program was sufficient, or because the TAs employed were considered very well trained.

In line with the IQO reports, the NQOs confirmed the quality of the survey process was generally good. There were issues reported in five cases, mostly errors with the testing system and data uploads, unruly/distracting behavior from students who finished the assessment early, or unsuitable seating arrangements. Additional comments highlighted that students sometimes felt the assessment was too long/too much to do in one day and often lost concentration, others additionally expressed the content was too difficult/hard to understand. In addition, there were some technical issues observed during administration, as previously mentioned.

Table 8.22: National quality assurance program (number of NRC responses, max 20)

	Yes	No
Did you implement a program of school visits by national quality observers?	17	3
Did you use the National Quality Observer (NQO) Manual and the Classroom Observation Record to conduct your national quality control program?	17	3
Did your national quality observers (NQOs) document any major problems or special circumstances that occurred frequently during the assessment administration?	5	15

Scoring Open-Ended Response Items and Coding Occupation Data

NRCs were also asked to comment on the persons responsible for scoring open-ended response items, the procedures used to verify scoring reliability, and the coding of occupation data (see Table 8.23). The participating countries used scorers with a variety of different backgrounds, including national center staff, teachers, university students or externals with degrees in relevant areas. Five NRCs reported experiencing difficulties with implementing the reliability (independent double scoring) procedure in their countries, often related to the scoring system or difficulty in finding consensus between scorers. These difficulties were generally resolved later on, sometimes with additional scoring practice.

The majority of countries (16) used the Occupation Data Excel file for coding occupations. Some PBA countries also mentioned coding occupations directly in the instruments. Countries used anywhere between 2–7 coders to code the occupation data. The coders were trained mostly through online and in-person sessions, where they were provided with concrete examples, and had the opportunity to ask questions. In some countries, coders also attended an online discussion group and/or received printed materials with instructions. In a few countries (two), no training was necessary as already experienced/trained coders were used. Four countries reported difficulties coding occupations, stemming mostly from the answers given by students being vague and not clearly referring to any of the ISCO-08 coding scheme definitions.

Table 8.23: Scoring open-ended response items and coding occupational data (number of NRC responses, max 20)

	Yes	No
Who scored your open-ended response items? (Check all that apply.)		
National center staff	12	8
Teachers/professional educators	9	11
University students	8	12
Other, please specify	4	16
Did you have any difficulties with the procedures for reliability scoring (double scoring) of open-ended response items?	5	15
Did you use the Occupation Data Excel file for coding occupations?	16	4
Did you have any difficulties with entering the occupation data to the Occupation Data Excel file? (where applicable)	1	15
Did you use any job classification scheme other than the ISCO-08 for coding occupations?	2	18
Did you translate the ISCO-08 coding scheme into your national language or use an already existing translation?	17	3
Did the coders of occupations have any experience in this area?	15	5
Did you have access to expert advice on occupational coding in your country?	7	13
Did you have any difficulties with coding the occupation data?	4	16

Entering, Collecting, and Submitting Data

As Table 8.24 shows, when asked about the persons responsible for entering paper questionnaire data, the NRCs reported using their own (national center) staff most often, followed by a combination of staff from the national center and an external data entry company. In some cases, data entry was done by university students or other external associates/assistants. Most countries used the IEA DME software to enter data manually. Those few that did not typically used scanning procedures or relied on an external company to carry out the data entry to a format that would later be imported into the IEA DME.

About two thirds (14) of the countries used the RM player for the student assessment. The most common form of delivery for CBA was online delivery (10), followed by USB (8), and portable computers taken to schools (7). Only two countries used a local server network to administer the assessment. More than half (10) of the countries administering CBA experienced difficulties using the RM player, although this is believed to only affect a small portion of the data. These difficulties included faulty USBs and technical issues with the online system (login issues, long loading times, buttons not working properly, too small font size, difficulty retrieving answers after accidental log outs). Some difficulties such as data not syncing up or other errors uploading the data from USB sticks or via a server method were encountered in five countries. Eight countries also reported difficulties monitoring the data collection, mentioning the data collection reports not working at the beginning of the data collection for a few weeks and/or only working using an Excel macro workaround.

Table 8.24: Entering and submitting data (number of NRC responses)

	Yes	No	No answer/ Not applicable
Did you use any paper-based versions of the Student, School and/or Teacher instruments?	15	5	0
Did you use your own staff or external staff to enter the data from the test booklets and questionnaires into computer files? (please mark only one choice, where applicable)			
Own staff	5	-	-
External data entry company	2	-	-
Combination of the above	3	-	-
Other	5	-	-
Did you use the IEA Data Management Expert (DME) to manually enter your data? (where applicable)	12	3	(5)
Did you use the RM player for the student assessment?	14	6	-
Which form of delivery was used for the computer-based assessment in your country?			
USB	8	6	6
Local network	2	11	7
Online	10	3	7
Portable computers taken to schools	7	7	6
Did you experience any difficulties using the RM Player during the assessment?	10	4	6
Did you experience any difficulties uploading the data from USB Sticks or via server method?	5	4	11
Did you experience any difficulties monitoring the data collection?	8	6	6

The NRCs' Assessment of the Quality of the ICCS Manuals and Study Resources

The survey activities questionnaire also provided an opportunity for NRCs to give feedback on the quality of the various ICCS manuals and study resources. The NRCs were generally very positive about the manuals and webinars, and nearly all of them described these resources as “very” or “somewhat” helpful for carrying out the survey (Table 8.25). When asked for their suggestions for improvement, some NRCs suggested that the information in the manuals could be more concise, extra webinar time dedicated to those new to the project, and that short accessible videos would be an appreciated additional resource.

In addition, most NRCs found the ICCS SharePoint website very or somewhat helpful, especially for exchanging data and materials and accessing meeting documents. The individual country schedules and country consultations were also perceived as useful for the preparations for the main survey and national material production.

Table 8.25: Other experiences (number of NRC responses, max 20)

	Very helpful	Somewhat helpful	A little helpful	Not helpful at all	Does not apply/not answered
Did you find the manuals and webinars helpful for carrying out the ICCS 2022 main survey?					
Survey Operations Procedures Unit 1	17	3	0	0	0
Survey Operations Procedures Unit 2	17	3	0	0	0
Survey Operations Procedures Unit 3	16	4	0	0	0
Survey Operations Procedures Unit 4	16	4	0	0	0
Survey Operations Procedures Unit 5	16	4	0	0	0
Assessment Master Manual	14	4	1	1	0
Assessment Master Webinar	11	3	2	2	2
School Coordinator Manual	15	4	1	0	0
Test Administrator Manual	15	4	1	0	0
Preparing computers for ICCS and USB compatibility test Manual(s)	10	2	1	1	6
National Quality Observer Manual	14	4	0	0	2
Scoring Guides	15	5	0	0	0
How helpful did you find the ICCS SharePoint?					
For exchanging contact information	14	4	2	0	0
For exchanging data and materials	19	0	1	0	0
For addressing scorer queries	10	6	2	2	0
For accessing meeting documents	19	1	0	0	0
	Very useful	Somewhat useful	A little useful	Not useful at all	
How useful did you find the Individual Country Schedule (Timeline Estimator) for the national instrument production?	10	9	1	0	0
How useful did you find the individual country consultations in preparation of the main survey?	14	5	1	0	0

8.5 Summary

The ICCS quality assurance programs conducted at both the international and national levels provided crucial documentation of participating countries' adherence to the standardized data collection procedures. The registered observations of the IQOs in the Classroom Observation Records indicate that in the vast majority of cases, the prescribed survey administration procedures were very closely followed. These included the preparatory activities, delivery of the administration script, timing of the sessions, and handling of the instruments.

The *Survey Activities Questionnaire* further documented a number of key aspects of the ICCS survey administration and provided information for improving the quality of future IEA surveys. Complementing

the IQO reports, the results of this questionnaire also show that NRCs generally felt well-supported by the study consortium and operational manuals, and that they were able to comply well with the international procedures related to sampling, external verification, and reliability scoring.

These results provide evidence of the high quality of the data collection sessions, despite being affected by various technical issues and COVID-19 related disruptions. The observed good behavior of most of the participating students, strong evaluations from national and international quality observers, School Coordinators, and Test Administrators, and positive attitudes of other school staff members toward the survey administration further attest to the careful planning and implementation of this cycle of ICCS.

Chapter 9:

Data management, cleaning, and creation of the ICCS international database

Hannah Kowolik and Yasin Afana

9.1 Introduction

This chapter describes the procedures for verifying the International Civic and Citizenship Education Study (ICCS 2022) data and creating the ICCS 2022 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), the associate research center at Lumsa University and the national centers of the participating countries.

Preparing the ICCS 2022 international database and ensuring its integrity was a challenging endeavor requiring extensive collaboration between IEA, the ISC, and the national centers of participating countries.

Participating countries followed standardized survey operations procedures to prepare instruments, administer the civic knowledge test and questionnaires, collect national context questionnaire data, and prepare data files (depending on the administration mode) and documentation. Once the team at IEA received the ICCS 2022 data and documentation, 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;
- All variables used for international comparisons were comparable across countries (after harmonization, where necessary); and
- All institutions involved in this process applied control measures throughout it to assure the quality and accuracy of the ICCS data.

9.2 Data sources

In ICCS 2022, participating education systems could choose between two different administration modes for the students’ assessment: computer-based administration (CBA) and paper-based administration (PBA). Countries that opted for the CBA option and that participated in ICCS 2016 were administering student instruments in paper format as part of the bridging study. For the school and teacher questionnaires, participating countries could choose to administer these online and/or paper-based.

Data collected as part of ICCS 2022 required work by participating countries before being submitted to IEA for processing and cleaning. This included data collected from: (i) the ICCS Player, which delivered the CBA, including student test and questionnaire; (ii) online administration, including school and teacher questionnaires (iii) paper instruments, including all instruments. This included the scoring of constructed response items, as well as verification procedures. While data from the computer-based and online administration were automatically stored on central servers, data from the PBA had to be manually entered into a national database.

Data required for processing and cleaning came from the following sources (depending on the administration mode):

- Data from the CBA for the student survey; and/or
- Data from the online data collection of school and teacher questionnaires; and/or
- IEA DME database including the response data for the PBA (student, school, and teacher);
- WinW3S database including participation, tracking, and sampling information.

Computer-based administration (CBA) for the student survey

As documented in Chapter 7 of this report, one major change was the decision to offer countries the option of a computer-based delivery for the student survey.

The CBA option allowed countries to administer the ICCS instruments in digital format and entailed the creation of players as applications to administer the ICCS instruments. The computer-based instruments could be administered in three different ways: (i) on USB sticks; (ii) online using a common internet browser; and (iii) via a local network server. Whereas web delivery ensured that data were immediately available on servers for further processing, USB delivery needed Test Administrators to upload the student response databases after the testing session.

To ensure confidentiality, student login information was provided to the Test Administrators with the student tracking forms. This login information corresponded to the ID and checksum provided by the IEA WinW3S, meaning that the identity validation step occurred when the respondents were logged-in to the survey.

Raw response data were saved on central servers hosted by RM Australia in native format (further information on how data were exactly stored is an intellectual property of RM Australia and cannot be described in more detail). Answers to constructed-response items were displayed in the AM Marker component of the Assessment Master™ system and assigned a numeric code by a human scorer (see details in Chapter 4 and Chapter 7). Answers to non-constructed-response items (e.g., multiple choice or drag and drop) were coded to numeric values. The data were transferred to flat files (CSV) for creating the data output suitable for import into the data processing system employed at IEA for all international large-scale assessments. Each row represented one respondent while each column represented a variable (i.e., a numeric representation of the answer). The produced data files matched the structure given in the international codebook. The data included all variables that held respondents' direct responses or scores (human scores or computer scores). Some additional variables holding information either from the login (e.g., School ID) or from the system itself (e.g., a unique code for the session) were included. One file per instrument and country was created. For countries administering the instruments in different languages, the data were merged into one file per instrument.

Online data collection of school and teacher questionnaires

Furthermore, ICCS offered an online collection of school and teacher questionnaire data instead of or in addition to using paper-based questionnaires.

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 by the IEA WinW3S, meaning that the identity validation step occurred when the respondents logged-in to the survey.

As respondents completed their online questionnaires, their data were automatically stored in central servers hosted by RM Australia and, therefore, no manual data entry was needed.

Raw response data were saved on central servers hosted by RM Australia in native format and then transferred to flat files (CSV) for creating the data output suitable for import into the data processing system employed at IEA for all international large-scale assessments (see the description of Computer-based administration (CBA) for the student survey above).

Data entry and verification of paper instruments

Each national center was responsible for entering 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 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 metadata information about the names, lengths, labels, valid ranges for continuous measures, 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 deferred 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 enrolment) 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 are 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 IEA DME were automatically validated. As each respondent ID number was entered, it was checked by the IEA 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 confirm the inconsistencies before resuming data entry. In addition, national centers were requested to perform several checks that identified inconsistent records across datasets (e.g., a student record was available in the student questionnaire, but not in the student test). 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.

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). 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. The IEA DME included a check to compare entries between different punchers and to list all discrepancies. 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 IEA, and it was found that in general the margin of error observed for processed data was well below the required threshold.

Data verification at the national centers

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

After data collection, the data manager or other staff at each national center used the information from the survey tracking forms to verify the completeness of the data. Participation information (e.g., whether a student, teacher or principal 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/or the uploaded data on the central servers.

For the paper-based administration, the national centres could export an Excel file from the IEA DME to verify paper data availability for each unique Student, Teacher, and School ID that was entered during manual data entry. However, for the computer-based administration, national centers could use a data monitoring tool that listed all records present on the upload server and allowed for downloading the list to be used for checking and updating the data availability status in WinW3S. The exported files were in turn imported into the IEA WinW3S and mapped against the list of Student, Teacher, and School IDs and reported participation statuses in the IEA WinW3S.

Data managers were requested to resolve all problems before the final data submission to IEA. If inconsistencies remained or the national center could not solve them, IEA asked the center to provide documentation on these problems. IEA used this documentation when processing the data at a later stage.

9.3 Data processing and data cleaning

Overview

As described earlier in this chapter, the data from the computer-based and online administration were transferred from central servers to CSV files for import into the data processing system employed at IEA. For the paper-based administration, the 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 IEA. In addition, each participating country had to submit the WinW3S database holding general information on respondents, as well as tracking and participation information.

Staff at IEA then subjected these data to a comprehensive process of checking and editing. To facilitate the data cleaning process, 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 7). 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 IEA and the NRCs. After each country had submitted its data and required documentation, 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. Context data cleaning (resolving inconsistencies in questionnaire data).

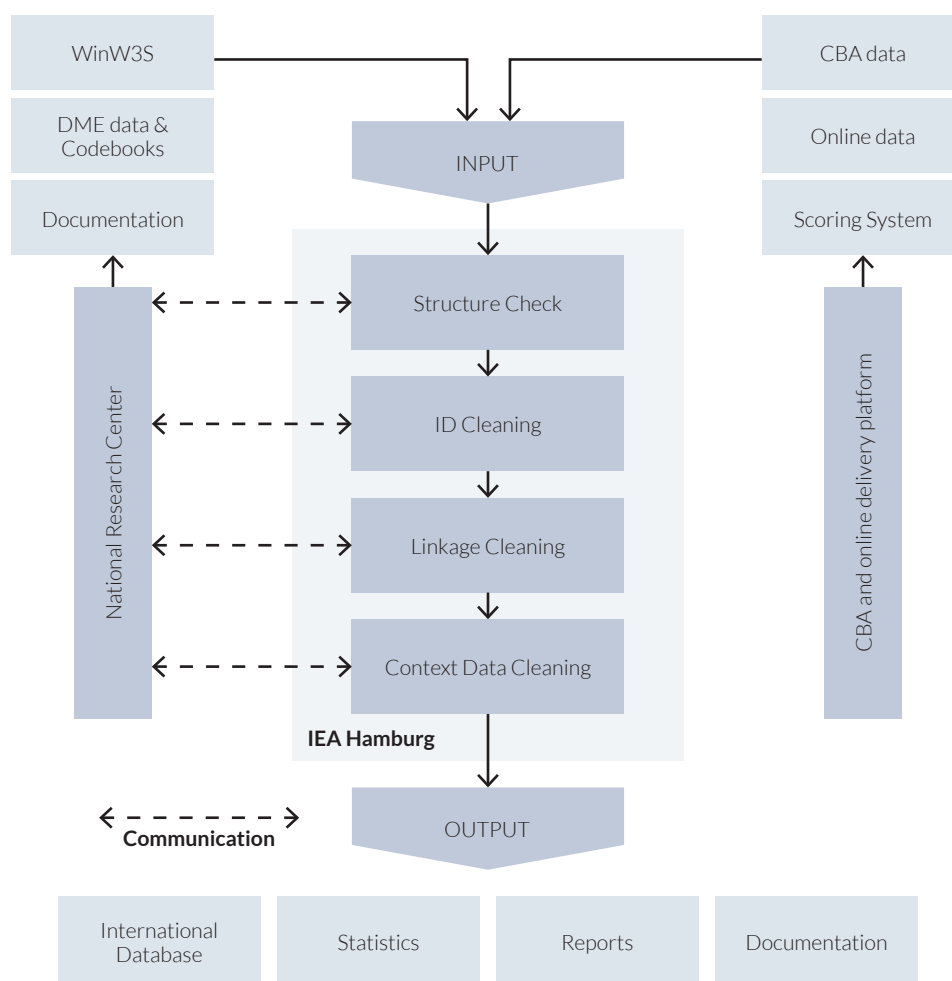
The cleaning process was an iterative process. Numerous iterations of the four-step cleaning procedure were completed on each national data set. This repetition ensured that all data were properly cleaned and that any new errors that could have been introduced during the data cleaning were rectified. 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 10 for details). Afterwards, the data, including sampling weights, were sent to the ICCS ISC so that scaling could be performed (see Chapter 11 and Chapter 12 for details). The NRCs were provided with interim data products to review at different points in the process.

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 accurately and consistently.

The program-based data cleaning consisted of a set of activities explained in the following subsections (Figure 9.1). IEA carried out all of these activities in close communication with the national centers.

Figure 9.1: 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 NAFs, student tracking forms, teacher tracking forms, and the survey activities questionnaire. IEA retrieved online and CBA data, including scored data from the scoring system, directly from the RM Assessment Master. National centers sent their WinW3S database, their DME database (if applicable) including scored data, and all required documentation to IEA. The fact that most centers sent this documentation greatly facilitated the data-cleaning process.

After the review, IEA first merged the tracking information and sampling information captured in IEA WinW3S database with the student-level databases containing the corresponding student instrument data from either paper- or computer-based assessment. 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, IEA asked the national centers to complete NAFs while they were adapting and translating the international version of the survey instruments. Where necessary, 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 2022 user guide for the international database (Afana et al., 2024).

IEA then discarded variables created purely for verification purposes 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 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.

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 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, 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.

Checking 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. Similarly, it was possible for a computer-based assessment login to be used (and responses uploaded) twice.

If two records in an ICCS database shared the same ID number and contained ‘exactly’ the same data, 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 more reliable or complete version of the 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 data files. 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.¹ These codes linked the students with their class and/or school membership, as well as teachers with their school.

Linkage cleaning consisted of several 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 were linked correctly with their corresponding schools. The

¹ 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 50120523 may be described as student 23 of class 05 in school 5012). The teacher ID includes the school ID (e.g., teacher 500103 may be described as teacher 3 in school 5001).

student tracking forms and teacher tracking forms were crucial in resolving any anomalies. 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.

Context data cleaning: 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. 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 21 in the school questionnaire asked for the total school enrollment (number of students) in all grades, while Question 22 asked for enrollment in the target grade only.

Clearly, the number given as a response to Question 22 could not possibly exceed the number provided by school principals in Question 21. IEA flagged inconsistencies of this kind and then asked the national centers to review these issues.

A filter question, which appeared in the teacher and the school questionnaires, directed respondents to a further question or section of the questionnaire. 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” 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 IEA recoded the filter question to “Yes.” This of course is only possible for dichotomous filter questions (e.g., with response options such as “Yes/No”).

IEA also applied what is known as split variable checks to questions where the answer was coded into several variables. For example, Question 30 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.

Two different sets of ICCS 2022 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, IEA queried the case with the national center, and the national center investigated which source of information was correct. If unresolved, questionnaire data were trusted over tracking data for students.

The teacher questionnaire did not ask teachers to provide birth year and month, but rather to choose between six age-ranges. The 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 was 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 2022 data capture: valid data values and missing data values. During data capture, missing data was assigned a value of “omitted,” “invalid,” or “not administered.” 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 very few cases, due to technical failure or 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** (only applicable for paper-based administration): 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. Following, 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’ (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 2022 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, IEA applied them to simulation datasets containing all possible problems and inconsistencies.
- **Registering all incoming data and documents in a specific database:** 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:** IEA recorded these in the ICCS main survey general cleaning documentation that was provided to NRCs and National Data Managers (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 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:** 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.

IEA compared national adaptations recorded in the documentation for the national dataset against the structure of the submitted national data files. Whenever possible, IEA recoded national deviations to match the international version to ensure consistency with the international data structure. However, if international comparability could not be guaranteed, IEA removed the corresponding data from the international database. IEA staff then documented any national adaptations made to the international version of the ICCS 2022 questionnaires in the ICCS 2022 user guide for the international database (Afana et al., 2024).

9.4 Interim data products

Building the ICCS 2022 international database was an iterative process. On completion of each major data-processing step, 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 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 (see details in Chapter 10). 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 (for more information, please refer to the ICCS 2022 user guide for the international database (Afana et al., 2024)). 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.

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 IEA; and a general cleaning documentation, describing in detail the data processing done at 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.

9.5 Final product: the ICCS 2022 international database

The ICCS international database incorporated all national data files from participating countries.

The data processing and cleaning effort implemented at IEA ensured that the ICCS 2022 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;
- 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 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 2022 user guide for the international database (Afana et al., 2024).

Reference

Afana, Y., Brese, F., Kowolik, H., Cortes, D., Schulz, W. (Eds.). (2024). *ICCS 2022 user guide for the international database*. International Association for the Evaluation of Educational Achievement (IEA).

Chapter 10:

Weighting procedures

Umut Atasever and Diego Cortes

10.1 Overview

A major objective of the International Civic and Citizenship Education Study (ICCS) is to derive and report accurate, precise, and internationally comparable estimates of population parameters for each participating country or benchmarking entity. To derive estimates with these characteristics, it is necessary to use complex estimators that account for several survey-design features. This chapter focuses on the role of estimation weights, which are the product of two features of our survey design. The first one is related to varying probabilities of inclusion across students or teachers, which makes probability weighting necessary to obtain unbiased estimates of population parameters. The second feature is related to weight adjustments that are necessary to obtain point estimates when the data collection process is subject to unit non-response. Adjustments are implemented under a non-response model that assumes that any unit non-response is ignorable within an adjustment cell. Since unit non-response plays an important role in the estimation of population parameters through weight adjustments, this chapter also discusses participation rates and outlines the different conditions under which estimates are reported in ICCS 2022.

This chapter begins by defining *school participation* in ICCS 2022. There is a long-standing policy implemented in ICCS that states that schools where only a minority of students or teachers participate in the study should not be considered as participating in the student or teacher study. This is applied because it is hard to justify that non-response is ignorable within a school when the participation rate is very low. Therefore, if a school had a participation rate of students or teachers lower than 50 percent, then the school was considered as not participating in the student or teacher survey.

The next three sections of the chapter contain a description of the sets of weights that were computed to ensure that results based on ICCS 2022 data resembled those in the underlying target populations. It describes how probability weights and adjustment weights were computed for participating students, classes, teachers, and schools. All reported estimates for ICCS 2022 were derived using weighted estimation and researchers conducting secondary analysis with this data 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 or benchmark entity achieved. The ICCS 2022 team regarded response rates as an important indicator of data quality. Although national centers made considerable efforts to ensure full participation, not all sampled units were included in the study. National samples were adjudicated accordingly with respect to sample participation requirements in the student and teacher surveys.

Finally, as described in Chapter 6, 11 countries transitioned into a computer-based assessment (CBA) in 2022. For these countries, an additional sample of students was surveyed in ICCS 2022 as part of the bridge study. Each of the processes described in the remainder of this chapter were implemented separately for both student surveys (core study and bridge study). This chapter does not distinguish between these two surveys when describing procedures in the text; however, the chapter does summarize and report statistics for ICCS 2022 and the ICCS 2022 bridge study separately.

10.2 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—provided that within-school weight adjustments are applied. Accordingly, ICCS defined a required minimum student participation rate within each school. This rate determined whether a school was considered as a “participant” in ICCS 2022.

In most participating countries or benchmark entities, only one class per school was selected for ICCS 2022. In this scenario, 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 eligible students had participated in the student study.

If a school did not meet this requirement, then all students within the school were regarded as non-participating.

In a small number of countries or benchmark entities, 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 eligible students participated.
- A sampled school was regarded as “participating school” if at least one sampled class was regarded as participating. If a school sampled two classes and one had a participation rate above 50 percent and the other below this threshold, then one class was considered as participating and the other as not participating.

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 2022 database.

Teacher survey participation requirements

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

- A school was regarded as a “participating school” in the teacher survey if at least 50 percent of its sampled teachers had participated.

If a school did not meet this requirement, it was regarded as a non-participant with respect to the teacher survey.

If the survey operation procedures in a school were not followed properly, the school was regarded as not 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, the teacher data from this school were not included in the ICCS 2022 database.

10.3 Weighting components for student survey

Student weights in ICCS 2022 databases are the product of several components, which could be grouped into two different types:

- *Base weights* reflect the inclusion probabilities of sampled units. At each stage of the sampling process, the base weight is the inverse of the inclusion probability of a sampled unit.
- *Non-response adjustments* reflect the probability of participation for each sampled unit, given that non-participation is *ignorable* within each adjustment cell (i.e., stratum or school). At each stage of the sampling process, non-response adjustments are the inverse of the probability of participation.

School base weight (WGTFAC1)

The first stage of sampling for ICCS 2022 involved selecting the schools in each country or benchmark entity. The school base weight reflects the selection probabilities of this sampling step. When explicit stratification was used, the school samples were selected independently in each explicit stratum h , with $h=1, \dots, H$, and H is the total number of strata in the country or benchmark entity. If no explicit strata were formed, the entire country was regarded as being one explicit stratum.

In most contexts, ICCS 2022 drew a systematic sample of schools with the selection probability of school i proportional to its 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. This was done to reduce the variability of school base weights across sampled schools.

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 .

School non-response adjustment (WGTADJ1S)

Given the fact that some schools refused to participate in ICCS 2022 or had to be removed from the international dataset, the school base weights had to be adjusted to account for the non-response. 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:

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

where $n_h^{s,e}$ is the number of sampled eligible schools and n_h^{p-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 2022. 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 2022. This could happen if a school had recently been closed, did not have any target grade students, or had enrolled only excluded students. In these cases, the ineligible school was not considered when calculating the non-response adjustment.

Class base weight (WGTFAC2S)

In each participating school, *Windows® Within-School Sampling Software* (WinW3S) was used to randomly select one or more classes (see Chapter 6). More specifically, this process involved a systematic random method with equal selection probabilities for each class. In this sampling step, the class base weight was 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 the number of sampled classes in school i in stratum h .

Class non-response adjustment (WGTADJ2S)

In most contexts, one class per school was selected to participate in ICCS 2022. 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. If two or more classes were sampled to participate per school and at least one did not participate, then a class weight adjustment was computed at the stratum level. The class weight adjustment, $WGTADJ2S_{hij}$, for each participating class j was calculated as:

$$WGTADJ2S_{hij} = \frac{C_h^s}{C_h^p}$$

where C_h^s is the total number of sampled classes and C_h^p is the total number of participating classes in explicit stratum h .

Student non-response adjustment (WGTADJ3S)

Since all students within sampled classes were selected to participate in ICCS 2022, all students with sampled classes had a probability of selection of one. The adjustment for non-response for student k , in school i and stratum h , $WGTADJ3S_{hijk}$, 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 adjustments, 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 classes had been sampled.

Student estimation weight (TOTWGTS)

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

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

10.4 Weighting components for teacher survey

School base weight (WGTFAC1)

Because ICCS 2022 included the same set of schools to participate in 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 for the student survey. 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 school non-response in the teacher survey, the school weight adjustment factor, $WGTADJ1T_{hi}$, 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 (see Chapter 6). 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. Under this scenario, those teachers that the national 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:

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

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:

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

where $t_{hi}^{s,e-cert}$ is the number of sampled eligible certainty teachers, t_{hi}^{p-cert} is the number of participating certainty teachers, $t_{hi}^{s,e-noncert}$ is the number of sampled eligible 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 computed as:

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

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 2022, 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 2022 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. 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.

Teacher estimation weight (TOTWGTT)

The teacher estimation weight of each teacher l of school i in stratum h , $TOTWGTT_{hil}$, was the product of the five teacher-level weight components:

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

10.5 Weighting components for school survey

ICCS was designed as a student and teacher survey, but not specifically as a school survey. Any statements about school-level variables must be treated cautiously because they can be subject to substantial sampling errors. However, school weights were calculated for ICCS 2022, and included in the international database, to also allow weighted analyses of data from school questionnaires.¹

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 there were some schools where their school principals or delegates had not completed the school questionnaire, but that had participated in the student or the teacher survey. Consequently, there could be schools which were regarded as participants for the student or teacher survey but as non-participants in the survey of school principals. 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 .

School estimation weight

The estimation weight of each school i in stratum h , $TOTWGTC_{hil}$, was the product of the two weight components:

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

¹ In the ICCS 2022 international report, all school level results were reported at the level of students (see Schulz et al., 2024).

10.6 Calculating participation rates

For ICCS 2022, 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. In addition, 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 was calculated as:

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

The unweighted school participation rate in the student survey after replacement was computed as:

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

Similarly, let cop denote the set of eligible and participating classes in all participating schools (i.e., including replacement schools), let cfp denote the full set of eligible classes in participating schools, and let np denote the set of eligible but non-participating classes in participating schools. In addition, let n^{cop} , n^{cfp} and n^{cnp} denote the numbers of classes in each of the respective sets. The unweighted class participation rate in the student survey was calculated as:

$$UPRS_{classes} = \frac{n^{cop}}{n^{cfp} + n^{cnp}}$$

Finally, 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, $UPRS_{students}$, was 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, $UPRS_{overall\_AR}$ was then given by:

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

Table 10.1: Unweighted participation rates in the student survey (core)

Country	School participation rate (%)		Class participation rate (%)	Student participation rate (%)	Overall participation rate (%)	
	Before replacement	After replacement			Before replacement	After replacement
Bulgaria	84.6	96.6	99.3	91.3	76.7	87.6
Brazil	73.1	86.8	96.3	79	55.6	66
Colombia	92.7	98.7	100	92.7	85.9	91.5
Chinese Taipei	97.3	98.7	100	93.6	91.1	92.4
Croatia	92.7	94	99.4	85.3	78.6	79.7
Cyprus	96.9	96.9	100	91.2	88.4	88.4
Denmark	29.6	62	98.4	88.9	25.9	54.3
Estonia	83.4	83.4	100	88.8	74.1	74.1
France	97.3	98	100	93.4	90.9	91.6
Italy	97.5	98.8	100	92.4	90.1	91.2
Lithuania	98.3	98.3	100	90.7	89.2	89.2
Latvia	93.4	96.7	100	88.6	82.8	85.7
Malta	100	100	100	89.3	89.3	89.3
Netherlands	57.7	83.2	100	89.5	51.6	74.5
Norway (9)	96.6	96.6	99.3	88.6	84.9	84.9
Poland	91.2	99.4	99.6	91	82.6	90.1
Romania	82.9	97.5	100	93.5	77.5	91.1
Serbia	88.7	92	98.6	88.1	77	79.9
Slovak Republic	91.9	96.2	100	93.2	85.6	89.7
Slovenia	90.2	96.6	100	94.9	85.6	91.6
Spain	96.9	98.1	100	90.2	87.4	88.6
Sweden	94.2	96.8	100	89.8	84.6	86.9
Benchmarking participant						
North Rhine-Westphalia	89.5	95.4	99.4	88.2	78.4	83.6
Schleswig-Holstein	57	73.7	98.8	80	45.1	58.3

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

Table 10.2: Unweighted participation rates in the student survey (bridge)

Country	School participation rate (%)		Class participation rate (%)	Student participation rate (%)	Overall participation rate (%)	
	Before replacement	After replacement			Before replacement	After replacement
Bulgaria	65.6	85.9	100	94.3	61.9	81.1
Colombia	95.8	97.9	100	93.1	89.2	91.1
Chinese Taipei	98.2	100	100	94.1	92.4	94.1
Estonia	88.7	88.7	100	86.2	76.4	76.4
Italy	94.4	100	100	97.3	91.9	97.3
Lithuania	100	100	100	91.7	91.7	91.7
Latvia	90.4	92.3	100	89.7	81.1	82.8
Malta	100	100	100	92.3	92.3	92.3
Norway (9)	98	98	98.9	89.6	86.8	86.8
Slovenia	89.7	93.6	100	95.4	85.6	89.2
Sweden	91.3	95.7	100	89.3	81.5	85.4

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

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 :

$$WPRS_{schools\_BR} = \frac{\sum_h \sum_{i \in op} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2_{hij} \times WGT ADJ2S_{hij} \times WGT ADJ3S_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT ADJ1S_{hi} \times WGT FAC2_{hij} \times WGT ADJ2S_{hij} \times WGT ADJ3S_{hijk}}$$

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

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

$$WPRS_{schools\_AR} = \frac{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2_{hij} \times WGT ADJ2S_{hij} \times WGT ADJ3S_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT ADJ1S_{hi} \times WGT FAC2_{hij} \times WGT ADJ2S_{hij} \times WGT ADJ3S_{hijk}}$$

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

$$WPRS_{classes} = \frac{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2_{hij} \times WGT ADJ3S_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2_{hij} \times WGT ADJ2S_{hij} \times WGT ADJ3S_{hijk}}$$

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

$$WPRS_{students} = \frac{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2_{hij} \times 1_{hijk}}{\sum_h \sum_{i \in fp} \sum_j \sum_{k \in sfp} WGT FAC1_{hi} \times WGT FAC2_{hij} \times WGT ADJ3S_{hijk}}$$

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:

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

Table 10.3: Weighted participation rates in the student survey (core)

Country	School participation rate (%)		Class participation rate (%)	Student participation rate (%)	Overall participation rate (%)	
	Before replacement	After replacement			Before replacement	After replacement
Bulgaria	84.7	96.5	99.3	91.3	76.8	87.5
Brazil	74.6	87.3	96.3	80.8	58	67.9
Colombia	92.7	98.4	100	92.8	86	91.3
Chinese Taipei	97.4	98.7	100	93.8	91.4	92.6
Croatia	92.5	93.9	99.4	86.6	79.6	80.8
Cyprus	97.4	97.4	100	91.5	89.1	89.1
Denmark	30.3	62	98.4	89.4	26.7	54.6
Estonia	83.4	83.4	100	89.5	74.6	74.6
France	97.3	98	100	93.3	90.8	91.4
Italy	97.5	98.8	100	92.2	89.9	91.1
Lithuania	97.8	97.8	100	91	89	89
Latvia	93.8	96.7	100	88.6	83.1	85.7
Malta	100	100	100	89.4	89.4	89.4
Netherlands	58.4	82.9	100	89.2	52.1	74
Norway (9)	98.1	98.1	99.1	88.6	86.1	86.1
Poland	90.8	99.4	99.6	91.3	82.6	90.4
Romania	83.6	98.5	100	92.6	77.4	91.2
Serbia	88.9	91.9	98.8	88.5	77.7	80.3
Slovak Republic	93.5	96.8	100	93.5	87.4	90.5
Slovenia	90.3	96.6	100	94.9	85.6	91.6
Spain	96.8	98.1	100	90.3	87.4	88.6
Sweden	93.9	96.8	100	90	84.5	87.2
Benchmarking participant						
North Rhine-Westphalia	89.3	95.4	99.8	89.2	79.5	84.9
Schleswig-Holstein	57.6	76.1	99.9	80.3	46.2	61.1

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

Table 10.4: Weighted participation rates in the student survey (bridge)

Country	School participation rate (%)		Class participation rate (%)	Student participation rate (%)	Overall participation rate (%)	
	Before replacement	After replacement			Before replacement	After replacement
Bulgaria	64.9	85.8	100	94.1	61.1	80.8
Colombia	97.3	99.5	100	92.8	90.3	92.3
Chinese Taipei	98.4	100	100	94.3	92.8	94.3
Estonia	86.6	86.6	100	85.9	74.4	74.4
Italy	94.5	100	100	97.2	91.9	97.2
Lithuania	100	100	100	92.4	92.4	92.4
Latvia	90	92	100	89.8	80.9	82.6
Malta	100	100	100	92.2	92.2	92.2
Norway (9)	97.9	97.9	98.9	90	87.1	87.1
Slovenia	89.7	93.6	100	95.3	85.6	89.2
Sweden	91.3	95.7	100	89.2	81.4	85.3

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

Unweighted participation rates in the teacher survey

To calculate unweighted participation rates in the teacher survey, 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 was computed as:

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

The unweighted school participation rate in the student survey after replacement was calculated as:

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

Let t^{fp} be the set of eligible and participating teachers in schools that constitute fp , t^{np} 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, $UPRT_{teachers}$, was defined as:

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

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

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

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

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

Table 10.5: 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
Bulgaria	81.2	94	87.3	70.9	82.1
Brazil	67.1	77.6	82.9	55.6	64.4
Colombia	76	80	86.2	65.5	69
Chinese Taipei	98	99.3	97.3	95.4	96.7
Croatia	97.3	98.7	91.1	88.7	89.9
Cyprus	86.7	86.7	74.5	64.6	64.6
Denmark	14.8	32.4	76.5	11.3	24.8
Estonia	72.9	72.9	81	59	59
France	64	64.7	70.6	45.2	45.7
Italy	96.9	98.1	92.2	89.3	90.5
Lithuania	97.8	97.8	90.8	88.8	88.8
Latvia	76.3	77.6	82.6	63	64.1
Malta	89.7	89.7	91.6	82.1	82.1
Netherlands	43.3	61.3	80.5	34.9	49.3
Norway (9)	83.7	83.7	86.9	72.7	72.7
Poland	90.6	98.2	94.7	85.8	93
Romania	82.9	96.8	97.5	80.8	94.4
Serbia	96.7	100	98.6	95.4	98.6
Slovak Republic	91.9	95.6	94.7	87	90.6
Slovenia	90.2	96.6	95.9	86.5	92.6
Spain	96.9	98.1	87.7	85	86.1
Sweden	83.1	85.1	76.8	63.8	65.3
Benchmarking participant					
North Rhine-Westphalia	78.9	84.9	82	64.7	69.6
Schleswig-Holstein	6.5	9.4	61.1	4	5.8

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

Weighted participation rates in the teacher survey

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

$$WPRT_{schools\_BR} = \frac{\sum_h \sum_{i \in op} \sum_{l \in tfp} WGT FAC1_{hi} \times WGT FAC2_{hil} \times WGT ADJ2T_{hil} \times WGT ADJ3T_{hil}}{\sum_h \sum_{i \in fp} \sum_{l \in tfp} WGT FAC1_{hi} \times WGT ADJ1T_{hi} \times WGT FAC2_{hil} \times WGT ADJ2T_{hil} \times WGT ADJ3T_{hil}}$$

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

$$WPRT_{schools\_AR} = \frac{\sum_h \sum_{i \in fp} \sum_{l \in tfp} WGT FAC1_{hi} \times WGT FAC2_{hil} \times WGT ADJ2T_{hil} \times WGT ADJ3T_{hil}}{\sum_h \sum_{i \in fp} \sum_{l \in tfp} WGT FAC1_{hi} \times WGT ADJ1T_{hi} \times WGT FAC2_{hil} \times WGT ADJ2T_{hil} \times WGT ADJ3T_{hil}}$$

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

$$WPRT_{teachers} = \frac{\sum_h \sum_{i \in fp} \sum_{l \in tfp} WGT FAC1_{hi} \times WGT FAC2_{hil} \times WGT ADJ3T_{hil}}{\sum_h \sum_{i \in fp} \sum_{l \in tfp} WGT FAC1_{hi} \times WGT FAC2_{hil} \times WGT ADJ2T_{hil} \times WGT ADJ3T_{hil}}$$

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

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

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

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

Table 10.6: 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
Bulgaria	83.5	94.1	87.7	73.2	82.5
Brazil	63.3	77.2	84.9	53.7	65.5
Colombia	69.8	76.8	86.2	60.2	66.2
Chinese Taipei	97.5	99.5	97.5	95	97
Croatia	97.7	98.7	91.4	89.3	90.2
Cyprus	86.8	86.8	74.1	64.3	64.3
Denmark	16.2	32.4	77.6	12.6	25.2
Estonia	72.8	72.8	80.2	58.4	58.4
France	64.2	64.7	70.1	45	45.4
Italy	96.9	98.2	92.4	89.5	90.7
Lithuania	97.6	97.6	90.1	87.9	87.9
Latvia	77.8	79.5	83.6	65.1	66.5
Malta	89.9	89.9	91.6	82.3	82.3
Netherlands	45.3	60	78.1	35.4	46.9
Norway (9)	84.6	84.6	87.3	73.8	73.8
Poland	90.4	98.5	94.9	85.9	93.6
Romania	81.1	98.2	97.1	78.7	95.3
Serbia	97.3	100	98.8	96.2	98.8
Slovak Republic	90.4	94.6	94.3	85.3	89.2
Slovenia	90	96.7	96.3	86.7	93.1
Spain	97.1	98.3	88.1	85.5	86.6
Sweden	82.7	85.2	76.4	63.1	65
Benchmarking participant					
North Rhine-Westphalia	79.5	87.2	81.8	65.1	71.3
Schleswig-Holstein	5.6	9.8	61.5	3.4	6

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

10.7 ICCS standards for sampling participation

Despite the efforts of national centers to achieve high participation rates in the aftermath of school closures during the COVID-19 pandemic across many countries, higher levels of non-response were evident in several participating countries or benchmarking entities. As is customary in IEA studies, ICCS 2022 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.

ICCS 2022 reporting strategy was the same as the one in previous cycles (see, for example, Weber et al., 2018). That is, countries grouped in Category 1 met the ICCS 2022 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 2022 sample participation requirements. During an ICCS sampling adjudication meeting in February 2023, the ICCS 2022 sampling referee Marc Joncas made binding decisions as to which country or benchmarking entity would be grouped in which category.

Student survey participation standards

Categories for sampling participation in the ICCS 2022 student survey were defined according to strict criteria (Figure 10.1).

Figure 10.1: 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 generally harder to obtain than for the student survey. However, there is no statistical justification for treating teacher data differently from student data with respect to an assessment of possible non-response bias, especially as teachers' motivation to participate in ICCS 2022 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 for the teacher survey were as strict as the ones for the student survey. Three categories for teacher sampling participation were defined (Figure 10.2).

Figure 10.2: 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 2022 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 appeared in the tables and figures in international reports without annotation.
- Category 2: Countries in this category were annotated in the tables and figures in international reports.
- Category 3: Countries in this category appeared in a separate section of the tables.

In Schleswig-Holstein (Germany), the teacher participation rate was exceptionally low, which made it impossible to generalize from sample data to population characteristics. Therefore, the sampling team did not calculate weights for these teacher data, and the benchmarking participant was not included in the analyses of teacher data for the ICCS 2022 international report.

References

- IEA (2015). Windows ® within-school sampling software (WinW3S) [Computer software]. International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., Damiani, V., Friedman, T. (2024). *Education for citizenship in times of global challenge. IEA international civic and citizenship education study 2022 international report*. Springer.
- Weber, S., Tieck, S., & Savaşçı, D. (2018). Weighting procedures. In: Schulz, W., Carstens, R., Losito, B., & Fraillon, J., (Eds.) (2018). *ICCS 2016 technical report*. International Association for the Evaluation of Educational Achievement (IEA).
- Zuehlke, O. (2011). Sampling design and implementation. In W. Schulz, J. Ainley, & J. Fraillon, (Eds.), *ICCS 2009 technical report*. International Association for the Evaluation of Educational Achievement (IEA).

Chapter 11:

Scaling procedures for civic knowledge test items

Wolfram Schulz and Greg Macaskill

11.1 Overview

This chapter describes the procedures used to analyze and scale the International Civic and Citizenship Education Study (ICCS) 2022 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 in the form of plausible values (PV); and
- Estimation of changes in civic content knowledge between previous cycles and 2022.

The development of the ICCS 2022 test items is described in Chapter 2 and was guided by the ICCS 2022 Assessment Framework (see Schulz et al., 2023).

Data used for the main survey scaling analyses presented in this chapter include those from the computer-based assessment (CBA), used in 16 countries and the two German benchmarking entities, and paper-based assessment (PBA), used in six countries. The CBA included 11 clusters of items common to both delivery modes plus three clusters of items used in the CBA only. The PBA included the 11 clusters of items common to the CBA only.

The main survey also included a bridging study for the purpose of investigating the effect of the delivery mode on measuring student outcome variables. All countries that participated in ICCS 2016 (for which all student assessments were delivered on paper) and were participating in ICCS 2023 using the CBA mode participated in the bridging study. The review of delivery mode effects used CBA data collected as part of the main survey these countries (using only the data from the 11 clusters of items common to both modes), as well as data from PBA conducted as part of the bridging study (using eight of the 11 clusters of items common to both modes). Comparisons between main survey and bridging study data in these 11 countries were used to review mode effects on how students responded to test items.

11.2 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:

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

Where $P_i(\theta_n)$ is the probability for person n to score 1 on item i , θ_n is the ability of person n , and δ_i is the location of item i on this dimension. For each item, item responses are modeled as a function of the latent trait θ_n .

More generally, if item i is a polytomous item with score categories $0, 1, 2, \dots, m_i$, we used the Partial Credit Model (Masters, 1982). The probability of person n scoring x on item i is then given by:

$$P_i(\theta_n) = \frac{\exp \sum_{k=0}^x (\theta_n - \delta_i + \tau_{ik})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^h (\theta_n - \delta_i + \tau_{ik})}, x_i = 1, 2, \dots, m_i$$

Here, $P_i(\theta_n)$ 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 while τ_{ik} is an additional step parameter.

The process of producing estimates of ability, difficulty (location), and step parameters is sometimes referred to as item calibration. Estimation of scaling parameters is not straightforward and requires elaborate iterative numerical algorithms and extensive quality control by experienced data analysts. To perform item calibration for the ICCS 2022 test data, we used the scaling software package ACER ConQuest, Version 5 (Adams et al., 2020).

11.3 Scaling outcomes

Test targeting and item dimensionality

Test targeting

When measuring cognitive abilities, it is important to use test items that cover the different levels of ability found in the target population. As a start point, we estimated the distribution of cognitive abilities among ICCS 2022 students and the location of item thresholds (with a response probability (rp) of 0.5; see Figure 11.1). Item thresholds were equal to item difficulties for dichotomous items. For partial credit items, a difficulty threshold was estimated for each score, indicated by .1, .2, or .3 appended to the item number. This mapping comparing student abilities and item thresholds is known as a *Wright Map*.¹

Figure 11.1: Mapping of student abilities and item thresholds



¹ This terminology was introduced by Wilson and Draney (2002) in honor of the psychometrics expert Ben Wright (University of Chicago).

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 2022 test items were slightly better targeted to students in the lower civic knowledge ranges in comparison to the higher civic knowledge ranges, a lot of items had difficulties between -1 and 0 logits.² The nature of this targeting varied across countries according to the distribution of students' civic knowledge within each country.

Item dimensionality

In response to the new areas of increased emphasis in the ICCS 2022 Assessment Framework (Schulz et al., 2023) the ICCS test instrument included 49 items mapped to the Global Citizenship Education/Education for Sustainable Development (GCED/ESD) themes—28 aligned with ESD and 21 aligned with GCED. Using items that were assigned to each of GCED, ESD and “Other ICCS” domains, we explored the associations across these three domains applying a multi-dimensional IRT model (see, for example, von Davier & Carstensen, 2007) which allows us to obtain estimated correlations between latent dimensions. This model is akin to a confirmatory factor analysis (Bollen, 1989) but within an IRT modeling framework.

As a criterion for judging dimensionality, we relied on reviewing correlations between the latent dimensions, given that statistical tests of relative model fit are cumbersome in view of the large sample sizes. However, it needs to be acknowledged that these correlations were calculated based on data from the pooled sample, and that lower correlations might be observed when using individual national samples.

When estimating the multi-dimensional IRT model with three factors reflecting item content related to GCED, ESD and other (non-GCED/ESD) aspects, we found latent correlations of 0.88 between GCED and ESD, 0.93 between GCED and other aspects, and 0.92 between ESD and other aspects. While these correlations are quite high, they suggest some potential difference between the sustainability dimension and the two other dimensions.

We further reviewed, using the same multi-dimensional IRT approach described above, two other possible dimensionality structures associated with the conceptual structure of the ICCS 2022 Assessment Framework³:

- Cognitive domains (knowing vs reasoning and applying, with 94 and 42 items respectively); and
- Content domains, civic institutions, and systems with civic roles and identities as a single domain (due to low item numbers for the latter domain and the conceptual coherence between these two domains) versus civic principles versus civic participation. The number of items in the resultant three domains are 43, 51, and 42 respectively.

A two-dimensional model was used to review the correlation between the two cognitive domains. The latent correlation (see Bollen, 1989) between the domains was 0.94, which suggests that items from each domain measure highly similar aspects of civic knowledge.

A three-dimensional model was used to review the correlation among the three content domains: (i) civic institutions and systems/civic roles and identities⁴; (ii) civic principles; and (iii) civic participation. The results also showed very high correlations between each pair of domains. The latent correlation between civic institutions and systems/civic roles and identities and each of civic principles and civic participation was 0.94 and 0.92 respectively. The latent correlation between civic principles and civic participation was 0.93. These results are consistent with those of previous ICCS cycles where civic knowledge was only reported as a unidimensional construct.

Analyses of dimensionality suggested that civic knowledge could be reported as a unidimensional scale. However, it is expected that secondary analyses will use ICCS 2022 cognitive data to further explore civic

² The logit of a probability p is defined as $\log(p/1-p)$. It maps the probability onto the full real line.

³ Table 1.1 in the ICCS 2022 international report (Schulz et al., 2024) includes information about the number of items in the intersection of these domains.

⁴ Content domains 1 and 4 from the ICCS assessment framework were combined in this analysis because of their conceptual congruence and the relatively small number of items associated with content domain 4.

knowledge sub-domains of specific research interest, especially regarding the ESD- and GCED-related sub-domains.

Student non-response data

There are four possible types of student non-response that could appear in the ICCS 2022 test data:

- *omitted items* (coded as 9): items to which a student was exposed but did not enter a response;
- *not administered items* (coded as 8): items to which a student was not exposed (most commonly those items in clusters that were not in the test form completed by a given student because of the rotated cluster design), but in some cases responses were also missing due to technical issues (CBA) or printing errors (PBA);
- *invalid responses* (coded as 7): responses that could not be scored because they are inconsistent with the interpretable response set—such as if a student were to select two responses to a multiple-choice item. These were relevant to the ICCS PBA only because the CBA environment controls the response formats available to students; and
- *“not reached” items* (coded as 6): this is a derived category of non-response based on evidence from strings of omitted items at the end of a test form completed by a student. If the final valid response in a test form completed by a student was followed by more than two omitted responses, the first omitted response remained coded as 9 while all remaining trailing omitted responses were deemed to have been not reached and were recoded accordingly.

Table 11.1 shows the average percentages of omitted responses for each item across each delivery mode (main survey CBA, main survey PBA, bridging study PBA) and the average percentages of invalid responses for items when delivered on paper (invalid responses did not occur when items were delivered on computer). The percentages for computer-delivered items were also calculated only for the 11 countries that participated in the bridging study (column 5). This makes it possible to compare them with the PBA bridging study averages (column 6) calculated using the paper-based bridging study data for the same set of countries.

Table 11.1: Percentages of omitted and invalid responses by delivery mode

Item	Format	Omitted Responses				Invalid Responses	
		CBA – 18 countries (%)	PBA – 6 countries (%)	CBA – 11 countries (%)	PBA bridging study – 11 countries (%)	PBA – 11 countries (%)	PBA bridging study – 11 countries (%)
CI2ASM1	MC	0.80	0.31	0.22	0.39	1.17	0.37
CI2ASM2	MC	0.51	0.48	0.14	0.61	1.15	0.42
CI2BIO1	CR	9.76	11.34	8.45	9.67	0.16	0.00
CI2BPM1	MC	0.60	0.26	0.41	0.69	1.92	0.53
CI2BPM2	MC	0.50	0.41	0.68	0.64	1.46	0.42
CI2CCM1	MC	0.47	0.41	0.39	0.30	0.88	0.36
CI2CCM2	MC	0.86	0.61	0.58	0.90	1.56	0.43
CI2CEM1	MC	0.69	0.52	0.86	0.83	1.41	0.54
CI2CEM2	MC	0.73	0.25	0.20	0.77	2.59	0.77
CI2CNM1	MC	0.19	0.09			2.51	
CI2CNM2	MC	0.37	0.19			0.63	
CI2ECM1	MC	0.63	0.25			1.06	
CI2ECM2	MC	0.79	0.31			1.77	
CI2FDM1	MC	0.49	0.62	0.37	0.61	1.00	0.46
CI2FSM1	MC	0.17	0.37	0.13	0.80	2.35	0.37
CI2GLM1	MC	0.46	0.67	0.54	1.03	2.18	0.58
CI2GLM2	MC	1.08	0.72	0.73	0.88	1.29	0.42

Item	Format	Omitted Responses				Invalid Responses	
		CBA – 18 countries (%)	PBA -6 countries (%)	CBA – 11 countries (%)	PBA bridging study – 11 countries (%)	PBA – 11 countries (%)	PBA bridging study – 11 countries (%)
CI2ORM1	MC	0.47	0.41			0.62	
CI2PGM1	MC	1.11	0.52	1.47	0.72	1.13	0.38
CI2PGM2	MC	1.87	0.89	0.71	1.78	1.35	0.74
CI2PRM1	MC	0.46	0.43	0.22	0.60	1.50	0.13
CI2RCM1	MC	1.14	0.68	0.64	0.75	0.66	0.26
CI2REM2	MC	0.29	0.41	0.26	0.94	3.74	1.08
CI2REM3	MC	0.27	0.32	0.19	0.72	2.80	0.87
CI2WFO1	CR	7.12	12.99	5.30	10.26	0.35	0.03
CI2WFO2	CR	6.60	11.60	5.88	8.88	0.34	0.01
CI308M1	MC	1.27	0.25	0.55	0.52	1.19	0.24
CI312M1	MC	0.78	0.74	0.59	0.45	1.17	0.20
CI314M1	MC	0.38	0.39			1.62	
CI3CAM1	MC	0.46	0.60	0.54	0.99	1.51	0.13
CI3CPO1	CR	5.93	9.98	4.00	8.33	0.17	0.02
CI3CPO2	CR	7.75	9.92	6.01	9.41	0.20	
CI3CSM1	MC	0.52	0.44	0.37	0.52	1.85	0.39
CI3DDM1	MC	0.62	0.40	0.55	1.36	0.84	0.26
CI3EPM1	MC	0.52	0.52			1.19	
CI3GMM1	MC	0.57	0.55	0.57	0.68	1.21	0.28
CI3IEM1	MC	0.37	0.37	0.30		0.92	
CI3IVM1	MC	0.50	0.63	0.47	0.79	1.04	0.22
CI3LPM1	MC	0.56	0.49	0.48	0.94	0.89	0.45
CI3LSM1	MC	0.31	0.26	0.26		1.42	
CI3LTM1	MC	0.70	0.54			2.12	
CI3MAM1	MC	0.43	0.29	0.37	1.06	0.86	0.47
CI3MDM1	MC	0.67	0.44			1.03	
CI3NPM1	MC	0.22	0.07	0.22	0.63	2.32	0.41
CI3NPM2	MC	0.25	0.16	0.30	0.51	0.83	0.29
CI3PAM1	MC	0.64	0.72	0.43	0.52	0.54	0.23
CI3PEM1	MC	1.25	0.62			0.71	
CI3PLM1	MC	0.25	0.20	0.14	0.75	0.72	0.10
CI3PRO1	CR	10.16	13.22			0.30	
CI3RRM1	MC	0.58	0.49	0.45	0.65	0.44	0.41
CI3SCM1	MC	0.42	0.36	0.26	0.71	1.37	0.43
CI3SDM1	MC	0.70	0.39			1.72	
CI3UHM1	MC	0.98	0.63	0.18	0.82	1.43	0.54
CI3ULM1	MC	0.43	0.62	0.40	0.72	1.37	0.50
CI3VGM1	MC	0.38	0.30			0.67	
CI4CFM1	MC	1.06	1.21	0.67	1.79	1.31	0.41
CI4CZM1	MC	0.39	0.38			3.37	
CI4DCM1	MC	0.59	0.36			1.34	

Item	Format	Omitted Responses				Invalid Responses	
		CBA – 18 countries (%)	PBA -6 countries (%)	CBA – 11 countries (%)	PBA bridging study – 11 countries (%)	PBA – 11 countries (%)	PBA bridging study – 11 countries (%)
CI4DDM1	MC	0.49	0.17	0.48	0.58	2.86	1.06
CI4DDM2	MC	0.57	0.19	0.51	1.06	0.78	0.09
CI4DVC1	CR	10.16	12.75	8.87	9.08	0.18	0.01
CI4EAM1	MC	0.63	0.36	0.46	0.53	1.30	0.51
CI4ECM1	MC	0.38	0.25	0.24	0.26	2.21	0.42
CI4ECM2	MC	0.25	0.46	0.29	1.04	3.14	0.43
CI4ECM3	MC	0.38	0.29	0.39	0.83	2.54	0.37
CI4EDC2	CR	9.51	11.30			0.47	
CI4EDM1	MC	0.59	0.54			0.97	
CI4EPM1	MC	0.58	0.36	0.32	0.51	1.65	0.45
CI4EPM2	MC	0.45	0.22	0.37	0.97	2.73	0.70
CI4ERM1	MC	0.62	0.18	0.28	0.62	0.97	0.46
CI4FFM1	MC	0.69	0.46			0.52	
CI4FKC1	CR	8.49	8.55	5.25	8.36	0.40	0.00
CI4FKM2	MC	0.58	0.48	0.23	0.32	3.09	0.93
CI4FRM1	MC	0.75	0.35	0.67	1.33	2.21	0.45
CI4FTM1	MC	0.56	0.34			1.89	
CI4FWC2	CR	6.74	6.38	6.11	8.45	0.18	0.00
CI4GCM1	MC	0.51	0.38	0.46	0.64	1.29	0.72
CI4GLM1	MC	0.66	0.56	0.71	1.66	0.80	0.37
CI4INM1	MC	0.58	0.20	0.11	0.47	0.82	0.38
CI4IPM1	MC	0.20	0.27	0.29	1.07	1.45	0.58
CI4JCM1	MC	0.35	0.37			4.94	
CI4JCM2	MC	0.55	0.49			1.93	
CI4JSM1	MC	0.68	0.43	0.38	0.83	2.63	1.05
CI4JSM2	MC	0.62	0.26	0.51	0.58	0.95	0.42
CI4LEM1	MC	1.00	0.67	0.27	1.37	1.23	0.32
CI4LEM2	MC	1.47	0.72	0.72	1.56	0.93	0.25
CI4MWM1	MC	0.54	0.24	0.45	0.91	1.15	0.29
CI4MWM2	MC	0.76	0.61	0.75	0.93	1.77	0.25
CI4MWM3	MC	0.58	0.46	0.47	1.23	1.95	0.47
CI4NDM1	MC	0.59	0.35	0.48	0.49	1.80	0.44
CI4NHC2	CR	11.74	12.13	9.22	11.07	0.30	0.00
CI4NHM1	MC	0.25	0.26			0.47	
CI4NTM1	MC	0.75	0.34	0.71	0.76	2.46	0.70
CI4NWM1	MC	0.73	0.53	0.76	0.50	2.80	0.37
CI4NWM2	MC	0.27	0.27	0.24	0.42	0.80	0.31
CI4OCM1	MC	0.86	0.31	0.83	0.97	2.02	0.55
CI4OCM2	MC	0.73	0.57	0.70	1.00	1.23	0.62
CI4PDM1	MC	0.99	0.46	0.72	0.53	1.50	0.40
CI4PDM2	MC	0.73	0.41	0.55	0.74	0.82	0.40

Item	Format	Omitted Responses				Invalid Responses	
		CBA – 18 countries (%)	PBA -6 countries (%)	CBA – 11 countries (%)	PBA bridging study – 11 countries (%)	PBA – 11 countries (%)	PBA bridging study – 11 countries (%)
CI4PIM1	MC	0.56	0.29			1.58	
CI4PRM1	MC	0.38	0.28	0.27	0.93	1.30	0.52
CI4RCM1	MC	0.85	0.70	0.45	0.54	2.55	1.00
CI4REM1	MC	0.40	0.57	0.34	0.69	1.12	0.67
CI4RFM1	MC	0.68	0.24	0.29	0.96	1.78	0.78
CI4RPM1	MC	0.97	0.56			0.57	
CI4SCM1	MC	0.68	0.29			2.74	
CI4SDM1	MC	0.58	0.24	0.42	0.50	1.01	0.75
CI4SDM2	MC	0.70	0.31	0.86	0.74	2.87	0.81
CI4SDM3	MC	0.67	0.67	0.48	1.09	1.47	0.81
CI4SJM1	MC	1.60	0.88	1.11	0.68	1.64	0.37
CI4SLC3	CR	9.85	16.72			0.48	
CI4SLC4	CR	12.57	18.38			0.44	
CI4SLM1	MC	0.86	0.50	0.35		1.82	
CI4SSM1	MC	0.26	0.51	0.14	0.30	0.70	0.28
CI4SWM1	MC	1.19	0.60	1.11		1.49	
CI4TDM1	MC	0.33	0.32	0.22	1.02	2.77	1.00
CI4TXM1	MC	0.63	0.49	0.51	0.48	0.47	0.29
CI4UWM1	MC	0.91	0.79			0.65	
CI4VRM1	MC	0.73	0.51	0.37	0.85	1.28	0.22
CI4VRM2	MC	0.56	0.47	0.54	0.23	1.41	0.28
CI4ZEC2_R	CR	0.41					
CI4ZEC4	CR	2.73					
CI4ZEC5	CR	11.17					
CI4ZED3L_R	DD	3.20					
CI4ZEM1	MC	0.23					
CI4ZHC1_R	CR	6.86					
CI4ZHC3	CR	6.45					
CI4ZHC4	CR	6.00					
CI4ZHL2_R	LT	0.21					
CI4ZHM5	MC	0.21					
CI4ZSC4	CR	0.51					
CI4ZSC5N	CR	4.09					
CI4ZSC5Y	CR	5.92					
CI4ZSD2	DD	9.59					
CI4ZSL3_R	LT	6.99					
CI4ZSM1	MC	2.36					

MC = Multiple choice; CR = Constructed Response; DD = Drag and drop; LT=Large task

For *multiple-choice* (MC) items, the average percentage of omitted responses was below one percent (Table 11.2) and varied only slightly by delivery mode. A slightly higher rate of omission was found for *large task* (LT) items and *drag and drop* (DD) items that had about five percent missing by omission. *Constructed response* (CR)

items, where students were required to write or type a response, had the highest proportions of omitted responses across all four item types, around about nine percent on average. This was fairly consistent with what was found in ICCS 2016 where omissions were around nine percent on average for CR items. For this item format, the average proportion of omitted responses was higher for paper-based responses (11.94% (PBA) and 9.28% (PBA bridging study) than for computer-based responses (7.71% (CBA) and 6.57% (CBA bridging study)). Similar observations about differences in missing proportions between computer- and paper-based test data were also found in other studies (see, for example, Kroehne et al., 2019).

The percentage of invalid responses was lower for PBA bridging study data than in the PBA mode, though both were fairly low. The very low percentages of invalid responses to CR items were to be expected, as there are very few ways in which an invalid response can be generated by students for this item format

Table 11.2: Average percentages of omitted and invalid responses by delivery mode and item format

Item Format	Omitted Responses						Invalid Responses		
	CBA countries (18 countries)		PBA countries (6 countries)		CBA & Bridging study (11 countries)		PBA countries (6 countries)		Bridging study (11 countries)
	Clusters 1 to 14	Clusters 1 to 11	Clusters 1 to 11	Clusters 1 to 8	Clusters 1 to 8	Clusters 1 to 8	Clusters 1 to 11	Clusters 1 to 8	Clusters 1 to 8
	CBA (%)	CBA (%)	PBA (%)	PBA (%)	CBA (%)	PBA (%)	PBA (%)	PBA (%)	PBA (%)
CR	7.71	8.95	11.94	10.72	6.57	9.28	0.31	0.26	0.01
DD	5.09								
LT	1.28								
MC	0.62	0.63	0.43	0.45	0.46	0.79	1.55	1.56	0.48

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 provides an indication of the extent to which each item is consistent with 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 assessed item fit by using a range of item statistic, including the item-rest correlations and the weighted mean square (*infit*) statistic given in the table below.

We determined the item-rest correlations of correct responses (or partial credit responses) and the weighted item fit statistics (Table 11.3). Almost all items had item-total correlations above 0.20 (items with lower discrimination than 0.20 are considered only weakly related to the latent variable that is measured), and item CI4SLM2 had a low item-rest correlation of 0.16. The constructed items CI2WFO2, CI4ZHL2_R, and CI4ZSL3_R had *weighted infit statistics* above 1.2 (indicating lower-than expected discrimination), however, all of them had item-rest correlation of 0.20 or more.

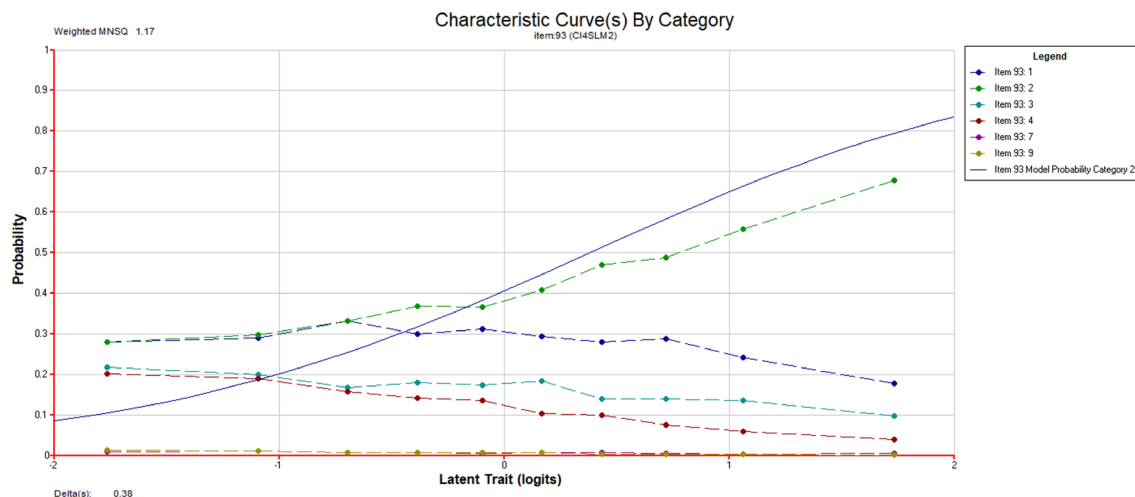
Table 11.3: Item facilities, total-score correlations, and weighted item fit for international sample

Item name	Facility	Item-Rest Correlation	Weighted MNSQ	Item name	Facility	Item-Rest Correlation	Weighted MNSQ	Item name	Facility	Item-Rest Correlation	Weighted MNSQ
CI3MAM1	79.7	0.28	1.06	CI3CAM1	49.9	0.27	1.11	CI4SLM2	43.4	0.16	1.17
CI4JSM1	55.7	0.37	1.02	CI4NHC2	39.0	0.48	1.02	CI4SLC3	32.7	0.35	1.01
CI4JSM2	68.3	0.47	0.90	CI4DDM1	66.5	0.44	0.96	CI4SLC4	28.7	0.33	1.01
CI2WFO2	42.8	0.36	1.30	CI4DDM2	80.7	0.44	0.92	CI314M1	74.8	0.35	1.00
CI4SDM1	77.0	0.44	0.93	CI3LPM1	77.1	0.50	0.88	CI2CEM1	75.8	0.38	0.97
CI4SDM2	74.5	0.43	0.93	CI4LEM1	66.1	0.40	1.00	CI2CEM2	48.2	0.21	1.15
CI4SDM3	69.2	0.43	0.96	CI4LEM2	52.4	0.36	1.03	CI3EPM1	71.7	0.40	0.97
CI2ECM1	71.3	0.47	0.92	CI3CSM1	64.8	0.39	1.00	CI4CZM1	66.9	0.33	1.06
CI2ECM2	64.3	0.29	1.08	CI312M1	63.2	0.48	0.92	CI2ORM1	72.2	0.46	0.92
CI308M1	76.3	0.45	0.90	CI4SSM1	66.3	0.36	1.02	CI3LSM1	73.2	0.44	0.93
CI4SJM1	60.7	0.52	0.88	CI3PAM1	69.3	0.45	0.93	CI3PRO1	42.4	0.44	1.08
CI3RRM1	79.0	0.41	0.95	CI3ULM1	59.4	0.36	1.01	CI3SDM1	60.6	0.43	0.97
CI2REM2	65.2	0.27	1.10	CI2BIO1	22.7	0.37	1.06	CI4JCM1	50.4	0.37	1.02
CI2REM3	77.9	0.44	0.93	CI4MVM1	69.4	0.43	0.95	CI4JCM2	66.6	0.34	1.03
CI4REM1	70.4	0.34	1.05	CI4MVM2	60.5	0.48	0.91	CI4PIM1	73.6	0.45	0.95
CI4FKC1	45.2	0.40	1.12	CI4MVM3	47.3	0.37	1.00	CI4RPM1	60.7	0.43	0.98
CI4FKM2	73.2	0.32	1.04	CI4NVM1	60.8	0.36	1.03	CI4EDM1	66.1	0.46	0.94
CI2CCM1	73.1	0.40	0.96	CI4NVM2	89.1	0.38	0.92	CI3PEM1	59.9	0.38	1.02
CI2CCM2	57.9	0.35	1.02	CI3DDM1	60.0	0.39	0.99	CI3LTM1	47.9	0.36	1.02
CI4EPM1	78.0	0.51	0.88	CI3GMM1	63.4	0.34	1.04	CI3IEM1	68.7	0.36	1.01
CI4EPM2	66.7	0.30	1.07	CI4IPM1	67.2	0.40	1.00	CI4FTM1	46.4	0.20	1.16
CI4CFM1	40.8	0.38	0.98	CI3NPM1	90.7	0.32	0.96	CI3MDM1	61.8	0.42	0.97
CI2FSM1	67.2	0.39	0.97	CI3NPM2	86.1	0.33	0.97	CI2CNM1	67.3	0.37	1.00
CI4TDM1	73.2	0.34	1.02	CI4ERM1	73.4	0.49	0.90	CI2CNM2	82.2	0.47	0.93
CI2PGM1	62.1	0.36	1.01	CI4GLM1	48.1	0.30	1.08	CI4FFM1	68.8	0.45	0.94
CI2PGM2	41.2	0.31	1.04	CI4GCM1	53.9	0.39	1.00	CI4EDC2	51.3	0.47	1.02
CI4FWC2	57.1	0.44	1.06	CI4DVC1	43.0	0.43	1.12	CI4SWM1	71.0	0.26	1.10
CI4NDM1	69.3	0.34	1.03	CI2GLM1	67.8	0.36	1.03	CI4SCM1	50.8	0.35	1.03
CI2BPM1	66.6	0.37	1.00	CI2GLM2	61.7	0.35	1.05	CI4UWM1	69.2	0.46	0.93
CI2BPM2	67.2	0.39	0.98	CI3SCM1	78.9	0.42	0.95	CI4ZHC1_R	55.7	0.52	1.00
CI4OCM1	50.5	0.34	1.04	CI4EAM1	71.3	0.46	0.94	CI4ZHL2_R	34.7	0.31	1.24
CI4OCM2	59.3	0.39	0.99	CI4INM1	90.5	0.34	0.98	CI4ZHC3	46.5	0.44	0.94
CI4NTM1	69.1	0.47	0.91	CI2ASM1	59.1	0.41	0.99	CI4ZHC4	33.7	0.37	0.99
CI3PLM1	85.8	0.38	0.92	CI2ASM2	75.0	0.45	0.93	CI4ZHM5	75.2	0.38	0.99
CI4ECM1	49.0	0.39	1.00	CI3UHM1	67.5	0.42	0.98	CI4ZEM1	83.9	0.34	0.99
CI4ECM2	54.7	0.33	1.04	CI4RFM1	72.1	0.40	0.99	CI4ZEC2_R	51.2	0.43	0.97
CI4ECM3	54.8	0.34	1.04	CI3CPO1	60.1	0.37	1.02	CI4ZED3L_R	37.2	0.42	1.07
CI2FDM1	70.0	0.39	0.99	CI3CPO2	51.6	0.41	0.98	CI4ZEC4	24.2	0.31	1.05
CI2WFO1	45.5	0.40	1.11	CI4PDM1	50.7	0.46	0.94	CI4ZEC5	33.8	0.52	0.97
CI2RCM1	67.0	0.39	1.00	CI4PDM2	64.2	0.47	0.94	CI4ZSM1	65.2	0.37	1.00
CI4TXM1	80.1	0.42	0.92	CI4FRM1	63.8	0.45	0.94	CI4ZSD2	10.9	0.20	1.02
CI4VRM1	58.3	0.40	1.00	CI4RCM1	62.4	0.42	0.97	CI4ZSL3_R	81.5	0.20	1.28
CI4VRM2	62.8	0.34	1.05	CI3VGM1	86.5	0.39	0.92	CI4ZSC4	55.3	0.41	0.98
CI3IVM1	71.9	0.44	0.95	CI4NHM1	75.2	0.25	1.07	CI4ZSC5Y	40.1	0.39	0.98
CI4PRM1	75.8	0.45	0.92	CI4DCM1	68.5	0.40	0.97	CI4ZSC5N	29.6	0.31	1.02
CI2PRM1	47.5	0.35	1.04	CI4SLM1	72.0	0.39	0.98				

We also used ACER 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 CI4SLM2 shows the discrimination was not satisfactory and that the proportion of students with correct answers did not increase as expected by the scaling model (Figure 11.2). Given this outcome, we decided not to retain this item for scaling.

Figure 11.2: Item characteristic curve by category for item CI4SLM2



Assessment of scorer agreement

In order to score the open-ended items in the ICCS 2022 cognitive test, we followed the scoring guides that were refined as an outcome of experiences in the international field trial of new test items. For the CBA mode, the RM AM Marker software was configured so that approximately 200 responses to each item were scored by a second scorer. In the online scoring context, scorers were unaware of whether any given response had been allocated to be double scored. For the PBA mode, subsamples of approximately 200 student responses were double scored. This required manual handling of the booklets. In this context, scorers were aware of whether or not a response they were scoring was allocated to be double scored but were not able to see any scores allocated by the other scorer. All double scoring across both delivery modes was therefore completed with scorers unaware of any other scores given to responses they may have scored. This double-scoring procedure allowed us to assess scorer agreement and the national percentages of scorer agreement, which ranged between 53 and 100 percent (Table 11.4). On average, scorer agreement for the items was between 70 and 90 percent.⁵ Percentages of 100 percent indicate that national centers may have misinterpreted the international guidelines and used reconciled scoring data for the recording of double-scoring results.

Table 11.4: Percentages of scorer agreement for open-ended ICCS test items

Country	CX2WFO2	CX4FKC1	CX4FWC2	CX2WFO1	CX4NHC2	CX2BIO1	CX4DVC1	CX3CPO1	CX3CPO2	CX4SLC3	CX4SLC4
Brazil	66	68	75	61	63	58	63	71	60	79	78
Bulgaria	74	62	68	71	80	76	69	76	81	85	88
Chinese Taipei	87	80	85	74	68	74	73	88	90	92	92
Colombia	70	73	75	72	75	80	77	70	83	84	89
Croatia	83	83	96	92	91	85	93	95	96	95	95
Cyprus	93	85	77	87	100	80	87	66	66	94	92
Denmark	95	91	96	92	92	95	92	95	95	94	97
Estonia	84	96	96	91	91	87	91	79	80	90	83
France	71	80	87	81	84	88	83	83	82	-	91
Italy	82	88	90	84	87	89	84	89	87	86	88
Latvia	67	80	88	68	63	75	71	64	59	79	84
Lithuania	95	96	95	97	94	98	98	91	90	94	96
Malta	82	79	77	78	80	84	84	66	70	88	85
Netherlands	80	79	81	79	80	85	87	84	91	87	90
North Rhine-Westphalia*	62	77	76	77	81	83	78	72	72	84	86
Norway	79	68	79	76	74	85	79	73	83	86	87
Poland	95	95	97	95	98	97	94	97	98	93	96
Romania	100	100	100	100	100	100	100	100	100	100	100
Schleswig-Holstein*	78	71	82	79	84	76	78	80	75	85	86
Serbia	97	94	90	92	88	94	97	95	95	91	93
Slovak Republic	65	65	76	65	70	80	64	73	81	80	88
Slovenia	92	98	99	99	99	100	99	96	94	100	100
Spain	94	97	95	96	97	95	93	95	94	97	94
Sweden	57	68	66	62	61	58	70	74	77	79	80
Average	81	82	85	82	83	84	84	82	83	89	90
Minimum	57	62	66	61	61	58	63	64	59	79	78
Maximum	100	100	100	100	100	100	100	100	100	100	100

*Participation in ICCS 2022 as a benchmarking entity.

⁵ 70 percent was the lower bound of expected reliability for this study.

Table 11.4: Percentages of scorer agreement for open-ended ICCS test items (cont.)

Country	CX3PRO1	CX4EDC2	CX4ZHC1	CX4ZHC3	CX4ZHC4	CX4ZEC2	CX4ZEC4	CX4ZEC5	CX4ZSC4	CX4ZSC5Y	CX4ZSC5N
Brazil	72	62	64	79	83	68	-	59	63	83	68
Bulgaria	78	76	81	85	86	84	87	75	63	82	77
Chinese Taipei	88	77	93	96	83	98	89	80	79	90	86
Colombia	73	77	75	87	83	82	79	85	53	81	67
Croatia	91	89	-	-	-	-	-	-	-	-	-
Cyprus	87	86	84	89	92	86	93	88	63	91	82
Denmark	93	96	-	-	-	-	-	-	-	-	-
Estonia	87	85	93	89	87	94	86	84	71	93	79
France	85	89	92	93	84	88	95	86	68	85	68
Italy	86	84	91	93	88	95	94	92	78	84	74
Latvia	76	76	83	79	84	80	82	71	67	82	62
Lithuania	97	95	96	97	97	95	96	94	69	98	85
Malta	83	80	85	91	87	85	81	82	72	80	78
Netherlands	77	72	-	-	-	-	-	-	-	-	-
North Rhine-Westphalia*	77	77	91	93	82	83	-	83	77	77	73
Norway	78	84	82	93	90	95	98	88	61	81	83
Poland	94	95	-	-	-	-	-	-	-	-	-
Romania	100	100	-	-	-	-	-	-	-	-	-
Schleswig-Holstein*	72	78	89	91	78	85	-	86	76	79	78
Serbia	92	94	-	-	-	-	-	-	-	-	-
Slovak Republic	82	74	82	93	86	95	90	82	72	73	78
Slovenia	100	100	100	100	100	100	100	100	81	100	91
Spain	97	94	97	95	91	93	92	95	77	98	90
Sweden	68	67	75	84	79	72	74	67	68	71	66
Average	85	84	86	90	87	88	89	83	70	85	77
Minimum	68	62	64	79	78	68	74	59	53	71	62
Maximum	100	100	100	100	100	100	100	100	81	100	91

*Participation in ICCS 2022 as a benchmarking entity.

Differential item functioning by gender

IRT scaling models assume the probability to be only a function of the measured latent variable and not of any group membership. Therefore, we explored the quality of the items by assessing differential item functioning (DIF). DIF occurs when sub-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*. In this study we reviewed DIF across groups defined by gender (female vs male) and national context (24 participating education systems).

We derived estimates of gender DIF by including interaction terms in the item response model. We used the following facet model to investigate the occurrence of gender DIF for dichotomously scored test items:

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

θ_n denotes the ability of person n and δ_i the 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.

For constructed items with more than two categories, this was modeled as:

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

The difference in item difficulty between gender groups was estimated by multiplying the item-by-gender interaction estimates by two to obtain the estimated difference between the two comparison groups and was reviewed for each item (Table 11.5). Based on pooled data from equally weighted countries that were included in the final calibration we observed 18 items for which the estimated differences in ability between gender groups was larger than 0.3 logits, which was used as a threshold equivalent to about a third of a standard deviation of the measured ability. However, occurrences of more considerable gender DIF were roughly equally balanced between females (8) and males (10).

Table 11.5: Gender DIF estimates for test items

Item No.	Item name	Difference between groups	Easier for	Difference larger than 0.3 logits
1	CI2ASM1	0.42	Males	Yes
2	CI2ASM2	0.05	Males	No
3	CI2BIO1	-0.28	Females	No
4	CI2BPM1	0.23	Males	No
5	CI2BPM2	-0.16	Females	No
6	CI2CCM1	0.03	Males	No
7	CI2CCM2	-0.04	Females	No
8	CI2CEM1	0.12	Males	No
9	CI2CEM2	-0.04	Females	No
10	CI2CNM1	0.18	Males	No
11	CI2CNM2	-0.28	Females	No
12	CI2ECM1	-0.12	Females	No
13	CI2ECM2	0.03	Males	No
14	CI2FDM1	0.00	Females	No
15	CI2FSM1	0.22	Males	No
16	CI2GLM1	0.15	Males	No
17	CI2GLM2	-0.03	Females	No
18	CI2ORM1	-0.30	Females	No
19	CI2PGM1	-0.12	Females	No
20	CI2PGM2	0.02	Males	No
21	CI2PRM1	0.05	Males	No
22	CI2RCM1	0.08	Males	No
23	CI2REM2	0.02	Males	No
24	CI2REM3	-0.19	Females	No
25	CI2WFO1	0.04	Males	No
26	CI2WFO2	-0.06	Females	No
27	CI308M1	-0.16	Females	No
28	CI312M1	-0.32	Females	Yes
29	CI314M1	0.18	Males	No
30	CI3CAM1	-0.05	Females	No
31	CI3CPO1	-0.38	Females	Yes
32	CI3CPO2	-0.23	Females	No
33	CI3CSM1	-0.11	Females	No
34	CI3DDM1	-0.18	Females	No
35	CI3EPM1	-0.05	Females	No
36	CI3GMM1	0.01	Males	No

Item No.	Item name	Difference between groups	Easier for	Difference larger than 0.3 logits
37	CI3IEM1	0.11	Males	No
38	CI3IVM1	-0.22	Females	No
39	CI3LPM1	-0.48	Females	Yes
40	CI3LSM1	-0.23	Females	No
41	CI3LTM1	0.36	Males	Yes
42	CI3MAM1	0.13	Males	No
43	CI3MDM1	0.17	Males	No
44	CI3NPM1	-0.30	Females	Yes
45	CI3NPM2	-0.39	Females	Yes
46	CI3PAM1	0.05	Males	No
47	CI3PEM1	0.31	Males	Yes
48	CI3PLM1	-0.19	Females	No
49	CI3PRO1	0.03	Males	No
50	CI3RRM1	-0.19	Females	No
51	CI3SCM1	-0.03	Females	No
52	CI3SDM1	0.17	Males	No
53	CI3UHM1	0.01	Males	No
54	CI3ULM1	0.05	Males	No
55	CI3VGM1	-0.22	Females	No
56	CI4CFM1	0.41	Males	Yes
57	CI4CZM1	0.09	Males	No
58	CI4DCM1	-0.13	Females	No
59	CI4DDM1	-0.08	Females	No
60	CI4DDM2	-0.28	Females	No
61	CI4DVC1	0.07	Males	No
62	CI4EAM1	-0.13	Females	No
63	CI4ECM1	0.15	Males	No
64	CI4ECM2	-0.03	Females	No
65	CI4ECM3	0.08	Males	No
66	CI4EDC2	-0.29	Females	No
67	CI4EDM1	-0.04	Females	No
68	CI4EPM1	-0.20	Females	No
69	CI4EPM2	-0.05	Females	No
70	CI4ERM1	-0.13	Females	No
71	CI4FFM1	0.31	Males	Yes
72	CI4FKC1	-0.07	Females	No
73	CI4FKM2	0.21	Males	No
74	CI4FRM1	0.14	Males	No
75	CI4FTM1	0.15	Males	No
76	CI4FWC2	0.00	Females	No
77	CI4GCM1	0.07	Males	No
78	CI4GLM1	0.20	Males	No
79	CI4INM1	-0.74	Females	Yes
80	CI4IPM1	0.30	Males	No

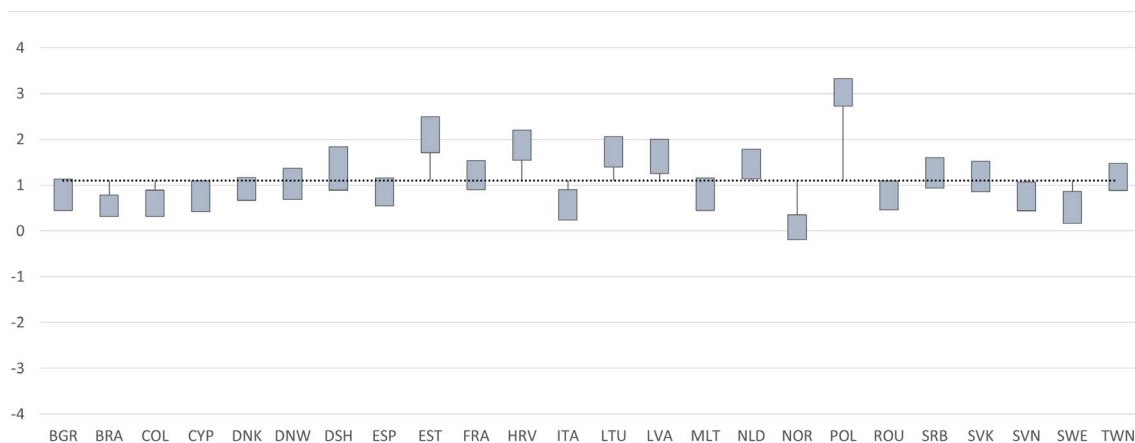
Item No.	Item name	Difference between groups	Easier for	Difference larger than 0.3 logits
81	CI4JCM1	0.00	Males	No
82	CI4JCM2	-0.17	Females	No
83	CI4JSM1	0.13	Males	No
84	CI4JSM2	-0.11	Females	No
85	CI4LEM1	-0.19	Females	No
86	CI4LEM2	0.10	Males	No
87	CI4MWM1	0.16	Males	No
88	CI4MWM2	-0.06	Females	No
89	CI4MWM3	-0.01	Females	No
90	CI4NDM1	0.09	Males	No
91	CI4NHC2	-0.04	Females	No
92	CI4NHM1	0.41	Males	Yes
93	CI4NTM1	0.12	Males	No
94	CI4NWM1	0.07	Males	No
95	CI4NWM2	-0.57	Females	Yes
96	CI4OCM1	0.15	Males	No
97	CI4OCM2	0.04	Males	No
98	CI4PDM1	0.42	Males	Yes
99	CI4PDM2	0.24	Males	No
100	CI4PIM1	-0.21	Females	No
101	CI4PRM1	-0.04	Females	No
102	CI4RCM1	-0.22	Females	No
103	CI4REM1	0.95	Males	Yes
104	CI4RFM1	0.05	Males	No
105	CI4RPM1	0.03	Males	No
106	CI4SCM1	0.19	Males	No
107	CI4SDM1	0.30	Males	Yes
108	CI4SDM2	-0.29	Females	No
109	CI4SDM3	0.14	Males	No
110	CI4SJM1	-0.12	Females	No
111	CI4SLC3	0.05	Males	No
112	CI4SLC4	-0.08	Females	No
113	CI4SLM1	0.14	Males	No
114	CI4SSM1	0.10	Males	No
115	CI4SWM1	-0.11	Females	No
116	CI4TDM1	0.14	Males	No
117	CI4TXM1	0.03	Males	No
118	CI4UWM1	-0.11	Females	No
119	CI4VRM1	0.00	Females	No
120	CI4VRM2	0.10	Males	No
121	CI4ZEC2_R	-0.39	Females	Yes
122	CI4ZEC4	0.01	Males	No
123	CI4ZEC5	-0.02	Females	No
124	CI4ZED3L_R	0.32	Males	Yes

Item No.	Item name	Difference between groups	Easier for	Difference larger than 0.3 logits
125	CI4ZEM1	0.02	Males	No
126	CI4ZHC1_R	-0.01	Females	No
127	CI4ZHC3	0.08	Males	No
128	CI4ZHC4	-0.05	Females	No
129	CI4ZHL2_R	-0.11	Females	No
130	CI4ZHM5	-0.03	Females	No
131	CI4ZSC4	-0.16	Females	No
132	CI4ZSC5N	-0.01	Females	No
133	CI4ZSC5Y	0.04	Males	No
134	CI4ZSD2	0.27	Males	No
135	CI4ZSL3_R	0.28	Males	No
136	CI4ZSM1	0.03	Males	No

Cross-national measurement equivalence

With any test used to assess student ability 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 correctly answering these questions. Items with considerable lack of measurement invariance across countries should undergo careful examination as they may indicate translation deviations or being not appropriate to national contexts. For the main survey analyses of ICCS 2022 test items, we compared national calibrations with international item difficulty parameters in order to assess the occurrence of item-by-country interaction. We then computed confidence intervals for each national item difficulty parameter, basing the computation on the respective standard errors and adjusting them for possible design effects and for multiple comparisons.

Figure 11.3: Example of item-by-country interaction graph for Item CI4SLM2



As an example, item CI4SLM2 showed significant item-by-country interaction in about half of the countries (Figure 11.3), indicating that there was considerable lack of measurement invariance for this item, which was not retained for scaling. We produced similar graphs for the test-item adjudication process at the international and national levels; information about occurrence of measurement invariance across countries was used to identify items for post-verification checks after completion of the main data collection.

Although most of the ICCS 2022 test items showed only limited item-by-country interactions, it needs to be acknowledged that there were some national item difficulties that deviated quite considerably from the international item difficulty. In previous cycles, we omitted these items from scaling for those national samples where larger deviations had been observed (1.3 logits or more).⁶ However, we conducted a review

⁶ This threshold was introduced in ICCS 2009 as being equivalent to about two standard deviations in the overall distribution of item parameters.

of the extent to which such deletions have an impact on scaling outcomes that showed only negligible differences. Therefore, for ICCS 2022, it was decided not to exclude any items based only on item DIF. Indications of deviations in item difficulty were extensively used to review possible issues with translation.

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. In ICCS 2022, we did not detect any systematic differences in relative item difficulties for scored items and kept all scored items for scaling at the national level.

Item adjudication

We conducted the adjudication of test items in two phases: first at the international level for the pooled ICCS international sample including all data from countries that had met IEA sample participation guidelines, 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 open-ended items, we also considered scorer reliability and the correct ordering of average ability estimates per category. Only one of the test items (CI4SLM2) had unsatisfactory scaling properties at the international level and was removed from the international scaling of civic knowledge.

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).

National centers were provided with item statistics (see example in Figure 11.4) and were asked to review flagged (“dodgy”) test items. For convenience, a summary of all the flagged test items for the country was also provided (Table 11.6), as well as an international summary of flagged items (not shown). 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.

Table 11.6: Summary of problematic items provided to national centers

ICCS 2022 Main Study, Dodgy Items				International Cognitive			BGR
	Item by Country Interactions			Discrimination			
	No of Valid Responses	Easier than Expected	Harder than Expected	Non-key PB is Positive	Key PB is Negative	low discrimination	Ability not Ordered
CI2CEM2	681	☐	☐	☐	☐	☒	☐
CI3CPO2	668	☐	☒	☐	☐	☐	☐
CI4ERM1	667	☐	☒	☐	☐	☐	☐
CI4FTM1	671	☐	☐	☐	☐	☒	☐
CI4JCM1	685	☐	☒	☐	☐	☐	☐
CI4NDM1	667	☒	☐	☐	☐	☐	☐
CI4SLM2	687	☐	☐	☐	☐	☒	☐
CI4SWM1	669	☒	☐	☐	☐	☐	☐
CI4ZHM5	669	☐	☒	☐	☐	☐	☐
CI4ZSD2	661	☐	☐	☐	☐	☒	☐

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 further 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 based on identified translation issues. In cases where items were translated differently from ICCS 2016, we also excluded those national items from the trend estimations but retained them in the international database. Table 11.7 lists those test items that were identified with translation issues and their corresponding treatment.

Table 11.7: National items adjudication

Country	Item	Treatment	Reason
France	CI4FKM2	Removed from database	Translation deviation
North Rhine-Westphalia/ Schleswig-Holstein	CI4INM1	Removed from database	Translation deviation
Latvia	CI2WFO2 CI4FKM2	Removed from trend scaling Removed from database	Translation changed from ICCS 2016 Translation deviation
Netherlands	CI4SWM1	Removed from database	Translation deviation
Norway	CI4SSM1	Removed from database	Translation deviation
Poland	CI4SLM2	Removed from database	Translation deviation
Slovenia	CI3DDM1	Removed from database	Translation deviation
Sweden	CI2FDM1 CI3PRO1	Removed from trend scaling Removed from trend scaling	Translation changed from ICCS 2016 Translation changed from ICCS 2016

11.4 Final scaling and equating procedures

International item calibration and test reliability

Item parameter estimates were obtained from a joint data file that included response data from both assessment modes used in ICCS 2022. National samples from countries meeting IEA sample participation requirements were equally weighted for the calibration and all items were included (except for items that were deleted nationally or internationally following the adjudication process). To balance the impact of modes in this calibration, we also used bridging study data from 11 countries. Countries with both paper-

based and computer-based data received modified weights so that their contribution was the same as for countries that only had one national sample (from either paper- or computer-based assessments) in the calibration. The mode effect was estimated based on comparisons with common item parameters across PBA and CBA. See von Davier et al. (2019) for more information on mode effect models.

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. Using this calibration, we identified a set of item parameters that we used to scale the ICCS 2022 test data (Table 11.8). The overall reliability of the international test, as obtained from the scaling model, was 0.85 (ACER ConQuest estimate).

Table 11.8: Final set of ICCS 2022 item parameters used to scale international test items

Item no.	Item name	Item parameter	Step 1	Item no.	Item name	Item parameter	Step 1	Item no.	Item name	Item parameter	Step 1	Step 2
1	CI3MAM1	-0.99		47	CI3CAM1	0.77		93	CI4SLC3	1.54		
2	CI4JSM1	0.41		48	CI4NHC2	1.20	-0.83	94	CI4SLC4	1.76		
3	CI4JSM2	-0.33		49	CI4DDM1	-0.12		95	CI314M1	-0.71		
4	CI2WFO2	1.01	0.70	50	CI4DDM2	-1.13		96	CI2CEM1	-0.80		
5	CI4SDM1	-0.79		51	CI3LPM1	-0.85		97	CI2CEM2	0.66		
6	CI4SDM2	-0.61		52	CI4LEM1	-0.20		98	CI3EPM1	-0.55		
7	CI4SDM3	-0.34		53	CI4LEM2	0.52		99	CI4CZM1	-0.28		
8	CI2ECM1	-0.61		54	CI3CSM1	-0.15		100	CI2ORM1	-0.44		
9	CI2ECM2	-0.05		55	CI312M1	-0.04		101	CI3LSM1	-0.56		
10	CI308M1	-0.72		56	CI4SSM1	-0.28		102	CI3PRO1	1.07	-0.85	
11	CI4SJM1	0.03		57	CI3PAM1	-0.35		103	CI3SDM1	0.08		
12	CI3RRM1	-0.99		58	CI3ULM1	0.25		104	CI4JCM1	0.79		
13	CI2REM2	-0.10		59	CI2BIO1	2.19	-0.91	105	CI4JCM2	-0.09		
14	CI2REM3	-0.92		60	CI4MWM1	-0.30		106	CI4PIM1	-0.69		
15	CI4REM1	-0.34		61	CI4MWM2	0.08		107	CI4RPM1	0.07		
16	CI4FKC1	0.95	-0.89	62	CI4MWM3	0.75		108	CI4EDM1	-0.29		
17	CI4FKM2	-0.60		63	CI4NWM1	0.02		109	CI3PEM1	0.18		
18	CI2CCM1	-0.52		64	CI4NWM2	-1.86		110	CI3LTM1	0.68		
19	CI2CCM2	0.24		65	CI3DDM1	0.03		111	CI3IEM1	-0.36		
20	CI4EPM1	-0.90		66	CI3GMM1	-0.05		112	CI4FTM1	0.77		
21	CI4EPM2	-0.32		67	CI4IPM1	-0.25		113	CI3MDM1	0.08		
22	CI4CFM1	1.24		68	CI3NPM1	-2.05		114	CI2CNM1	-0.23		
23	CI2FSM1	-0.16		69	CI3NPM2	-1.49		115	CI2CNM2	-1.43		
24	CI4TDM1	-0.56		70	CI4ERM1	-0.57		116	CI4FFM1	-0.37		
25	CI2PGM1	-0.06		71	CI4GLM1	0.83		117	CI4EDC2	0.56	-1.00	
26	CI2PGM2	1.06		72	CI4GCM1	0.47		118	CI4SWM1	-0.68		
27	CI4FWC2	0.30	-0.82	73	CI4DVC1	1.02	-0.84	119	CI4SCM1	0.66		
28	CI4NDM1	-0.36		74	CI2GLM1	-0.29		120	CI4UWM1	-0.43		
29	CI2BPM1	-0.12		75	CI2GLM2	0.08		121	CI4ZHC1_R	0.39	0.04	
30	CI2BPM2	-0.24		76	CI3SCM1	-0.94		122	CI4ZHL2_R	1.42	-1.56	0.0532
31	CI4OCM1	0.59		77	CI4EAM1	-0.51		123	CI4ZHC3	0.70		
32	CI4OCM2	0.21		78	CI4INM1	-2.21		124	CI4ZHC4	1.48		
33	CI4NTM1	-0.36		79	CI2ASM1	0.16		125	CI4ZHM5	-0.86		
34	CI3PLM1	-1.47		80	CI2ASM2	-0.69		126	CI4ZEM1	-1.37		
35	CI4ECM1	0.67		81	CI3UHM1	-0.40		127	CI4ZEC2_R	0.60		
36	CI4ECM2	0.53		82	CI4RFM1	-0.50		128	CI4ZED3L_R	1.41	-1.11	
37	CI4ECM3	0.52		83	CI3CPO1	0.21		129	CI4ZEC4	2.17		
38	CI2FDM1	-0.40		84	CI3CPO2	0.74		130	CI4ZEC5	1.57	-0.78	
39	CI2WFO1	0.94	-0.91	85	CI4PDM1	0.66		131	CI4ZSM1	-0.18		
40	CI2RCM1	-0.29		86	CI4PDM2	-0.07		132	CI4ZSD2	2.93		
41	CI4TXM1	-0.96		87	CI4FRM1	0.02		133	CI4ZSL3_R	-0.32	-0.48	
42	CI4VRM1	0.22		88	CI4RCM1	0.10		134	CI4ZSC4	0.36		
43	CI4VRM2	-0.02		89	CI3VGM1	-1.59		135	CI4ZSC5Y	1.10		
44	CI3IVM1	-0.55		90	CI4NHM1	-0.77		136	CI4ZSC5N	1.70		
45	CI4PRM1	-0.63		91	CI4DCM1	-0.39						
46	CI2PRM1	0.79		92	CI4SLM1	-0.55						

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 2022, 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 et al., 2009).

Using item parameters anchored at their estimated values from the international calibration, we generated plausible values as random draws $\hat{\theta}_{ng}$

$$\hat{\theta}_{ng} \sim N \left(E_g(\theta | x_n; z_n), \sqrt{V_g(\theta | x_n; z_n)} \right)$$

that depend on response data x_n (the conditional item response model), other available context data z_n (conditioning variables) and group membership g , which in the case of ICCS 2022 is determined by the education system in which an individual was assessed. Here, $E_g(\theta | x_n; z_n)$ represents the posterior distribution, while $V_g(\theta | x_n; z_n)$ denotes the posterior variance, which provides a measure of uncertainty (see more detailed descriptions in Adams et al., 1997; Adams, 2002; von Davier et al., 2020). In order to obtain plausible values as estimates of students' civic knowledge, we used the ACER ConQuest software (see Adams et al., 2020). Group membership g was accounted for by drawing plausible values separately for each national sample.

We used all available international student questionnaire variables and also those derived from regional instruments for conditioning. 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). The numbers of conditioning variables varied across countries depending on the availability of information from the international student questionnaire (core and optional variables) or additional regional data (from the European or Latin American questionnaire).

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 averages of classroom ability estimates (WLEs), adjusted for the contribution of a student's own ability estimate.

Review of delivery mode effects

As ICCS 2022 used both PBA and CBA, it has been necessary to estimate the effect of the delivery mode on student performance. This was done to support decisions about the best way to report ICCS 2022 data regardless of delivery mode within countries, after setting it on the ICCS historical scale established in ICCS 2009, that was (like ICCS 2016) completed using PBA.

Eleven countries that participated in previous ICCS cycles on paper participated in ICCS 2022 using CBA. In 2022, CBA countries that had also participated in ICCS 2016 also collected data for an additional representative sample of students on paper as part of a *bridging study*.

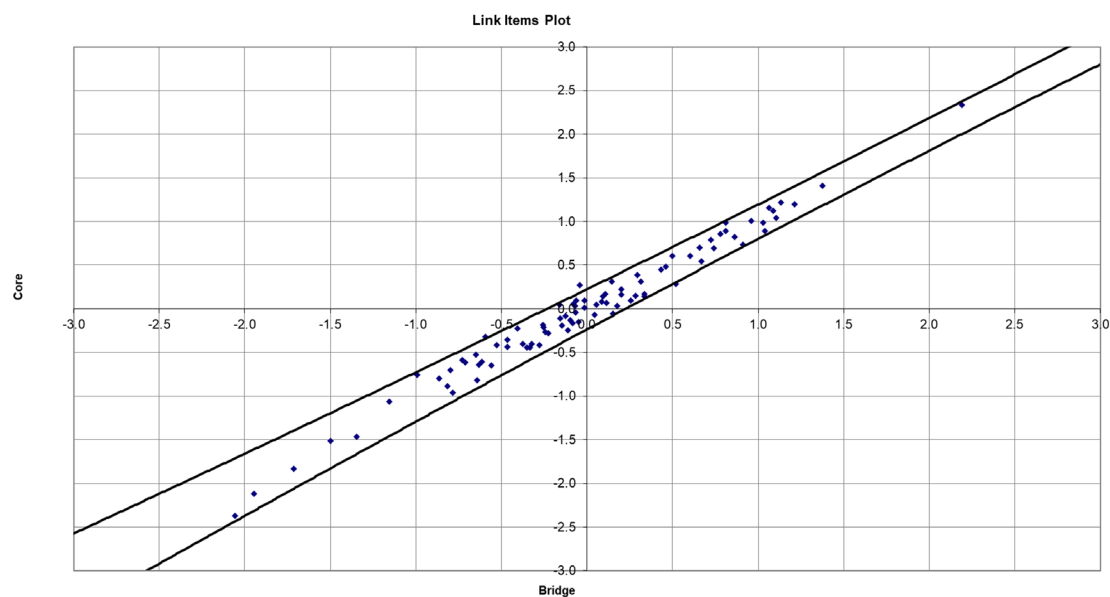
The bridging study instrument included eight of the 11 clusters common to CBA and PBA delivery across countries presented using a partially balanced incomplete block design (pBIB) distributed among eight paper-based bridging test forms. The smaller number of PBA bridging study clusters and test forms was used to minimize the necessary size of the bridging sample of students within countries while, maintaining a sufficient breadth of item coverage within the PBA-bridging instrument content to adequately reflect the ICCS civic knowledge construct.

The two item sets (CBA/core and PBA/bridging study) were calibrated separately, once with CBA and once with PBA-bridging study data. Figure 11.5 shows the correspondence of item parameters estimated based on CBA and PBA-bridging study data, where both sets of parameters have the same average difficulty of 0. The scatterplot suggests that there are only minor differences in terms of item parameters estimated with each respective sub-sample. The two lines indicate the degree to which estimated item parameters are

within the expected range due to sample variation (adjusted for an assumed sampling design effect of 4⁷), and all items appear within or (in a few cases) only slightly outside this range.

The data shown in Figure 11.5 suggested that the item difficulty parameters from the computer-based main survey and paper-based bridging study were well aligned and provided a sound basis for supporting the equating procedures required to report all ICCS 2022 data on the 'historical' ICCS civic knowledge scale established in ICCS 2009. The curved lines were computed to provide rough confidence intervals based on the original standard errors adjusted for approximate sampling design effects.

Figure 11.5 Scatterplot of item parameters for CBA and PBA bridging study items



We further looked at differences in outcomes across the two delivery modes. Initial analyses showed that there were more considerable differences between the (relatively small) sub-samples in Latvia that were caused by some 'outlier' schools in the bridging study sample. In liaison with the Latvian national team, we identified those schools and applied a re-weighting procedure to neutralize these effects, whereby seven out of 51 schools in the bridging sample received zero weights with weight adjustments for other schools where appropriate. The Latvian data from the bridging study and main survey were obtained from separate school samples.

Using a dataset including the Latvian data with modified sample weights, we then computed the national averages of percentages with correct responses within each sample for the 11 countries that had implemented the bridging study and compared them across the two delivery modes (Table 11.9). The results showed that, on average, there were slightly higher percentages in samples from paper-based assessment, the difference was on average about one percentage point. Further, we also observed some variation across countries, ranging from zero differences to 2-3 point-differences in Italy, Latvia, and Norway, a phenomenon that has also been observed in other international studies of mode effects (see, for example, von Davier et al., 2020). In view of the slight negative overall effect of computer mode in comparison with a paper-based assessment, it was decided to apply a mode effect adjustment as part of the final scaling and equating procedures.

⁷ Approximation of average design effect across countries.

Table 11.9: Average percentage correct for CBA and PBA data

Country	Percentages of correct responses		
	Core	Bridge	Difference
Bulgaria	54 (1.0)	54 (1.7)	0.3 (2.0)
Chinese Taipei	76 (0.4)	77 (0.7)	-1.1 (0.8)
Colombia	54 (0.7)	55 (1.9)	-1.3 (2.0)
Estonia	69 (1.0)	70 (1.0)	-0.4 (1.4)
Italy	67 (0.6)	70 (0.8)	-2.4 (1.0)
Latvia ¹	60 (0.5)	63 (1.0)	-2.6 (1.1)
Lithuania	63 (0.7)	64 (1.0)	-0.3 (1.3)
Malta	60 (1.4)	61 (1.3)	-0.9 (1.9)
Norway (9) ¹	68 (0.5)	70 (0.7)	-2.1 (0.9)
Slovenia	64 (0.5)	64 (0.7)	-0.2 (0.9)
Sweden ¹	74 (0.6)	74 (0.7)	-0.5 (0.9)
Average	64 (0.2)	65 (0.3)	-1.0 (0.4)

Notes:

() Standard errors appear in parentheses.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

Equating test scores

For ICCS 2009 (Schulz & Fraillon, 2011), 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

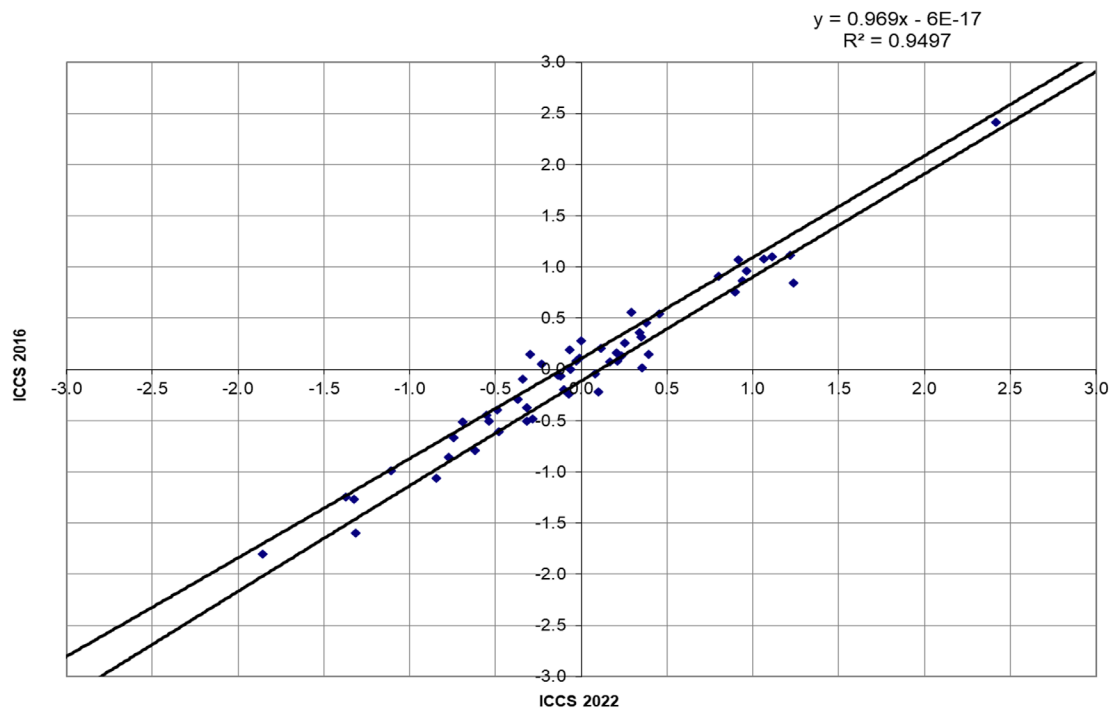
$$\hat{\theta}'_n = 500 + 100 \left(\frac{\hat{\theta}^{09}_n - \bar{\theta}}{\sigma_{\theta}} \right)$$

where $\hat{\theta}'_n$ were the student civic knowledge estimates (plausible values) in the international reporting metric, $\hat{\theta}^{09}_n$ were the original estimates (in logits), $\bar{\theta}$ and σ_{θ} were the corresponding ICCS 2009 mean (-0.01) and standard deviation (0.95) computed with weights assigning equal weight to each national sample. This transformation was applied to each of the five plausible values.

In 2016, the international project team conducted a concurrent calibration of ICCS 2009 and 2016 data on the ICCS 2016 scale using all items from both cycles. All ICCS 2009 item parameters were re-estimated concurrently during the ICCS 2016 calibration process and used to re-scale the ICCS 2009 data to obtain a linear transformation that was applied to set the ICCS 2016 estimates to the reporting metric established in the first cycle (Gebhardt & Schulz, 2018). Given the dual delivery mode including both paper- and computer-based assessment across different countries in a design where not all trend items were included in the bridging study, it was not possible to directly apply the same linking method as in the previous cycle.

Before implementing our linking procedures, 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 55 common items (Figure 11.6) revealed that most items were within or relatively close to the confidence interval that was computed around the assumption of measurement invariance across cycles. None of the differences between relative items difficulties were more than half a logit and there were no obvious outliers. Therefore, we decided to keep all the common items as link items after removing all national items for which we had detected discrepancies in translation between 2016 and 2022.

Figure 11.6 Scatterplot for link item parameter estimates from ICCS 2022 and ICCS 2016



In addition to the calibration of ICCS 2022 test items based on the combined dataset of paper- and computer-based test data (see Table 11.7), we also conducted a joint calibration of ICCS 2016 and paper-based ICCS 2022 data from the current cycle that included national core samples from six countries and bridging study data from eleven countries. All national samples were given equal weights and only ICCS 2016 data from countries participating in both cycles were included. This calibration was done to ensure that the trend estimation was based on the same delivery mode and the parameters for the items used in ICCS 2022 are displayed in Table 11.10.

Table 11.10: ICCS 2022 item parameters from joint calibration with ICCS 2016 data

Item no.	Item name	Item parameter	Step 1	Item no.	Item name	Item parameter	Step 1	Item no.	Item name	Item parameter	Step 1
1	CI3MAM1	-0.99		41	CI4TXM1	-0.81		81	CI3UHM1	-0.34	
2	CI4JSM1	0.55		42	CI4VRM1	0.37		82	CI4RFM1	-0.37	
3	CI4JSM2	-0.28		43	CI4VRM2	0.15		83	CI3CPO1	0.43	
4	CI2WFO2	1.16	0.73	44	CI3IVM1	-0.46		84	CI3CPO2	1.02	
5	CI4SDM1	-0.75		45	CI4PRM1	-0.51		85	CI4PDM1	0.79	
6	CI4SDM2	-0.46		46	CI2PRM1	1.01		86	CI4PDM2	0.09	
7	CI4SDM3	-0.24		47	CI3CAM1	0.84		87	CI4FRM1	0.19	
8	CI2ECM1	-0.58		48	CI4NHC2	1.37	-0.89	88	CI4RCM1	0.30	
9	CI2ECM2	0.10		49	CI4DDM1	0.03		89	CI3VGM1	-1.36	
10	CI308M1	-0.59		50	CI4DDM2	-1.05		90	CI4NHM1	-0.55	
11	CI4SJM1	0.01		51	CI3LPM1	-0.84		91	CI4DCM1	-0.15	
12	CI3RRM1	-0.92		52	CI4LEM1	-0.12		92	CI4SLM1	-0.34	
13	CI2REM2	-0.14		53	CI4LEM2	0.62		93	CI4SLC3	1.76	
14	CI2REM3	-0.78		54	CI3CSM1	-0.12		94	CI4SLC4	1.98	
15	CI4REM1	-0.17		55	CI312M1	0.05		95	CI314M1	-0.52	
16	CI4FKC1	1.03	-0.87	56	CI4SSM1	-0.12		96	CI2CEM1	-0.62	
17	CI4FKM2	-0.47		57	CI3PAM1	-0.25		97	CI2CEM2	0.75	
18	CI2CCM1	-0.30		58	CI3ULM1	0.52		98	CI3EPM1	-0.54	
19	CI2CCM2	0.47		59	CI2BIO1	2.44	-0.90	99	CI4CZM1	0.03	
20	CI4EPM1	-0.78		60	CI4MWM1	-0.16		100	CI2ORM1	0.04	
21	CI4EPM2	-0.18		61	CI4MWM2	0.14		101	CI3LSM1	-0.25	
22	CI4CFM1	1.24		62	CI4MWM3	0.89		102	CI3PRO1	0.93	-0.83
23	CI2FSM1	-0.05		63	CI4NWM1	0.13		103	CI3SDM1	0.36	
24	CI4TDM1	-0.44		64	CI4NWM2	-1.72		104	CI4JCM1	1.06	
25	CI2PGM1	0.06		65	CI3DDM1	0.18		105	CI4JCM2	0.41	
26	CI2PGM2	1.16		66	CI3GMM1	0.08		106	CI4PIM1	-0.69	
27	CI4FWC2	0.44	-0.83	67	CI4IPM1	-0.05		107	CI4RPM1	-0.04	
28	CI4NDM1	-0.24		68	CI3NPM1	-1.91		108	CI4EDM1	-0.09	
29	CI2BPM1	0.17		69	CI3NPM2	-1.31		109	CI3PEM1	0.10	
30	CI2BPM2	-0.15		70	CI4ERM1	-0.40		110	CI3LTM1	1.07	
31	CI4OCM1	0.70		71	CI4GLM1	1.05		111	CI3IEM1	-0.06	
32	CI4OCM2	0.29		72	CI4GCM1	0.63		112	CI4FTM1	0.88	
33	CI4NTM1	-0.28		73	CI4DVC1	1.12	-0.86	113	CI3MDM1	0.23	
34	CI3PLM1	-1.52		74	CI2GLM1	-0.20		114	CI2CNM1	0.06	
35	CI4ECM1	0.74		75	CI2GLM2	0.22		115	CI2CNM2	-1.27	
36	CI4ECM2	0.65		76	CI3SCM1	-0.78		116	CI4FFM1	-0.10	
37	CI4ECM3	0.57		77	CI4EAM1	-0.38		117	CI4EDC2	0.75	-1.07
38	CI2FDM1	-0.15		78	CI4INM1	-2.04		118	CI4SWM1	-0.55	
39	CI2WFO1	1.09	-0.87	79	CI2ASM1	0.20		119	CI4SCM1	0.71	
40	CI2RCM1	-0.15		80	CI2ASM2	-0.49		120	CI4UWM1	-0.43	

The scaling, equating, and mode effect procedures were implemented in three steps, each of which consisted of obtaining three different linear transformations (T1, T2, and T3) that were applied to civic knowledge estimates (plausible values, $\hat{\theta}_n$ to transform them to estimates in a different metric ($\hat{\theta}'_n$) using different intercept (A1, A2, and A3) and slope parameters (B1, B2, and B3) applying the formula:

$$\hat{\theta}'_n = A_i + B_i * \hat{\theta}_n$$

The procedures consisted of the following steps:

- I. Are-scaling of ICCS 2016 data (using data from countries also participating in ICCS 2022) through the generation of PVs with full conditioning based on the joint calibration of paper-based test data from ICCS 2022 and ICCS 2016. Based on this new set of PVs and the original PVs used for reporting ICCS 2016 results we computed the linear transformation T1 that transformed the paper-based ICCS 2022 test data to be on the historical civic knowledge scales ($A1 = 434.521$, $B1 = 98.138$).
- II. Paper-based ICCS 2022 data (from core and bridging study samples) were used to draw to sets of plausible values with full conditioning, the first set of estimates ($PV_{PBA2016/22}$) was scaled based on the joint calibration based on ICCS 2022 and 2016 data, the second set ($PV_{PBA2022}$) was scaled based on the ICCS 2022 calibration (using a combined database of all computer-based and paper-based ICCS 2022 data including bridging study data). These sets of PVs from both calibrations were used to compute a second linear transformation that aligned the ICCS 2022/2016 paper-based calibration with the calibration based on all ICCS 2022 test data ($A2 = 0.129$, $B2 = 1.004$).
- III. Computer-based ICCS 2022 test data were scaled using the ICCS 2022 calibration through the computation of plausible values ($PV_{CBA2022}$) with full conditioning. For the 11 countries with plausible values from computer-based ($PV_{CBA2022}$) and paper-based assessments ($PV_{PBA2022}$), we then computed

the third linear transformation (T3), which was applied for computer-based civic knowledge data to adjust for the observed overall mode effect ($A3 = 0.033$, $B3 = 1.021$).

Estimates of civic knowledge ($\hat{\theta}_n^{P22}$) from paper-based assessment based on the combined ICCS 2022 calibration ($PV_{PBA2022}$) were set on the historical civic knowledge applying the following transformation:

$$\hat{\theta}_n^E = 438.51 + 98.138 \times (0.129 + 1.004 * \hat{\theta}_n^{P22})$$

Estimates of civic knowledge ($\hat{\theta}_n^{C22}$) based on computer-based assessments using the combined ICCS 2022 calibration ($PV_{CBA2022}$) were equated to previous cycles applying the following transformation, which included the mode effect adjustment:

$$\hat{\theta}_n^E = 438.51 + 98.138 \times (0.129 + 1.004 * (0.033 + 1.021 * \hat{\theta}_n^{C22}))$$

Uncertainty in the link

Because the transformation equating the ICCS 2022 data with the ICCS 2016 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 in part 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. It should be noted that there are also other factors that may cause uncertainty in the link (such as model error), and that the chosen approach described in the following paragraph constitutes only an approximation.

The uncertainty resulting from link-item sampling is often referred to as linking error, and it is an error that analysts should consider 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 *component of the uncertainty*.

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. Use i to index items in a unit and j to index units so that $\hat{\delta}_{ij}^y$ is the estimated difficulty of item i in unit j for year y , and let:

$$c_{ij} = \hat{\delta}_{ij}^{2022} - \hat{\delta}_{ij}^{2016}$$

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

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

Further let:

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

and then the link error, considering the clustering was computed as follows:

$$LinkError_{2022,2016} = \sqrt{\frac{\sum_{j=1}^K m_j^2 (c_{\cdot j} - \bar{c})^2}{K(K-1)\bar{m}^2}} = \frac{\sum_{j=1}^K m_j^2 (c_{\cdot j} - \bar{c})^2}{L^2} \frac{K}{K-1}$$

It is important to note that this way of estimating the linking error is only appropriate for the Rasch model, which was used in this study. The procedure does not consider changes in item functions as reflected in parameter changes in more general models. The linking error was computed after setting item parameters from 2022 and 2016 on the same scale (by shifting link items so that the average of item parameters were 0 for both item sets).

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 ongoing 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 comparisons of survey results across study cycles. In the second ICCS cycle, we applied some changes to the way ICCS proficiency were described (see Schulz et al., 2018).

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 ability 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?

In ICCS 2009, 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.⁸

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.

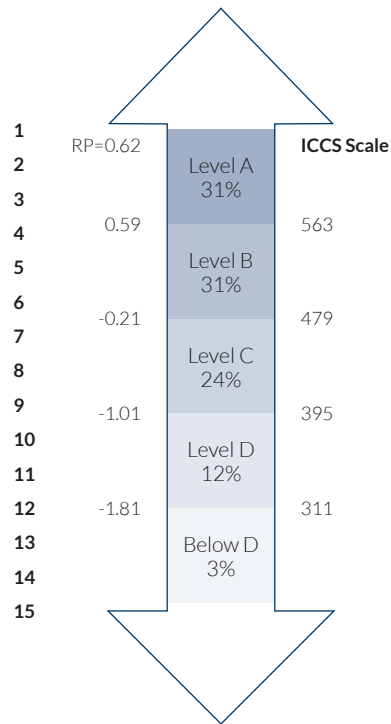
The approach that we chose was essentially an attempt to apply an appropriate choice of mastery by placing item locations at $rp = 0.62$ while simultaneously ensuring that the approach would be understood by the readers of ICCS reports.⁹

In ICCS 2016, we identified four proficiency levels that could be used when reporting student performances from the assessment, extending the original three levels that had been applied in ICCS 2009. In ICCS 2022, we used the same proficiency levels as in ICCS 2016, and Figure 11.7 shows the cut-points for these levels and the percentage of students at each proficiency level across the participating ICCS 2022 countries.

⁸ If the variance of ability estimates (in logits) is 1, five levels of width 0.8 logits will cover the range -2 to 2 logits, or roughly 95 percent of student abilities. With a variance of 1.3 as observed in ICCS 2009, six levels would cover approximately the same range.

⁹ A description of this is given in OECD (2002), PISA 2000 Technical Report, OECD.

Figure 11.7: Proficiency level cut-points and percentage of students at each level



When reporting released test items and mapping them against proficiency levels, we transformed 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 was obtained by applying the following transformation to the item difficulty in its original metric (δ_i):

$$\delta'_i = 438.51 + 98.138 \times \left(0.129 + 1.004 * (\delta_i + \ln \left(\frac{0.62}{0.38} \right)) \right)$$

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). 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*. Boston College.
- Adams, R. J., Wu, M. L., Cloney, D., Berezner, A., & Wilson, M. (2020). *ACER ConQuest: Generalised Item Response Modelling Software* (Version 5.29) [Computer software]. Australian Council for Educational Research. <https://www.acer.org/au/conquest>
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.
- Gebhardt, E. & Schulz, W. (2018). Scaling procedures for ICCS 2016 test items. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.), *ICCS 2016 technical report* (pp. 117–137). International Association for the Evaluation of Educational Achievement (IEA).
- Kroehne, U., Buerger, S., Hahnel, C., & Goldhammer, F. (2019). Construct equivalence of PISA reading comprehension measured with paper-based and computer-based assessments. *Educational Measurement: Issues and Practice*, 38(3), 97–111. <https://doi.org/10.1111/emip.12280>
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56(2), 177–196. <https://doi.org/10.1007/BF02294457>
- 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). 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.
- Rost, J., & von Davier, M. (1994). A Conditional Item-Fit Index for Rasch Models. *Applied Psychological Measurement*, 18(2), 171–182. <https://doi.org/10.1177/014662169401800206>
- 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*. Springer. <https://doi.org/10.1007/978-3-319-73963-2>
- Schulz, W. & Fraillon, J. (2011). Scaling procedures for ICCS test items. In W. Schulz, J. Ainley, & J. Fraillon (Eds.), *ICCS 2009 technical report* (pp. 129–156). International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Fraillon, J., Losito, B., Agrusti, G., Ainley, J., Damiani, V., & Friedman, T. (2023). *IEA International Civic and Citizenship Education Study 2022 assessment framework*. Springer. <https://doi.org/10.1007/978-3-031-20113-4>
- von Davier, M., & Carstensen, C. H. (2007). *Multivariate and mixture distribution Rasch models*. Springer Science+ Business Media, LLC.
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, 2, 9–36.
- von Davier, M., Foy, P., Martin, M., & Mullis, I.V.S. (2020). Examining eTIMSS country differences between eTIMSS data and bridging study data. A Look at country-level mode of administration effects. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report*. Retrieved

from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/methods>

von Davier, M., Gonzalez, E., & Schulz, W. (2020). Validity and international comparisons: The statistical roots of measurement as used in international educational surveys. In H. Wagemaker (Ed.), *Validity and reliability: quality control strategies in IEA's large scale assessments* (pp. 187–219). Springer.

von Davier, M., Khorramdel, L., He, Q., Shin, H. J., & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics*, 44(6), 671–705. <https://doi.org/10.3102/1076998619881789>

Wilson, M. & Draney, K. (2002). A technique for setting standards and maintaining them over time. In S. Nishisato, Y. Baba, H. Bozdogan, K. Kanefuji (Eds.), *Measurement and multivariate analysis* (pp. 325–332). Springer. https://doi.org/10.1007/978-4-431-65955-6_35

Wright, B., & Masters, G. (1982). *Rating scale analysis*. MESA Press.

Chapter 12:

Scaling procedures for ICCS 2022 questionnaire items

Wolfram Schulz and Tim Friedman

12.1 Introduction

This chapter describes the scaling procedures used for scaling questionnaire scales that were reported in the International Civic and Citizenship Education Study (ICCS) 2022 publications (Schulz et al., 2023; Damiani et al., 2024). 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 the Rasch model to derive scale scores from items with two or more categories.

The first part of this chapter describes the simple indices that were derived from ICCS 2022 data and how they were created. The second part outlines the procedures used for the scaling of questionnaire data in ICCS 2022. The third part includes a description of scaled indices based on student, teacher, and school level questionnaire data, providing statistical information on the factor structure of related item sets, scale reliabilities, and parameters used for IRT (Item Response Theory) scaling. The chapter concludes with a section about the comparability of questionnaire data from computer-based and paper-based delivery modes.

Results from an analysis of cross-country validity of item dimensionality and constructs were already part of ICCS 2022 field trial analyses, as part of which the International Study Center at the Australian Council for Educational Research (ACER) conducted a review of the extent to which measurement models were invariant across participating countries for draft item material. The review of field trial data 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).

12.2 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 9 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., “2022” or “2007”), 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, choosing between ‘Girl’, ‘Boy’ and in some countries, a third option, ‘Other’ (countries had a choice whether to include this third option or not, and how it can be named). These were captured in the indicator variable *student gender* (SD_GENDER) with a value of 1 assigned to

females and value of 0 assigned to males. In cases where students had omitted data for this question, we used the information from the tracking form instead. For this indicator variable, student data corresponding to the optional third category were treated as missing information. However, the variable with the original is also included in the international database.

Question 3 asked students to 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's degree or equivalent), ISCED level 7 (master's degree or equivalent) or ISCED level 8 (doctoral degree 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 2022 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:

- (0) The language spoken at home most of the time differed from the language of assessment;
- (1) The language spoken at home most of the time was the language of assessment.

Students were asked to provide details about the jobs of their two parents or guardians¹ 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_P1ISCO (student's parent/guardian 1) and S_P2ISCO (student's parent/guardian 2). We further mapped these codes to the international socioeconomic index of occupational status (ISEI) (Ganzeboom et al., 1992). The three indices that we obtained from these scores were *occupational status of parent/guardian 1* (S_P1ISEI), *occupational status of parent/guardian 2* (S_P2ISEI), and the *highest occupational status of both parents or guardians* (S_HISEI), with the last corresponding to the higher ISEI score of either parent/guardian 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 2022 we classified levels of parental education according to the International Standard Classification of Education (ISCED) (UNESCO, 2006).

¹ The two parents were named <parent/guardian 1> and <parent/guardian 2> in the English source version, and national centers adapted them as appropriate for their national contexts.

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's degree or equivalent), ISCED level 7 (master's degree or equivalent) or ISCED level 8 (doctoral degree or equivalent).

For each student, indices with these categories are available for the *highest educational attainment of parent/guardian 1* (S_P1ISCED) and *highest educational attainment of parent/guardian 2* (S_P2ISCED). The index for *highest educational level of parental education* (S_HISCED) corresponds to the higher educational attainment of either parent or guardian.

Question 10 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).

Question 11 requested students to indicate how interested they were and how interested their parent(s) or guardian(s) were in political and social issues. The (recoded) indices for *student interest* (S_SINT), *interest of parent/guardian 1* (S_P1INT) and *interest of parent/guardian 2* (S_P2INT) 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 computing the maximum value of both S_P1INT and S_P2INT.

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 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 4 of the teacher questionnaire. We assigned “less than 25” a value of 23 and coded “60 or over” as 63.

The *sex of teacher* (TD_GENDER) was computed from the data captured from Question 5 of the teacher questionnaire. Female teachers were coded as 1, whereas male teachers were coded as 0. Teachers who selected the third option were set to missing for this indicator variable while the original data are also available in the database.

School questionnaire

Question 20 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 21 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 (IC4G22A) and the total number of girls enrolled on the same date (IC4G22B).

Question 22 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_GENROL) by adding the number of boys (IC4G22A) and the number of girls (IC4G22B) 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_GENROL) by the *total number of enrolled students at the school* (C_SCSIZE). We then multiplied this value by 100.

Question 23 collected information about the 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.

12.3 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). We investigated the spread of reliabilities across countries. 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

Confirmatory factor analysis (CFA) (Brown, 2015; Jöreskog, 1969) allows the confirmation of theoretically expected dimensions and, at the field-trial stage, the re-specification of the expected dimensional structure.² 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 more general Structural Equation Model (SEM) (see, for example, Kaplan, 2009) framework, latent variables are linked to observable variables via measurement equations.

In view of the categorical nature of the questionnaire items used for scaling, the ICCS analysis team relied on robust weighted least squares estimation (WLSMV) (see Muthén et al., 1997; Flora & Curran, 2004) to estimate the confirmatory factor models. The software package used for estimation was MPLUS 7 (Muthén & Muthén, 2012).

² In the initial stages of field trial analyses, we also employed exploratory factor analysis (Tucker & MacCallum, 1997; Fabrigar et al., 1999) to determine item dimensionality of larger item pools that had been newly developed for this cycle.

A random variable y_j reflects the observed categorical responses to item j , that take the form of $K+1$ values $0, 1, \dots, K$. The threshold model (Muthén, 1984) assumes underlying continuous responses y^* for all items, for which the observed categorical response variable y_j is discretized through thresholds τ_{jk} ($j=1, 2, \dots, p$; $k=0, 1, \dots, K+1$). Assuming $\tau_{j0}=-\infty$ and $\tau_{j(K+1)}=\infty$, and $y_j=k$ if $\tau_{jk} < y_j^* < \tau_{j(k+1)}$, the latent responses for all items y^* are modeled as:

$$y^* = v + \Lambda\xi + \delta,$$

where ξ denotes a vector of m common factors (latent variables), δ a $p \times 1$ vector of unique error variables, v a $p \times 1$ vector of intercepts, and Λ a $p \times m$ matrix of factor loadings. The covariance matrix is fitted according to the theoretical factor structure.

When conducting the confirmatory factor analyses for ICCS 2022 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 2022 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 poor, 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 et al., 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, with values in between regarded as marginally satisfactory (see Bentler & Bonett, 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_j can be interpreted in the same way as standardized correlation coefficients between the indicator variable and 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_j (these provide an estimate of the unexplained proportion of variance) as:

$$\delta_j = (1 - l_j^2).$$

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. Both unidimensional and multidimensional models were applied when confirming expected questionnaire item dimensionality, the latter was more prominent for the analysis of items that measured related latent variables or that were included in the same questionnaire item batteries. 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 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 countries, and it may not be appropriate to assume the same factor structure for each population. To test parameter invariance, multiple-group modeling, as an extension of CFA, offers an approach to test the equivalence of measurement models across sub-samples (Little, 1997; Byrne, 2008).

If one considers a model where respondents belong to different groups indexed as $g=1, 2, \dots, G$, the multiple-group factor model becomes:

$$y_g^* = v_g + \Lambda_g \xi_g + \delta_g$$

For example, a test of factorial invariance (H_A) where factor loadings are defined as being equal, often

referred to as “metric equivalence” (Horn & McArdle, 1992), can be defined as:

$$H_A: \Lambda_1 = \Lambda_1 = \Lambda_2 = \dots = \Lambda_g$$

It is possible to compare model-fit indices across different multiple-group models, each with an increasing degree of constraints, including relaxed models with no constraints through more constrained models to models with largely invariant model parameters. While such comparisons can include many different ways of constraining parameters across different groups of cases or variables, in this report, for all international and regional student questionnaire scales, three different multiple-group models for CFA were estimated with the following 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 one that indicates the fullest level of measurement comparability across participating countries. Hypothesis testing using tests of statistical significance is rather problematic (especially with data from larger samples), therefore, when reviewing results, it was deemed more appropriate to focus on relative changes in model fit across the three models with different levels of constraints. The multiple-group analyses presented in this report provide a rough indication of the extent to which measurement models tended to be invariant across countries. While there are no clear scientific criteria about the point at which lack of measurement invariance causes scale scores to become incomparable, for those scales for which a lack of measurement invariance was noted, comparisons across national contexts should be interpreted with due caution.

Item response modeling

In line with the scaling of test item data (see Chapter 11), item response modeling was used to derive scales from ICCS 2022 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:

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

Here, $P_i(\theta_n)$ 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; Masters, 1982), which takes the form of:

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

where we define $\exp \sum_{k=0}^0 (\theta_n - \delta_i + \tau_{ik}) = 1$, $P_{x_i}(\theta_n)$ denotes the probability of person n scoring x on item i , θ_n the person's latent trait, the item parameter δ_i the location of the item on the latent continuum, and τ_{ik} an additional step parameter for each step k between adjacent categories.

We used *weighted mean-square* statistics based on model residuals, as one indicator of assessing general fit of items to the scaling model (Wright & Masters, 1982). The residual-based statistics were used in conjunction with a wide range of further item statistics provided the basis for an assessment of IRT model fit. ICCS 2022 used the ACER Conquest software package (Adams et al., 2020) 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

where each of the national samples from 20 countries that met sample participation requirements for the student survey received the same weight.

- B. Calibration of item parameters for the teacher questionnaire: This was done based on a pooled database where each of the national samples from 13 countries that met sample participation requirements for the teacher survey received the same weight.
- C. Calibration of item parameters for school questionnaire: This was done based on a pooled database where each of the national samples from 20 countries that met sample participation requirements for the student survey received the same weight.
- D. Calibration of item parameters for the European questionnaire: This was done based on a pooled database where each of the national samples from 17 European countries that met sample participation requirements for the student survey received the same weight.

Following the estimation of international item parameters from the calibration sample, we computed weighted likelihood estimates of person parameters to obtain individual student scores. So-called weighted likelihood estimates (Warm, 1989) are computed by minimizing the equation:

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

for each case n , where r_x is the sum score obtained from a set of j items with m_i+1 categories and k steps. This was achieved by applying the Newton-Raphson method with fixed item parameters from the international item calibrations. 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 (Warm, 1989). ACER ConQuest software was used to calibrate the item parameters before deriving scale scores.

For scales that were new to ICCS 2022, the transformation of weighted likelihood estimates to an international metric resulted in reporting scales with an ICCS 2022 average of 50 and a standard deviation of 10 for equally weighted datasets from the countries that met sample participation requirements. This was achieved by applying the following formula:

$$\theta'_n = 50 + 10 \frac{\theta_n - \mu_{\theta(ICC22)}}{\sigma_{\theta(ICC22)}} ,$$

where θ'_n denotes the scores in the international metric, θ_n the weighted likelihood estimates in their original logit metric as derived from the estimation, and $\mu_{\theta(ICC22)}$ the international mean of these logit scores using weights that ensured equal contributions for each of the national samples. Derived using the same weight, $\sigma_{\theta(ICC22)}$ represents the corresponding international standard deviation of the original likelihood estimates in their logit metric. The means and standard deviations (Table 12.1) were used to transform the original scale scores for the international and regional student, teacher, and school questionnaires into the international metric.

Table 12.1: Transformation parameters for new ICCS 2022 questionnaire scales (means and standard deviations of original IRT logit scores)

Questionnaire	Scale	IRT logit score	
		Mean	SD
International student questionnaire	S_ATTENV	1.65	1.65
	S_COMPART	-1.49	1.08
	S_DEMTHRT	1.01	1.31
	S_ENGDM	-1.93	1.19
	S_ENREST	0.04	1.14
	S_ENVACT	0.80	2.01
	S_ENVCON	-0.69	1.43
	S_GLOBCIT	1.36	1.74
	S_IMMPOS	1.93	1.98
	S_INFDEC	0.53	1.69
	S_SCHPART	-0.61	1.09
	S_SYSCRT	0.74	1.64
	S_SYSSAT	-0.26	2.06
Teacher questionnaire	T_TCHPRT	1.50	1.76
	T_PROBSC	-2.94	1.73
	T_STDCOM	-0.19	1.72
	T_PCCLIM	0.84	4.31
	T_STDINV	0.28	1.63
	T_DIVACT	1.69	1.96
	T_POSCDIF	1.92	2.29
	T_NEGCDIF	-1.48	2.52
	T_POSSDIF	0.82	2.09
	T_NEGSDIF	-1.72	2.86
	T_ACTDIG	-0.50	1.56
	T_CITCON	0.07	1.46
	T_CITSOC	2.60	2.00
	T_GLOBCIT	2.92	2.01
	T_CIVCLAS	0.08	1.06
	T_ACTGLOB	0.26	1.51
	T_PRPCCE	2.07	1.88
	T_PDACCE	0.05	2.41
	T_OPPLRN	2.17	2.08
	T_PDATCH	1.32	1.79

Questionnaire	Scale	IRT logit score	
		Mean	SD
School questionnaire	C_TCPART	1.60	1.83
	C_STDINV	0.38	1.89
	C_PARINV	0.26	1.53
	C_COLLCOM	1.11	1.33
	C_STDCOM	0.21	1.05
	C_TCTACTD	0.15	1.43
	C_ENPRAC	0.36	0.93
	C_ACTENGC	0.65	1.54
	C_ACTDIGT	0.68	1.44
	C_AVRESCOM	1.21	1.93
	C_COMETN	-3.36	2.33
	C_COMPOV	-1.71	2.47
	C_COMCRI	-2.89	2.37
European student questionnaire	E_CCOOP	1.63	1.57
	E_COOPENV	2.48	2.07
	E_ETHCON	-0.28	1.83
	E_EUATNEG	0.77	1.63
	E_EUATPOS	1.51	1.95
	E_EUFNEG	0.47	1.33
	E_EUFPOS	0.79	1.34
	E_EULRN	0.86	1.90
	E_RESMIG	0.44	1.76
	E_SUSBEH	0.37	1.34

Linking questionnaire scale indices to previous cycles

The ICCS 2022 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 link the scale scores based on ICCS 2022 data to those from ICCS 2016 and/or 2009. Scale linking was applied to 17 scales from the international questionnaire and three scales from the regional European questionnaire.³ For some of the scales individual items had been discarded or had been augmented by new items.⁴ In these cases, the linking 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 cycle 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 2022, item parameters were calibrated using the IRT (Rasch) partial credit

³ Given that only two Latin American countries implemented the ICCS 2022 main survey, four scales from the Latin American regional questionnaire were equated through scaling using the same parameters from the previous cycle.

⁴ In all but one case, there was only one item exchanged or added.

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 2016, we used a two-step transformation process. The first step placed the ICCS 2022 logit scores on the ICCS 2016 logit metric (using linear constants denoted A_1 and B_1), while the second step (using linear constants denoted A_2 and B_2) transformed the latter to the same reporting metric as in ICCS 2016. 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.⁵ The transformation parameters in the first step were calculated on the means and standard deviations of these combined parameters as:

$$B_1 = \frac{SD_{ICCS16}}{SD_{ICCS22}}$$

$$A_1 = MN_{ICCS16} - \frac{SD_{ICCS16}}{SD_{ICCS22}} \times MN_{ICCS22}$$

MN_{ICCS22} and SD_{ICCS22} are the mean and standard deviation of the combined item location and step parameters from the calibration of ICCS 2022, respectively, while MN_{ICCS16} and SD_{ICCS16} are the corresponding (combined) parameters from the calibration based on ICCS 2016 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 2022 data. For scales that were established in ICCS 2016, the transformation parameters A_2 and B_2 were calculated as:

$$B_2 = 10 / \sigma_{\theta(ICCS2016)}$$

$$A_2 = 50 - (\mu_{\theta(ICCS2016)} \times B_2)$$

Here, $\mu_{\theta(ICCS2016)}$ is the mean and $\sigma_{\theta(ICCS2016)}$ is the standard deviation of logit scores for equally weighted countries in ICCS 2016. In case of ICCS 2022 scales that had already been equated to ICCS 2009, the transformation parameters A_2 and B_2 were the same as those used for the transformation of scale scores (A and B) that were applied to set the logit scores from ICCS 2016 to the reporting metric established in ICCS 2009, where 50 represents the average of equally weighted ICCS 2009 countries and 10 the corresponding standard deviation (see Table 11.2 in Schulz & Friedman, 2018).

For all linked scales, the final transformation of ICCS 2022 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 their original logit metric.

Given the (albeit relatively minor) changes in item parameters between the different 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 2022 and 2016 using the following formula:

$$\sigma_{(linking\_error)} = \sqrt{\frac{\sigma^2}{N}}$$

⁵ 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.

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 12.2) and was considered when estimating changes in scale scores between cycles (see Chapter 13).

Table 12.2: 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	44.67	8.28	0.24	0.89	46.66	7.33	0.033	0.240
S_CITEFF	47.78	6.48	-0.03	0.94	47.61	6.10	0.033	0.202
S_CITSOC	41.66	5.62	0.00	0.92	41.67	5.18	0.033	0.169
S_CIVLRN	46.49	7.63	0.11	1.04	47.36	7.98	0.030	0.240
S_ELECPART	42.25	4.28	0.00	0.96	42.25	4.10	0.052	0.211
S_ETHRGHT	39.53	4.87	0.00	0.92	39.53	4.50	0.071	0.318
S_GENEQL	38.55	6.37	0.00	1.09	38.55	6.93	0.047	0.329
S_INTACT	46.98	7.19	0.00	1.09	43.97	5.10	0.032	0.270
S_INTRUST	47.99	5.12	0.00	0.95	47.99	4.85	0.032	0.133
S_ILLACT	57.89	3.98	0.00	1.06	57.89	4.21	0.029	0.168
S_LEGACT	49.78	5.62	-0.03	1.02	49.63	5.73	0.025	0.212
S_OPDISC	45.54	8.05	0.00	1.04	45.54	8.37	0.021	0.127
S_POLDISC	58.22	6.43	0.00	0.93	58.22	5.97	0.014	0.061
S_POLPART	54.14	4.70	0.17	0.90	54.95	4.25	0.026	0.115
S_RELINF	51.03	4.59	0.00	0.96	51.03	4.39	0.026	0.173
S_SCACT	46.83	5.99	-0.18	1.11	45.78	6.65	0.022	0.113
S_STUTREL	43.28	5.33	0.00	0.96	43.28	5.11	0.022	0.113
E_EUIDENT	41.02	4.77	0.00	1.06	41.02	5.03	0.035	0.177
E_FREEMOVE	37.19	4.61	0.00	1.05	37.19	4.86	0.008	0.040
E_INDFUT	37.04	5.10	0.00	1.10	37.04	5.60	0.057	0.319

Describing questionnaire scale indices

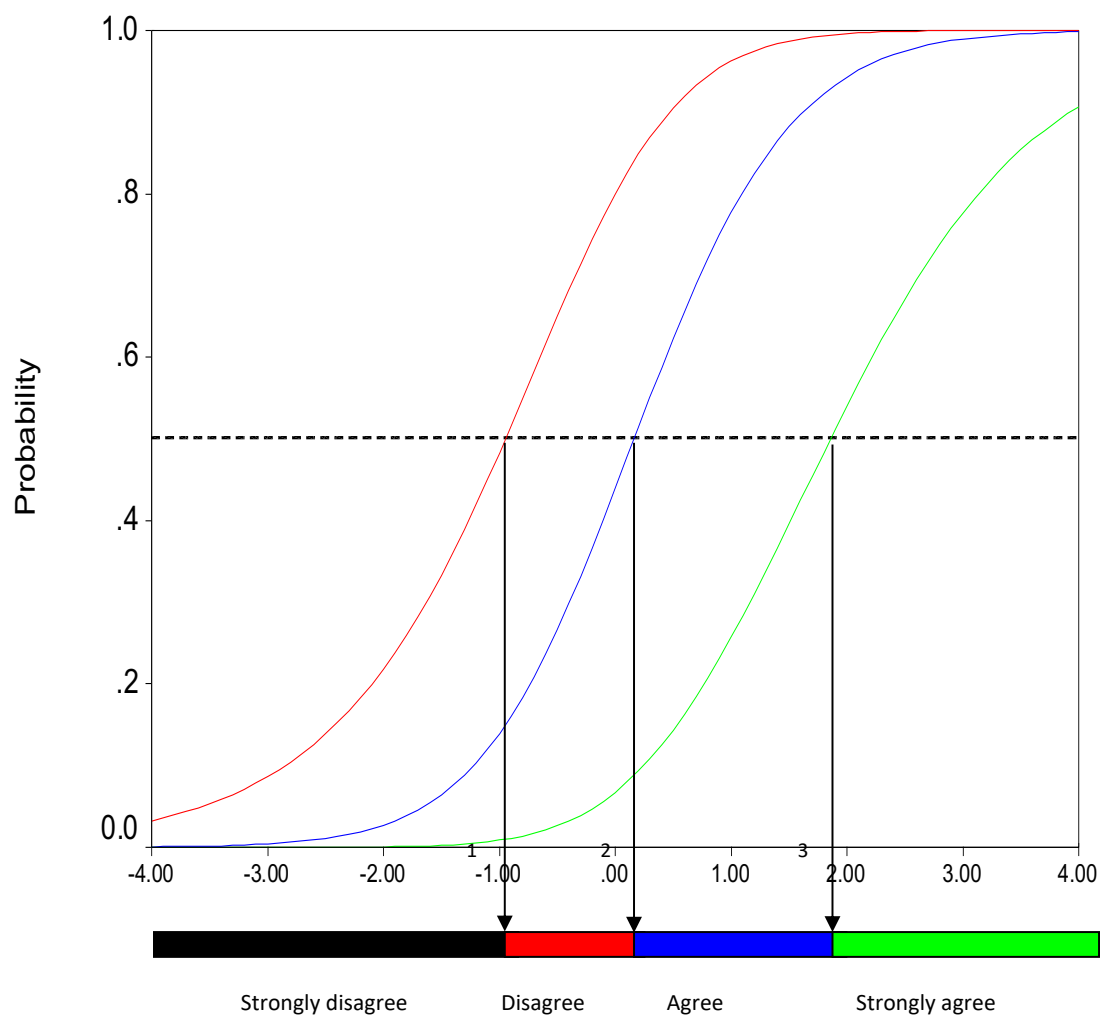
Questionnaire scales derived from weighted likelihood estimates (logits) present values on a continuum with an ICCS 2022 (or, where equated, 2016 or 2009) average of 50 and a standard deviation of 10 (for equally weighted national samples). This allows an interpretation of these scores by comparing individual scores or group average scores with the ICCS 2022 average, or the respective averages for ICCS 2016 and 2009 in the case of equated student questionnaire scales. 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.⁶

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 higher 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 12.1). The three vertical lines denote those points on the latent continuum where it becomes more likely to score > 0 , > 1 , or > 2 . These locations τ_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.

⁶ This approach was also used in the IEA ICCS 2009 and 2016 surveys (see Schulz & Friedman, 2011, 2018).

Figure 12.1: Summed category probabilities for fictitious item

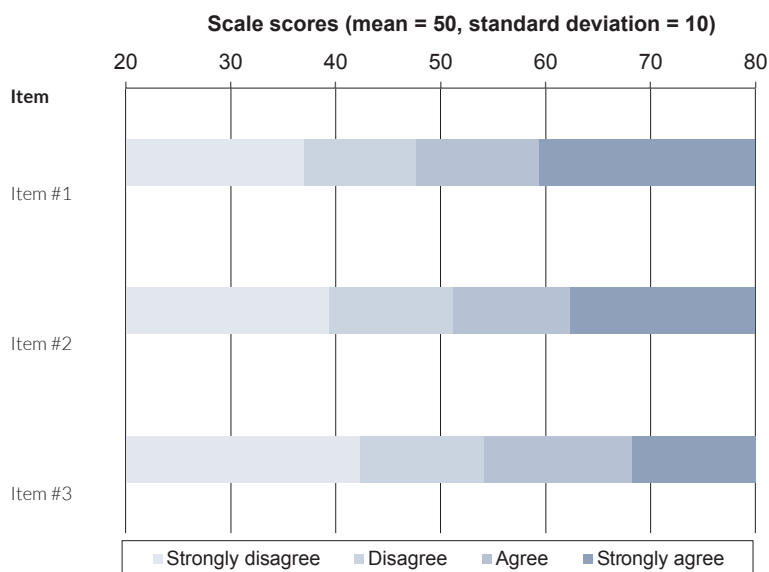


Summed probabilities are not identical to expected item scores and have to be understood in terms of the probability of scoring at least a particular category.⁷ Thurstonian thresholds can be used to indicate for each item category those points on a scale at which respondents have a 0.5 probability of 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 2022 reports predict the minimum coded score (e.g., 0 = “strongly disagree,” 1 = “disagree,” 2 = “agree,” and 3 = “strongly agree”) a respondent would obtain on a Likert-type item. For example, we could predict that students with a certain scale score would have a 50 percent probability of agreeing (or strongly agreeing) with a particular item (see the example item-by-score map in Figure 12.2). For each item, it is thus possible to determine Thurstonian thresholds, 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.

⁷ 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).

Figure 12.2: Example of questionnaire item-by-score map

**Example of how to interpret the item-by-score map**

- #1. A respondent with score 30 has more than 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 items 1 and to disagree with items 2 and 3
- #4. A respondent with score 60 has more than 50% probability to strongly agree with items 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

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 2022 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”).

12.4 Scaled indices

Student questionnaire

National index of students’ socioeconomic background

The bivariate and multivariate analyses presented in the international report (Schulz et al., 2024) include data from a composite index reflecting students’ socioeconomic background, which was derived in the same way in the previous ICCS cycles in 2016 and 2009.⁸ The *national index of students’ socioeconomic background*

⁸ Similar approaches to deriving socio-economic background indices have been applied in the IEA International Computer and Information Literacy Study (ICILS) and the OECD Programme for International Student Assessment PISA (see, for example, Schulz & Friedman, 2020; OECD, 2023).

(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). To avoid small numbers of cases in these categories, for the S_HISCED index, we collapsed the lowest two categories resulting in an indicator variable with four categories: lower-secondary or below, upper-secondary, tertiary non-university, and university education. To achieve a more equal distribution of proportions, the S_HOMLIT index was also 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. Previous analyses had shown approximately linear associations across these categories with civic knowledge test scores and the other indicators of socioeconomic background.

We imputed values for students who had missing data for only one of the three indicators based on a regression on the other two variables that had been estimated for students with values on all three variables using predicted values plus a random component. This simple 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. Table 12.3 shows the factor loadings and reliabilities for each national sample.

Table 12.3: Factor loadings and reliabilities for the national index of students' socioeconomic background in ICCS 2022

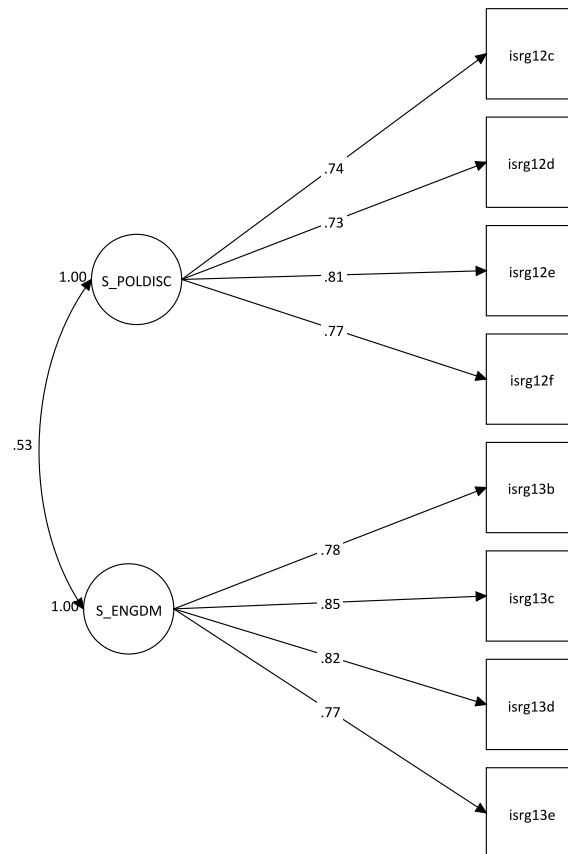
Country	Factor loadings for:			Cronbach's alpha
	Highest parental occupation	Highest parental education	Books at home	
Brazil	0.70	0.78	0.70	0.56
Bulgaria	0.81	0.83	0.76	0.72
Chinese Taipei	0.78	0.81	0.73	0.66
Colombia	0.82	0.79	0.57	0.56
Croatia	0.84	0.86	0.67	0.70
Cyprus	0.80	0.78	0.68	0.62
Denmark	0.83	0.82	0.68	0.67
Estonia	0.79	0.79	0.68	0.62
France	0.78	0.77	0.70	0.61
Italy	0.81	0.83	0.72	0.69
Latvia	0.80	0.79	0.67	0.62
Lithuania	0.81	0.79	0.71	0.66
Malta	0.81	0.76	0.68	0.61
Netherlands	0.78	0.81	0.67	0.62
North Rhine-Westphalia*	0.78	0.69	0.75	0.59
Norway	0.78	0.78	0.71	0.62
Poland	0.82	0.84	0.70	0.70
Romania	0.87	0.87	0.76	0.78
Schleswig-Holstein*	0.76	0.65	0.72	0.51
Serbia	0.82	0.84	0.68	0.68
Slovak Republic	0.81	0.82	0.75	0.70
Slovenia	0.78	0.81	0.69	0.64
Spain	0.82	0.82	0.70	0.68
Sweden	0.77	0.77	0.70	0.59
ICCS 2022 average	0.80	0.81	0.70	0.65

Notes: * German benchmarking participant.

Students' participation in out-of-school activities

Question 12 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 including media use and participation in discussions of political or social issues 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." Using the same response options, Question 13 was designed to capture students' engagement with digital media.

Figure 12.3: Confirmatory factor analysis of items measuring students' participation in out-of-school activities



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.092	0.114	0.107	0.092
CFI	0.95	0.947	0.940	0.93
TLI	0.92	0.92	0.932	0.95

Two scales were derived from these two item sets. Four of the items in Question 12 were used to derive a scale reflecting *students' discussion of political and social issues outside of school* (S_POLDISC), while four of the items included in Question 13 were used to derive a scale of *students' engagement with digital media* (S_ENGDM). Higher scale scores on both scales reflect more frequent engagement in each type of these activities.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' participation in out-of-school activities (Figure 12.3). The model fit was marginally satisfactory and the two latent factors (S_POLDISC and S_ENGDM) were highly correlated ($r = 0.53$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was poor to marginally satisfactory across multiple-group models.

Table 12.4: Reliabilities for scales measuring students' participation in out-of-school activities

Country	S_POLDISC	S_ENGDM
Brazil	0.73	0.73
Bulgaria	0.79	0.80
Chinese Taipei	0.82	0.84
Colombia	0.72	0.77
Croatia	0.77	0.71
Cyprus	0.77	0.77
Denmark	0.80	0.71
Estonia	0.80	0.70
France	0.74	0.73
Italy	0.72	0.67
Latvia	0.82	0.74
Lithuania	0.81	0.78
Malta	0.77	0.75
Netherlands	0.78	0.72
North Rhine-Westphalia*	0.78	0.72
Norway	0.83	0.77
Poland	0.77	0.67
Romania	0.72	0.73
Schleswig-Holstein*	0.79	0.67
Serbia	0.76	0.78
Slovak Republic	0.74	0.74
Slovenia	0.76	0.83
Spain	0.70	0.73
Sweden	0.82	-
Average	0.77	0.74
Minimum value	0.70	0.67
Maximum value	0.83	0.84

The average reliabilities (Cronbach's alpha) across national samples⁹ were 0.77 for S_POLDISC (ranging from 0.70 to 0.83) and 0.74 for S_ENGDM (ranging from 0.67 to 0.84) (see Table 12.4). These scales were based on the item parameters shown in Table 12.5. The scale scores for S_POLDISC were linked to the scale established in the ICCS 2009 so that they were comparable across cycles. The item parameters used for scaling are displayed in Table 12.5.

⁹ In line with the reporting of study data, the ICCS 2022 average was calculated only based on data from countries that had met IEA sample response guidelines. Minimum and maximum values were computed based on all national samples.

Table 12.5: Partial credit model item parameters for scales measuring students' activities outside school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
<i>S_POLDISC</i>	<i>How often are you involved in each of the following activities?</i>				
IS4G12C	Talking with your parent(s) about political or social issues	0.10	-1.09	-0.24	1.33
IS4G12D	Talking with your parent(s) about what is happening in other countries	-0.67	-1.47	-0.22	1.69
IS4G12E	Talking with friends about political or social issues	0.61	-0.95	-0.27	1.22
IS4G12F	Talking with friends about what is happening in other countries	-0.05	-1.49	-0.15	1.64
<i>S_ENGDM</i>	<i>How often are you involved in each of the following activities related to the internet and social media?</i>				
IS4G13B	Posting your own content about a political or social issue on the internet or social media	0.53	-0.11	-0.40	0.51
IS4G13C	Sharing content about a political or social issue by someone else	0.34	-0.64	-0.20	0.84
IS4G13D	Commenting on an online post about a political or social issue	0.20	-0.34	-0.32	0.66
IS4G13E	[Liking] an online post about a political or social issue	-1.07	-0.77	-0.16	0.92

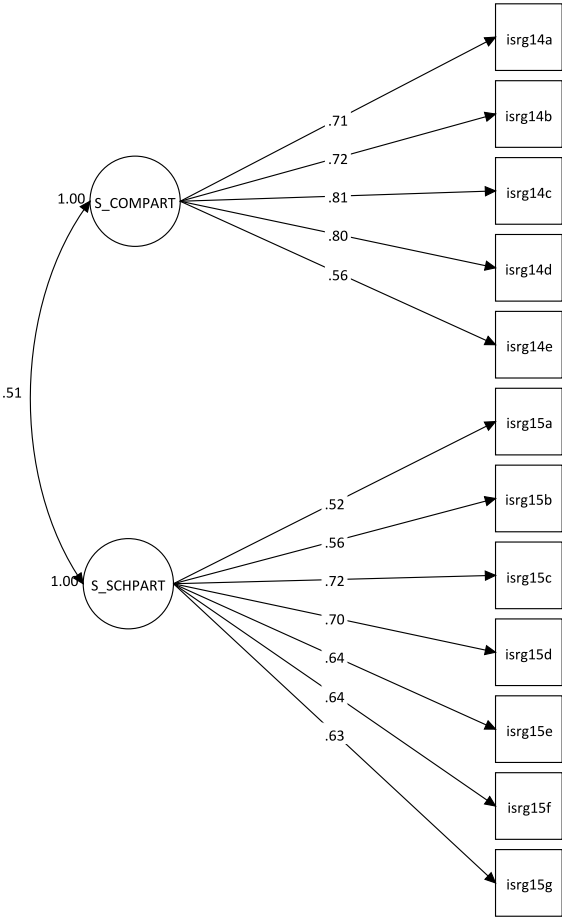
Students' civic engagement in the community and at school

Question 14 asked students to indicate their participation ("yes, I have done this within the last twelve months," "yes, I have done this but more than a year ago," or "no, I have never done this") in different types of groups or organizations in the community. Using the same response categories, Question 15 measured student reported participation in the six different types of civic-related school activities.

Five of the items in Question 14 were used to derive a scale reflecting *students' participation in wider community groups or organizations* (*S_COMPART*), while seven of the items included in Question 15 were used to derive a scale of *students' participation in civic activities at school* (*S_SCHPART*). Higher scale scores on both scales reflect more frequent engagement in each type of these activities. Both scales were not equated to the previous cycles and treated as new scales for ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' participation in the community and at school (Figure 12.4). The model fit was satisfactory for the pooled item set and the two latent factors (*S_COMPART* and *S_SCHPART*) were highly correlated ($r = 0.51$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was less satisfactory for more constrained models, suggesting a lack of measurement invariance across countries.

Figure 12.4: Confirmatory factor analysis of items measuring students' participation in civic activities in the community and at school



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.054	0.051	0.077	0.085
CFI	0.94	0.95	0.87	0.81
TLI	0.92	0.94	0.86	0.83

Table 12.6: Reliabilities for scales measuring students' participation in civic activities in the community and at school

Country	S_COMPART	S_SCHPART
Brazil	0.71	0.71
Bulgaria	0.71	0.79
Chinese Taipei	0.70	0.75
Colombia	0.75	0.73
Croatia	0.64	0.59
Cyprus	0.75	0.79
Denmark	0.65	0.69
Estonia	0.71	0.77
France	-	0.66
Italy	0.59	0.66
Latvia	0.73	0.77
Lithuania	0.72	0.71
Malta	0.68	0.75
Netherlands	0.70	0.70
North Rhine-Westphalia*	0.66	0.65
Norway	0.77	0.75
Poland	0.56	0.68
Romania	0.67	0.70
Schleswig-Holstein*	0.65	0.66
Serbia	0.78	0.70
Slovak Republic	0.69	0.73
Slovenia	0.72	0.76
Spain	0.72	0.64
Sweden	0.70	0.73
Average	0.69	0.71
Minimum value	0.56	0.59
Maximum value	0.78	0.79

The average reliabilities (Cronbach's alpha) across national samples were 0.69 for S_COMPART (ranging from 0.56 to 0.78) and 0.71 for S_SCHPART (ranging from 0.59 to 0.79) (see Table 12.6). These scales were based on the item parameters shown in Table 12.7.

Table 12.7: Partial credit model item parameters for scales measuring students' participation in civic activities in the community and at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)
S_COMPART	Have you ever been involved in activities of any of the following organisations, clubs or groups?			
IS4G14A	A youth organisation affiliated with a political party or union	0.66	0.76	-0.76
IS4G14B	A voluntary group doing something to help the [local community]	-0.57	-0.31	0.31
IS4G14C	A group or organization campaigning for a particular cause (e.g. environmental protection, human rights, or animal rights)	-0.28	-0.24	0.24
IS4G14D	A global campaign for a particular issue (e.g. action on climate change)	0.44	0.11	-0.11
IS4G14E	A community youth group (e.g.[scouts, YMCA])	-0.26	0.02	-0.02
S_SCHPART	At school, have you ever done any of the following activities?			
IS4G15A	Active participation in an organised debate	-0.29	0.19	-0.19
IS4G15B	Voting for [class representative] or [school parliament/council]	-1.26	0.01	-0.01
IS4G15C	Taking part in decision-making about how the school is run	0.21	0.15	-0.15
IS4G15D	Taking part in discussions at a [student assembly/gathering]	0.17	0.16	-0.16
IS4G15E	Becoming a candidate for [class representative] or [school parliament/council]	-0.01	0.05	-0.05

IS4G15F	Participating in an activity to make the school more [environmentally friendly]	0.05	-0.23	0.23
IS4G15G	Participating in writing articles on political and social issues for a [school publication or website]	1.12	0.36	-0.36

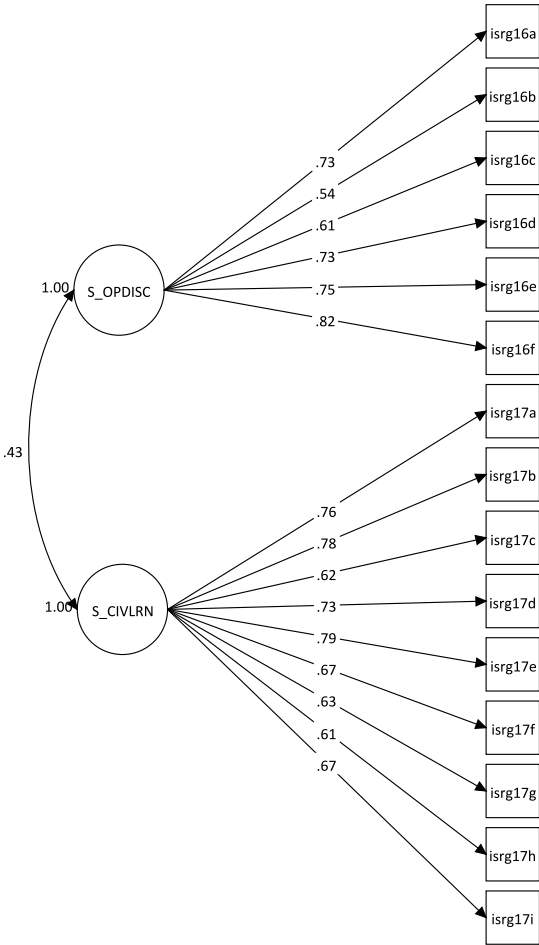
Students' perceptions of classroom climate and civic learning

Question 16 of the student questionnaire required students to report the frequency (“never,” “rarely,” “sometimes,” or “often”) of things happening in their classrooms that indicated openness to the discussion of political and social issues. Question 17 was designed to capture students' reports on their learning (“to a large extent,” “to a moderate extent,” “to a small extent,” or “not at all”) about different civic issues at school.

Two scales were derived from these two item sets. Six of the items in Question 12 were used to derive a scale reflecting *students' perceptions of openness in classroom discussions* (S_OPDISC), while nine of the items included in Question 17 were used to derive a scale of *students' civic learning* (S_CIVLRN). Higher scale scores on both scales reflect more openness in classroom climate or a larger extent of reported civic learning, and both scales were equated to ICCS 2009 and ICCS 2016, respectively.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' perceptions of classroom climate and civic learning (Figure 12.5). The model fit was satisfactory and the two latent factors (S_OPDISC and S_CIVLRN) were moderately correlated ($r = 0.43$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was less satisfactory for more constrained models, suggesting a lack of measurement invariance across countries.

Figure 12.5: Confirmatory factor analysis of items measuring students' perceptions of open classroom climates and civic learning



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.060	0.065	0.070	0.089
CFI	0.96	0.96	0.94	0.88
TLI	0.95	0.95	0.94	0.90

Table 12.8: Reliabilities for scales measuring students' perceptions of open classroom climates and civic learning

Country	S_OPDISC	S_CIVLRN
Brazil	0.78	0.83
Bulgaria	0.84	0.87
Chinese Taipei	0.88	0.92
Colombia	0.80	0.84
Croatia	0.80	0.85
Cyprus	0.82	0.88
Denmark	0.81	0.79
Estonia	0.81	0.86
France	0.82	0.82
Italy	0.73	0.79
Latvia	0.86	0.88
Lithuania	0.82	0.87
Malta	0.78	0.84
Netherlands	0.81	0.84
North Rhine-Westphalia*	0.77	0.80
Norway	0.85	0.87
Poland	0.78	0.83
Romania	0.70	0.80
Schleswig-Holstein*	0.77	0.81
Serbia	0.75	0.83
Slovak Republic	0.75	0.82
Slovenia	0.83	0.88
Spain	0.79	0.84
Sweden	0.83	0.85
Average	0.80	0.84
Minimum value	0.70	0.79
Maximum value	0.88	0.92

The average reliabilities (Cronbach's alpha) across national samples were 0.80 for S_OPDISC (ranging from 0.70 to 0.88) and 0.84 for S_CIVLRN (ranging from 0.79 to 0.92) (see Table 12.8). These scales were derived based on the item parameters shown in Table 12.9.

Table 12.9: Partial credit model item parameters for scales measuring students' perceptions of open classroom climates and civic learning

Scale or item	Question/item wording	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?				
IS4G16A	Teachers encourage students to make up their own minds.	-0.29	-0.95	-0.33	1.27
IS4G16B	Students [bring up] current political events for discussion in class.	0.82	-1.46	-0.09	1.54
IS4G16C	Students express opinions in class even when their opinions are different from most of the other students.	-0.34	-1.17	-0.14	1.32
IS4G16D	Teachers encourage students to discuss the issues with people having different opinions.	0.41	-1.00	-0.21	1.21
IS4G16E	Teachers present several sides of the issues when explaining them in class.	-0.05	-1.03	-0.24	1.27
IS4G16F	Teachers encourage students to express their opinions.	-0.56	-0.63	-0.15	0.78
S_CIVLRN	At school, to what extent have you learned about the following topics?				
IS4G17A	How citizens can vote in local or national elections	0.11	-1.05	-0.29	1.35
IS4G17B	How laws are introduced and changed in [country of test]	0.15	-1.33	-0.20	1.53

IS4G17C	How to protect the environment (e.g. through energy-saving or recycling)	-1.22	-1.13	-0.17	1.30
IS4G17D	How to contribute to solving problems in the [local community]	0.15	-1.43	-0.15	1.59
IS4G17E	How citizen rights are protected in [country of test]	-0.10	-1.20	-0.16	1.36
IS4G17F	Political issues and events in other countries	0.07	-1.51	-0.07	1.58
IS4G17G	How the economy works	0.07	-1.28	-0.13	1.42
IS4G17H	How to check whether online information can be trusted	-0.32	-1.08	-0.09	1.17
IS4G17I	How to become a candidate in a [local election]	1.08	-0.78	-0.19	0.96

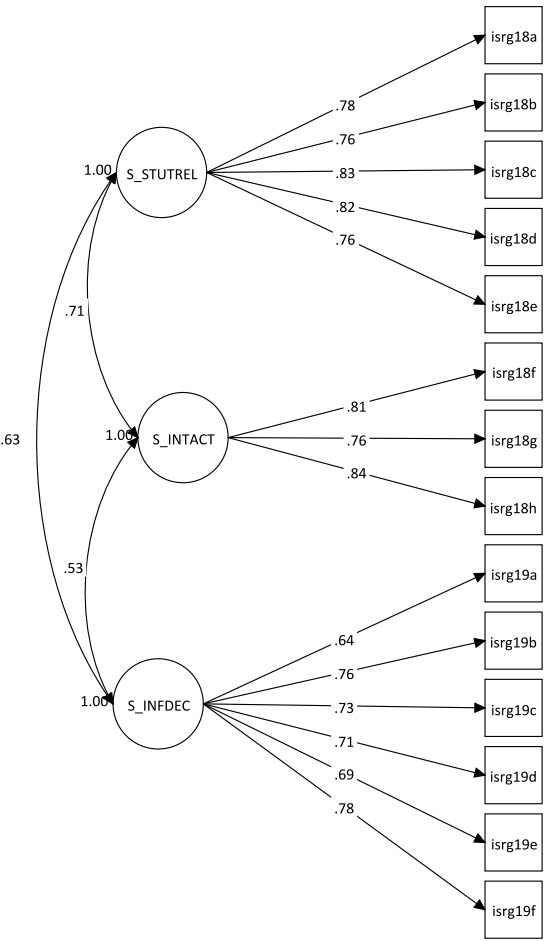
Students' perceptions of school interactions

Question 18 of the student questionnaire required students to indicate their level of agreement (“strongly agree,” “agree,” “disagree,” or “strongly disagree”) with statements about social interactions at their school. Using the same response categories, Question 19 asked students about their agreement with statements about student participation at their school.

Three scales were derived from these two item sets. Five of the items were used to derive a scale reflecting *students' perceptions of student-teacher relations at school* (S_STUTREL), while three of the items were used to derive a scale of *students' perceptions of student interaction at school* (S_INTACT). The six items included in Question 19 were used to derive a scale reflecting *students' beliefs about their influence on decision-making at school* (S_INFDEC). Higher scale scores on both scales reflect higher levels of agreement. While S_STUTREL and S_INTACT were equated to ICCS 2009 and ICCS 2016, respectively, S_INFDEC was a new scale for ICCS 2022.

We used confirmatory factor analyses assuming a three-dimensional model for these items to assess students' perceptions of school interactions (Figure 12.6). The model fit was satisfactory, and all three latent factors were highly correlated with each other. When reviewing measurement invariance across multiple-group models with different constraints, the model fit was still satisfactory for more constrained models, suggesting a relatively high level of measurement invariance across countries.

Figure 12.6: Confirmatory factor analysis of items measuring students' perceptions of school interactions



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.069	0.070	0.067	0.073
CFI	0.96	0.97	0.97	0.95
TLI	0.95	0.96	0.96	0.96

Table 12.10: Reliabilities for scales measuring students' perceptions of school interactions

Country	S_STUTREL	S_INTACT	S_INFDEC
Brazil	0.79	0.75	0.77
Bulgaria	0.83	0.78	0.82
Chinese Taipei	0.91	0.85	0.90
Colombia	0.82	0.75	0.83
Croatia	0.83	0.75	0.79
Cyprus	0.86	0.72	0.84
Denmark	0.85	0.82	0.79
Estonia	0.83	0.79	0.82
France	0.82	0.71	0.78
Italy	0.82	0.74	0.78
Latvia	0.81	0.79	0.83
Lithuania	0.83	0.76	0.83
Malta	0.82	0.77	0.82
Netherlands	0.82	0.79	0.82
North Rhine-Westphalia*	0.80	0.71	0.75
Norway	0.90	0.84	0.87
Poland	0.84	0.72	0.78
Romania	0.81	0.74	0.72
Schleswig-Holstein*	0.78	0.72	0.73
Serbia	0.86	0.76	0.84
Slovak Republic	0.83	0.71	0.76
Slovenia	0.84	0.74	0.86
Spain	0.82	0.75	0.81
Sweden	0.87	0.83	0.84
Average	0.83	0.76	0.81
Minimum value	0.78	0.71	0.72
Maximum value	0.91	0.85	0.90

The average reliabilities (Cronbach's alpha) across national samples were 0.83 for S_STUTREL (ranging from 0.78 to 0.91), 0.76 for S_INTACT (ranging from 0.71 to 0.85), and 0.81 for S_INFDEC (ranging from 0.72 to 0.90) (see Table 12.10). These scales were based on the item parameters shown in Table 12.11.

Table 12.11: Partial credit model item parameters for scales measuring students' perceptions of school interactions

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_STUTREL	<i>How much do you agree or disagree with the following statements about teachers and students at your school?</i>				
IS4G18A	Most of my teachers treat me fairly.	-0.26	-1.99	-0.76	2.75
IS4G18B	Students get along well with most teachers.	0.48	-2.86	-0.43	3.29
IS4G18C	Most of my teachers are interested in students' well-being.	0.11	-2.28	-0.63	2.90
IS4G18D	Most of my teachers listen to what I have to say.	-0.02	-2.31	-0.64	2.95
IS4G18E	If I need extra help, I receive it from my teachers.	-0.30	-2.09	-0.84	2.93
S_INTACT	<i>How much do you agree or disagree with the following statements about teachers and students at your school?</i>				
IS4G18F	Most students at my school treat each other with respect.	0.31	-2.38	-0.48	2.86
IS4G18G	Most students at my school get along well with each other.	-0.21	-2.48	-0.64	3.12
IS4G18H	My school is a place where students feel safe.	-0.10	-2.16	-0.80	2.95
S_INFDEC	<i>How much do you agree or disagree with the following statements about student participation at school?</i>				
IS4G19A	Students' participation in decision-making contributes to make my school better.	-0.66	-1.83	-0.71	2.54
IS4G19B	There are clear rules about how students can be involved in decision-making at my school.	-0.14	-2.31	-0.48	2.79
IS4G19C	My school encourages students to organize in groups to express their opinions.	0.14	-2.24	-0.17	2.41
IS4G19D	Students can influence decisions that affect our whole school.	0.35	-2.13	-0.31	2.44
IS4G19E	Voting in student elections makes a difference to what happens at my school.	0.19	-1.98	-0.40	2.38
IS4G19F	Students' interests are usually considered when decisions are made at my school	0.13	-2.01	-0.58	2.59

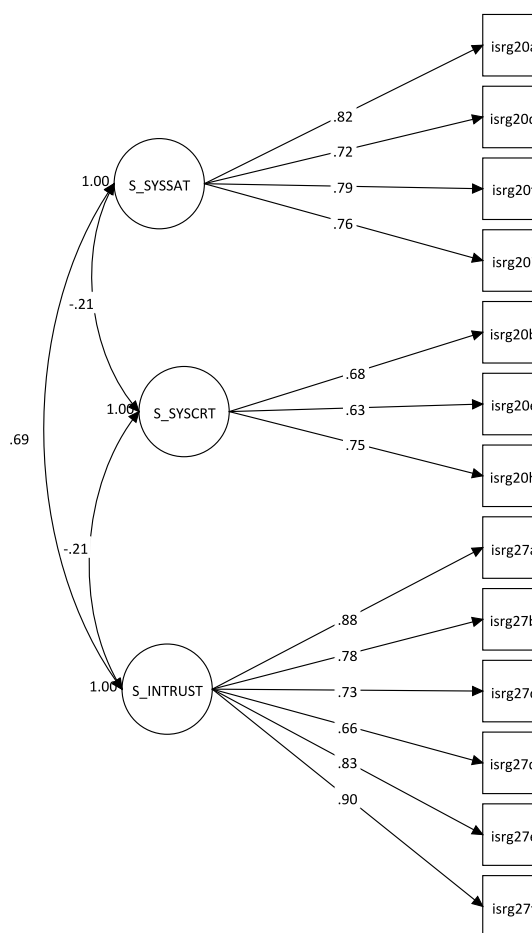
Students' perceptions of political systems and institutions

Question 20 of the student questionnaire required students to indicate their level of agreement (“strongly agree,” “agree,” “disagree,” or “strongly disagree”) with positive and negative statements about the political system in their countries. Question 27 was designed to measure how much trust (“completely,” “quite a lot,” “a little,” or “not at all”) students expressed in different institutions or groups in their country.

Three scales were derived from these two item sets. Four of the items in Question 20 were used to derive a scale reflecting *students' satisfaction with the political system* (S_SYSSAT), while three of the items included in this question were used to derive a scale of *students' critical views of the political system* (S_SYSCRT). Six items included in Question 27 provided a basis for deriving a *scale of students' trust in civic institutions* (S_INTRUST). Higher scale scores on these scales reflect more positive or negative views, or higher levels of trust. S_INTRUST was equated to ICCS 2009.

We used confirmatory factor analyses assuming a three-dimensional model for these items to assess students' perceptions of political systems and institutions (Figure 12.7). The model fit was satisfactory for the pooled data set, and while S_SYSSAT and S_INTRUST were positively correlated ($r = 0.69$), S_SYSCRT had negative correlations of -0.21 with both other latent variables. When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained satisfactory across models with slightly poorer fit for the most constrained model.

Figure 12.7: Confirmatory factor analysis of items measuring students' perceptions of political systems and institutions



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.068	0.074	0.071	0.079
CFI	0.98	0.97	0.97	0.95
TLI	0.97	0.97	0.97	0.96

Table 12.12: Reliabilities for scales measuring students' perceptions of political systems and institutions

Country	S_SYSSAT	S_SYSCRT	S_INTRUST
Brazil	0.77	0.58	0.88
Bulgaria	0.83	0.67	0.89
Chinese Taipei	0.79	0.71	0.90
Colombia	0.82	0.62	0.92
Croatia	0.80	0.68	0.85
Cyprus	0.81	0.69	0.90
Denmark	0.67	0.56	0.85
Estonia	0.78	0.66	0.87
France	0.75	0.63	0.86
Italy	0.78	0.59	0.84
Latvia	0.83	0.67	0.88
Lithuania	0.81	0.67	0.87
Malta	0.80	0.66	0.86
Netherlands	0.74	0.58	0.86
North Rhine-Westphalia*	0.68	0.48	0.85
Norway	0.77	0.69	0.90
Poland	0.80	0.62	0.83
Romania	0.78	0.57	0.84
Schleswig-Holstein*	0.70	0.50	0.83
Serbia	0.82	0.62	0.88
Slovak Republic	0.81	0.57	0.85
Slovenia	0.81	0.73	0.88
Spain	0.76	0.61	0.83
Sweden	0.73	0.63	0.89
Average	0.78	0.63	0.87
Minimum value	0.67	0.48	0.83
Maximum value	0.83	0.73	0.92

The average reliabilities (Cronbach's alpha) across national samples were 0.78 for S_SYSSAT (ranging from 0.67 to 0.83), 0.63 for S_SYSCRT (ranging from 0.48 to 0.73), and 0.87 for S_INTRUST (ranging from 0.83 to 0.92) (see Table 12.12). These scales were based on the item parameters shown in Table 12.13.

Table 12.13: Partial credit model item parameters for scales measuring students' perceptions of political systems and institutions

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_SYSSAT	How much do you agree or disagree with the following statements about political leaders, members of parliament/congress and political decision-making in this country?				
IS4G20A	The political system of [country of test] works well.	-0.20	-2.60	-0.42	3.01
IS4G20D	Members of [parliament/congress] are good at representing the interests of young people.	0.21	-3.12	-0.11	3.23
IS4G20F	Members of [parliament/congress] generally represent the interests of people in their country well.	-0.18	-3.17	-0.44	3.60
IS4G20I	Members of [parliament/congress] treat all people in society fairly.	0.17	-2.77	-0.13	2.91
S_SYSCRT	How much do you agree or disagree with the following statements about political leaders, members of parliament/congress and political decision-making in this country?				
IS4G20B	Members of [parliament/congress] do not care enough about the wishes of the people.	0.27	-2.48	-0.15	2.63
IS4G20E	Political leaders have too much power compared to other people.	-0.15	-2.15	-0.16	2.31
IS4G20H	Members of [parliament/congress] usually forget the needs of the people who voted for them.	-0.13	-2.19	-0.22	2.42
S_INTRUST	How much do you trust each of the following groups, institutions or sources of information?				
IS4G27A	The [national government] of [country of test]	0.11	-2.44	-0.18	2.62
IS4G27B	The [local government] of your town or city	-0.28	-2.87	-0.36	3.23
IS4G27C	Courts of justice	-0.51	-2.62	-0.36	2.98
IS4G27D	The police	-0.66	-2.06	-0.43	2.49
IS4G27E	Political parties	0.93	-2.89	0.03	2.86
IS4G27F	[Parliament/congress]	0.40	-2.61	-0.20	2.80

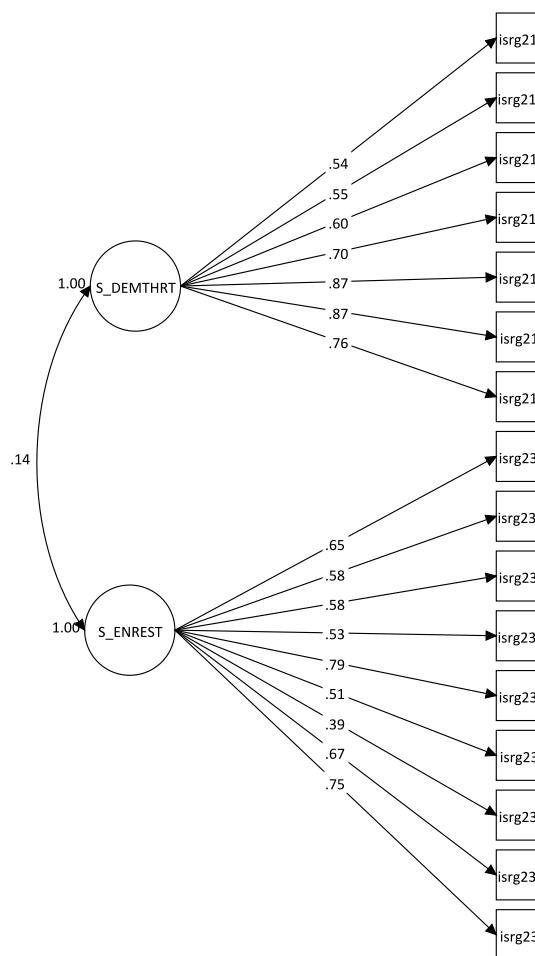
Students' beliefs about threats to democracy and restrictions in national emergencies

Question 21 of the student questionnaire required students to indicate to which extent they considered certain situations as bad for democracy (“very bad,” “quite bad,” “somewhat bad,” or “not bad at all”). Question 23 of the student questionnaire required students to indicate their level of agreement (“strongly agree,” “agree,” “disagree,” or “strongly disagree”) with actions governments may take in time of a national emergency (such as the COVID-19 pandemic) in their countries.

Two scales were derived from each of these two item sets. All seven items in Question 21 were used to derive a scale reflecting *students' beliefs about threats to democracy* (S_DEMTHRT), while the nine items included in Question 23 were used to derive a scale of *students' endorsement of restrictions in a national emergency* (S_ENREST). Higher scale scores reflect more concern about threats for democracy (S_DEMTHRT) or higher levels of endorsement of restrictions in a national emergency (S_ENREST). Both scales were new to ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' beliefs about threats to democracy and restrictions in national emergencies (Figure 12.8). The model fit was satisfactory for the pooled data set, and the two latent factors were weakly correlated ($r = 0.14$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was slightly less satisfactory for the most restricted (scalar) model.

Figure 12.8: Confirmatory factor analysis of items measuring students' beliefs about threats to democracy and restrictions in national emergencies



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.053	0.062	0.064	0.076
CFI	0.95	0.95	0.94	0.89
TLI	0.95	0.94	0.93	0.91

Table 12.14: Reliabilities for scales measuring students' beliefs about threats to democracy and restrictions in national emergencies

Country	S_DEMTHRT	S_ENREST
Brazil	0.79	0.83
Bulgaria	0.82	0.83
Chinese Taipei	0.90	0.81
Colombia	0.81	0.83
Croatia	0.74	0.75
Cyprus	0.83	0.83
Denmark	0.79	0.77
Estonia	0.83	0.79
France	0.76	0.81
Italy	0.72	0.78
Latvia	0.82	0.80
Lithuania	0.81	0.80
Malta	0.79	0.77
Netherlands	0.83	0.76
North Rhine-Westphalia*	0.75	0.77
Norway	0.88	0.83
Poland	0.74	0.76
Romania	0.71	0.78
Schleswig-Holstein*	0.78	0.78
Serbia	0.80	0.79
Slovak Republic	0.79	0.78
Slovenia	0.86	0.85
Spain	0.79	0.78
Sweden	0.88	0.79
Average	0.80	0.79
Minimum value	0.71	0.75
Maximum value	0.90	0.85

The average reliabilities (Cronbach's alpha) across national samples were 0.80 for S_DEMTHRT (ranging from 0.71 to 0.90) and 0.79 for S_ENREST (ranging from 0.75 to 0.85) (see Table 12.14). These scales were based on the item parameters shown in Table 12.15.

Table 12.15: Partial credit model item parameters for scales measuring students' beliefs about threats to democracy and restrictions in national emergencies

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_DEMTHRT	How bad would it be for democracy if the following situations were to happen?				
IS4G21A	Political leaders give government jobs to family members.	0.36	-1.38	0.12	1.26
IS4G21B	The government breaks a law to fulfil a promise they made before they were elected.	-0.03	-1.46	0.14	1.32
IS4G21C	Opposition leaders are arrested because they openly criticized a new law.	0.26	-1.39	0.15	1.24
IS4G21D	Only government supporters are appointed as judges.	0.12	-1.25	0.19	1.06
IS4G21E	The government closes newspapers, radio and television stations that have been critical of its policies.	-0.30	-0.94	0.34	0.60
IS4G21F	The government blocks social media to prevent users from criticizing its policies.	-0.41	-0.82	0.33	0.49
IS4G21G	The government controls all newspapers, radio and television stations in a country.	0.00	-0.75	0.27	0.48
S_ENREST	To what extent do you agree or disagree that the government should have the right to take the following actions during a national emergency?				
IS4G23A	Close shops and businesses	0.27	-1.43	0.15	1.28
IS4G23B	Fine people whose behavior might put others at risk	-0.72	-0.89	-0.62	1.51
IS4G23C	Oblige people to provide information about their movements	0.38	-1.55	0.16	1.39

IS4G23D	Postpone meetings of the [parliament/congress]	-0.03	-1.83	-0.08	1.91
IS4G23E	Impose travel restrictions	-0.11	-0.96	-0.48	1.44
IS4G23F	Make peaceful protests, marches or rallies illegal	0.25	-1.30	-0.07	1.37
IS4G23G	Close schools	-0.53	-0.77	0.01	0.76
IS4G23H	Make it illegal for people to leave their homes without sufficient cause	0.60	-1.13	-0.02	1.15
IS4G23I	Prohibit larger gatherings of people at sporting and entertainment events	-0.11	-0.72	-0.49	1.21

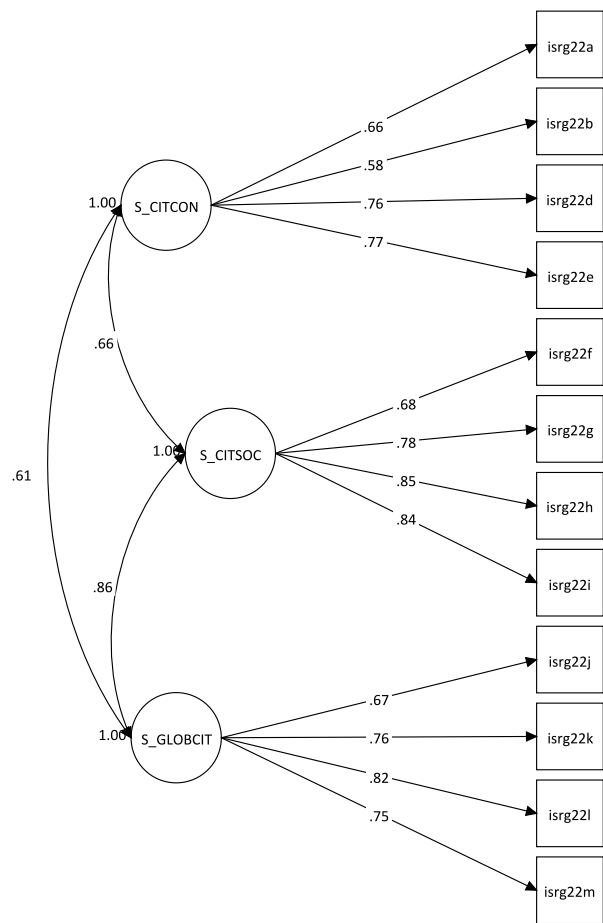
Students' perceptions of the importance of good citizenship behavior

Question 22 of the student questionnaire required students to indicate how important (“very important,” “quite important,” “not very important,” “not at all important”) they considered different behaviors for good citizenship.

Three scales were derived from this item set. Three sets of four items each were used to derive scales reflecting *students' beliefs about the importance of conventional citizenship activities* (S_CITCON), *students' beliefs about the importance of social movement related citizenship activities* (S_CITSOC), and *students' beliefs about the importance of globally oriented citizenship activities* (S_GLOBCIT). Higher scale scores on these scales reflect higher levels of importance assigned to these types of citizenship. S_CITCON and S_CITSOC were equated to ICCS 2009.

We used confirmatory factor analyses assuming a three-dimensional model for these items to assess students' perceptions of the importance of good citizenship behavior (Figure 12.9). The model fit was marginally satisfactory for the pooled data set, and the three latent factors were highly correlated with each other, with the highest correlation ($r = 0.86$) observed between S_CITSOC and S_GLOBCIT. When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained marginally satisfactory all across models with different levels of constraints.

Figure 12.9: Confirmatory factor analysis of items measuring students' perceptions of the importance of good citizenship behavior



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.080	0.089	0.081	0.086
CFI	0.95	0.95	0.96	0.93
TLI	0.94	0.94	0.95	0.94

Table 12.16: Reliabilities for scales measuring students' perceptions of the importance of good citizenship behavior

Country	S_CITCON	S_CITSOC	S_GLOBCIT
Brazil	0.70	0.79	0.75
Bulgaria	0.70	0.81	0.79
Chinese Taipei	0.78	0.88	0.86
Colombia	0.74	0.79	0.79
Croatia	0.71	0.76	0.76
Cyprus	0.78	0.86	0.82
Denmark	0.63	0.80	0.75
Estonia	0.68	0.80	0.81
France	0.70	0.76	0.75
Italy	0.67	0.78	0.69
Latvia	0.77	0.83	0.79
Lithuania	0.72	0.81	0.77
Malta	0.74	0.82	0.80
Netherlands	0.68	0.75	0.77
North Rhine-Westphalia*	0.67	0.79	0.68
Norway	0.75	0.83	0.82
Poland	0.67	0.77	0.70
Romania	0.64	0.73	0.72
Schleswig-Holstein*	0.67	0.76	0.72
Serbia	0.77	0.77	0.75
Slovak Republic	0.63	0.74	0.73
Slovenia	0.75	0.82	0.81
Spain	0.73	0.81	0.77
Sweden	0.72	0.85	0.80
Average	0.71	0.80	0.77
Minimum value	0.63	0.73	0.68
Maximum value	0.78	0.88	0.86

The average reliabilities (Cronbach's alpha) across national samples were 0.71 for S_CITCON (ranging from 0.63 to 0.78), 0.80 for S_CITSOC (ranging from 0.73 to 0.88) and 0.77 for S_GLOBCIT (ranging from 0.68 to 0.86) (see Table 12.16). These scales were based on the item parameters shown in Table 12.17.

Table 12.17: Partial credit model item parameters for scales measuring students' perceptions of the importance of good citizenship behavior

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_CITCON	How important are the following behaviours for being a good adult citizen?				
IS4G22A	Voting in every national election	-0.97	-1.63	-0.17	1.81
IS4G22B	Joining a political party	0.83	-2.31	0.41	1.90
IS4G22D	Following political issues in the newspaper, on the radio, on TV or on the Internet	-0.35	-1.98	-0.23	2.22
IS4G22E	Engaging in political discussions	0.49	-2.35	0.36	1.99
S_CITSOC	How important are the following behaviours for being a good adult citizen?				
IS4G22F	Participating in peaceful protests against laws believed to be unjust	0.92	-2.28	-0.10	2.38
IS4G22G	Participating in activities to benefit people in the [local community]	0.01	-2.31	-0.41	2.72
IS4G22H	Taking part in activities promoting human rights	-0.38	-2.03	-0.40	2.43
IS4G22I	Taking part in activities to protect the environment	-0.56	-1.91	-0.40	2.30
S_GLOBCIT	How important are the following behaviours for being a good adult citizen?				
IS4G22J	Showing interest in different cultures and languages	0.33	-2.22	-0.14	2.36
IS4G22K	Making changes to one's personal lifestyle in order to become more [environmentally friendly]	-0.02	-1.85	-0.40	2.25
IS4G22L	Supporting initiatives that promote equal opportunities for all people across the world	-0.12	-1.83	-0.40	2.23
IS4G22M	Helping people in less developed countries	-0.19	-1.59	-0.51	2.10

Students' support for equal rights for immigrants and gender equality

Question 24 of the student questionnaire required students to indicate their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with statements about rights and opportunities for immigrants. Using the same response categories, Question 25 asked students to rate their agreement with four positive and three negative statements about gender rights.

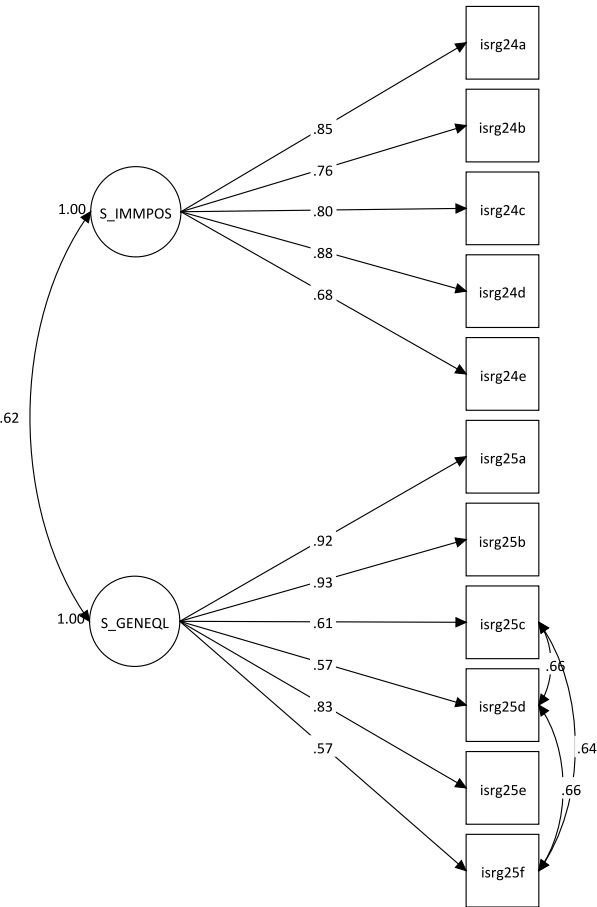
Two scales were derived from these two item sets. All five items in Question 24 were used to derive a scale reflecting *students' positive attitudes toward immigrants* (S_IMMPOS), while six of the items included in Question 25 were used to derive a scale of *students' endorsement of gender equality* (S_GENEQL). Higher scale scores on both scales reflect higher levels of endorsement of equal rights or opportunities for immigrants and gender equality. S_GENEQL was equated to ICCS 2009, however, due to changes in the item set and because the item set about immigrant rights and opportunities had only been included in the European questionnaire in ICCS 2016, S_IMMPOS was not equated to previous cycles.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' support for equal rights for immigrants and gender equality (Figure 12.10). Due to the mix of positive and negative items, correlated error terms were included to consider the measurement effects of this item set format. The model fit was very satisfactory for the pooled item set, and the two latent factors were highly correlated ($r = 0.62$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was satisfactory for both configural and scalar models.¹⁰

The average reliabilities (Cronbach's alpha) across national samples were 0.83 for S_IMMPOS (ranging from 0.77 to 0.93) and 0.81 for S_GENEQL (ranging from 0.73 to 0.87) (see Table 12.18). These scales were based on the item parameters shown in Table 12.19.

¹⁰ The scalar model did not converge for this item set and could not be reported.

Figure 12.10: Confirmatory factor analysis of items measuring students’ support for equal rights for immigrants and gender equality



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.048	0.060	*	0.055
CFI	0.99	0.99	*	0.98
TLI	0.99	0.98	*	0.99

* Model did not converge.

Table 12.18: Reliabilities for scales measuring students' support for equal rights for immigrants and gender equality

Country	S_IMMPOS	S_GENEQL
Brazil	0.81	0.76
Bulgaria	0.82	0.74
Chinese Taipei	0.93	0.78
Colombia	0.79	0.73
Croatia	0.81	0.83
Cyprus	0.84	0.80
Denmark	0.84	0.84
Estonia	0.85	0.85
France	0.82	0.82
Italy	0.80	0.81
Latvia	0.83	0.83
Lithuania	0.81	0.84
Malta	0.86	0.82
Netherlands	0.85	0.81
North Rhine-Westphalia*	0.81	0.82
Norway	0.88	0.83
Poland	0.78	0.87
Romania	0.77	0.79
Schleswig-Holstein*	0.83	0.81
Serbia	0.83	0.76
Slovak Republic	0.79	0.82
Slovenia	0.86	0.83
Spain	0.83	0.79
Sweden	0.87	0.80
Average	0.83	0.81
Minimum value	0.77	0.73
Maximum value	0.93	0.87

Table 12.19: Partial credit model item parameters for scales measuring students' activities outside school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_IMMPOS	How much do you agree or disagree with the following statements about immigrants?				
IS4G24A	[Immigrant] children should have the same opportunities for education that other children in the country have.	-0.78	-0.88	-0.93	1.81
IS4G24B	[Immigrants] who live in a country for several years should have the opportunity to vote in elections.	0.31	-2.06	-0.48	2.54
IS4G24C	[Immigrants] should have the opportunity to continue their own customs and lifestyle.	-0.06	-1.79	-0.71	2.50
IS4G24D	[Immigrants] should have the same rights that everyone else in the country has.	-0.26	-1.33	-0.65	1.99
IS4G24E	[Immigrants] bring many cultural, social and economic benefits to [country of test].	0.79	-2.22	-0.39	2.61
S_GENEQL	How much do you agree or disagree with the following statements?				
IS4G25A	Men and women should have equal opportunities to take part in government.	-0.63	-0.22	-0.91	1.13
IS4G25B	Men and women should have the same rights in every way.	-0.52	-0.77	-0.49	1.25
IS4G25C	Women should stay out of politics.	0.24	-0.62	-0.53	1.15
IS4G25D	When there are not many jobs available, men should have more right to a job than women.	0.50	-0.96	-0.29	1.24
IS4G25E	Men and women should get equal pay when they are doing the same jobs.	-0.34	-0.67	-0.49	1.17
IS4G25F	Men are better qualified to be political leaders than women.	0.74	-1.13	-0.22	1.35

Students' endorsement of equal rights for all ethnic groups in society

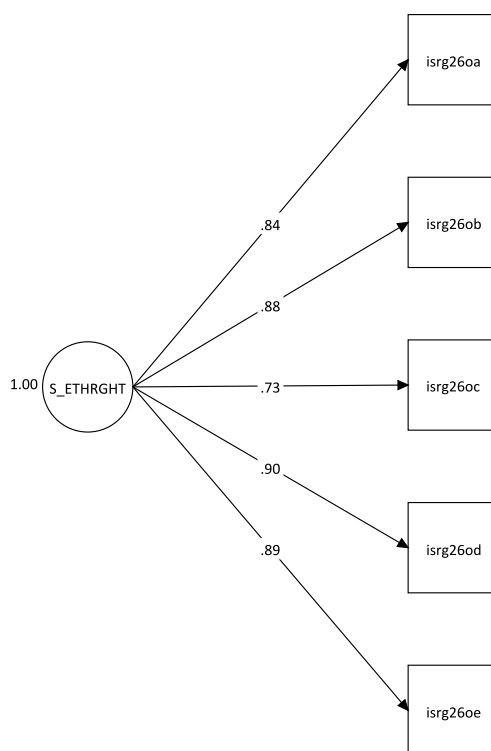
Question 26O of the student questionnaire was an international option requiring students to indicate

their level of agreement (“strongly agree,” “agree,” “disagree,” or “strongly disagree”) with statements about rights and opportunities for all ethnic groups in their country. These five items were used to derive a scale reflecting *students’ endorsement of equal rights for all ethnic groups in society* (S_ETHGRP). Higher scale scores on both scales reflect higher levels of endorsement and the scale was equated with ICCS 2009.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess students’ endorsement of equal rights for all ethnic groups in society (Figure 12.11). The model fit was satisfactory for the pooled sample. When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained satisfactory across models with different constraints, indicating a high degree of measurement invariance.

The average reliability (Cronbach’s alpha) across national samples for S_ETHRGT was 0.87 for (ranging from 0.82 to 0.93) (see Table 12.20). The scale was derived using item parameters shown in Table 12.21.

Figure 12.11: Confirmatory factor analysis of items measuring students’ endorsement of equal rights for all ethnic groups in society



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.067	0.075	0.050	0.052
CFI	1.00	1.00	1.00	1.00
TLI	0.99	0.99	1.00	1.00

Table 12.20: Reliabilities for scale measuring students' endorsement of equal rights for all ethnic groups in society

Country	S_ETHRGHT
Brazil	0.85
Bulgaria	0.86
Chinese Taipei	0.91
Colombia	0.84
Croatia	0.85
Cyprus	0.87
Denmark	0.87
Estonia	0.90
France	-
Italy	0.85
Latvia	0.87
Lithuania	0.89
Malta	0.90
Netherlands	0.87
North Rhine-Westphalia*	-
Norway	0.93
Poland	0.84
Romania	0.82
Schleswig-Holstein*	-
Serbia	-
Slovak Republic	0.85
Slovenia	0.89
Spain	0.86
Sweden	0.93
Average	0.87
Minimum value	0.82
Maximum value	0.93

Table 12.21: Item parameters for scale measuring students' endorsement of equal rights for all ethnic groups in society

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_ETHRGHT	How much do you agree or disagree with the following statements?				
IS4G26OA	All [ethnic groups] should have an equal chance to get good jobs in [country of test].	-0.36	-1.58	-1.31	2.89
IS4G26OB	Schools should teach students to respect [members of all ethnic groups].	-0.24	-1.97	-0.87	2.84
IS4G26OC	[Members of all ethnic groups] should be encouraged to run in elections for political office.	0.88	-3.03	-0.52	3.55
IS4G26OD	All [ethnic groups] should have an equal chance to get a good education in [country of test].	-0.13	-1.54	-1.13	2.67
IS4G26OE	[Members of all ethnic groups] should have the same rights and responsibilities.	-0.15	-1.64	-0.98	2.62

Students' support for environmental protection and their environmental concerns

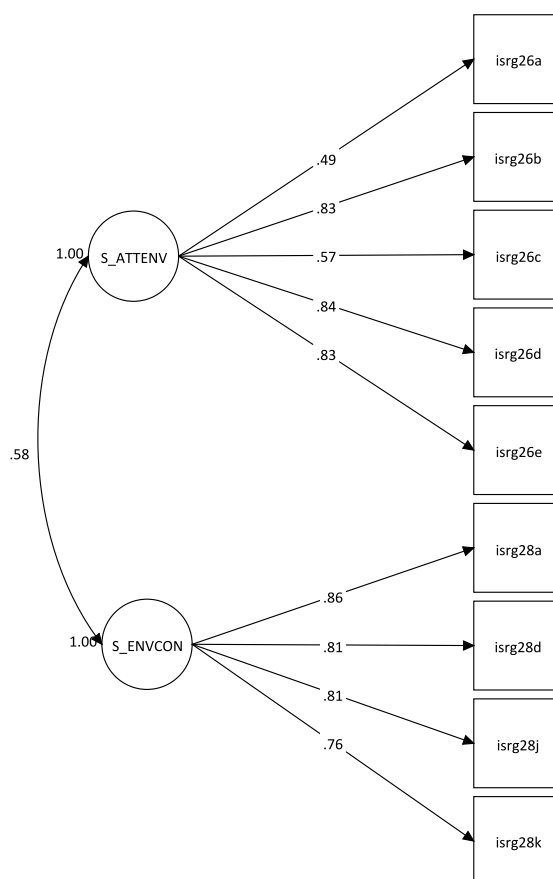
Question 26 of the student questionnaire asked students about their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with statements about environmental protection, while Question 28 required students to indicate to which extent they thought of global issues as threats to the world's future ("to a large extent," "to a moderate extent," "to a small extent," or "not at all").

Two scales were derived from these two item sets. All five items in Question 26 were used to derive a scale reflecting *students' positive attitudes toward environmental protection* (S_ATTENV), while four of the items included in Question 28 were used to derive a scale of *students' concern about threats to the global environment* (S_ENVCON). Higher scale scores reflect more positive attitudes for S_ATTENV, and higher levels of concern for S_ENVCON.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' support for environmental protection and their environmental concerns (Figure 12.12). The model fit was satisfactory for the pooled sample and the two latent factors were highly correlated ($r = 0.58$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was slightly poorer for the most constrained (scalar) model.¹¹

The average reliabilities (Cronbach's alpha) across national samples were 0.77 for S_ATTENV (ranging from 0.59 to 0.85) and 0.78 for S_ENVCON (ranging from 0.69 to 0.85) (see Table 12.22). These scales were both new for ICCS 2022 and based on the item parameters shown in Table 12.23.

Figure 12.12: Confirmatory factor analysis of items measuring students' support for environmental protection and their environmental concerns



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.062	0.070	*	0.074
CFI	0.98	0.98	*	0.95
TLI	0.97	0.97	*	0.96

* Model did not converge.

¹¹ The scalar model did not converge for this item set and could not be reported.

Table 12.22: Reliabilities for scales measuring students' support for environmental protection and their environmental concerns

Country	S_ATTENV	S_ENVCON
Brazil	0.76	0.84
Bulgaria	0.75	0.83
Chinese Taipei	0.83	0.80
Colombia	0.78	0.84
Croatia	0.75	0.76
Cyprus	0.82	0.83
Denmark	0.77	0.71
Estonia	0.78	0.80
France	0.79	0.82
Italy	0.70	0.76
Latvia	0.75	0.81
Lithuania	0.77	0.80
Malta	0.79	0.84
Netherlands	0.81	0.77
North Rhine-Westphalia*	0.76	0.70
Norway	0.85	0.80
Poland	0.73	0.69
Romania	0.59	0.69
Schleswig-Holstein*	0.77	0.77
Serbia	0.72	0.76
Slovak Republic	0.71	0.80
Slovenia	0.82	0.85
Spain	0.80	0.80
Sweden	0.83	0.80
Average	0.77	0.78
Minimum value	0.59	0.69
Maximum value	0.85	0.85

Table 12.23: Partial credit model item parameters for scales measuring students' support for environmental protection and environmental concerns

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_ATTENV	How much do you agree or disagree with the following statements?				
IS4G26A	Governments should focus more on protecting the environment than on supporting economic growth.	0.36	-2.10	-0.40	2.50
IS4G26B	Every citizen needs to contribute to the reduction of pollution.	-0.35	-1.35	-0.78	2.13
IS4G26C	[Country of test] should contribute to protecting the environment in other countries.	0.64	-2.09	-0.29	2.39
IS4G26D	All human beings should take responsibility for preserving the natural world.	-0.22	-1.06	-0.80	1.86
IS4G26E	Countries need to work together to preserve the world's natural resources.	-0.44	-0.91	-0.89	1.80
S_ENVCON	To what extent do you think the following issues are a threat to the world's future?				
IS4G28A	Pollution	-3.41	-0.66	-0.36	1.02
IS4G28D	Climate change	-2.85	-1.19	-0.24	1.43
IS4G28J	[Loss of biodiversity, extinction of living species]	-2.55	-1.38	-0.22	1.60
IS4G28K	Water shortages	-2.70	-0.89	0.10	0.79

Students' citizenship self-efficacy and willingness to participate at school

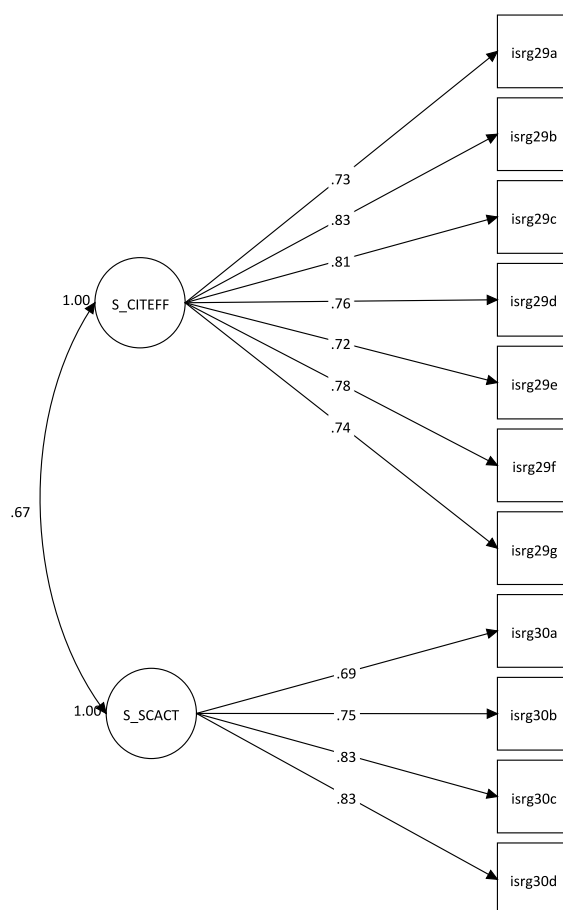
Question 29 of the student questionnaire required students to indicate how well ("very well," "fairly well," "not very well," or "not at all") they would do different activities related to civic engagement. Questions 30 asked students to report how likely ("very likely," "quite likely," "not very likely," or "not at all likely") they expected to undertake different civic engagement activities at school.

Two scales were derived from these two item sets. All seven items in Question 29 were used to derive a scale reflecting *students' citizenship self-efficacy* (S_CITEFF), while all four items included in Question 30 were used to derive a scale of *students' willingness to participate in school activities* (S_SCACT). Higher scale scores reflect higher levels of self-efficacy for S_CITEFF, and higher levels of willingness for S_SCACT. S_CITEFF was equated to ICCS 2009, while S_SCACT was linked to ICCS 2016, when it was first introduced.

We used confirmatory factor analyses assuming a two-dimensional model for these items (Figure 12.13). The model fit was satisfactory for the pooled data set and the two latent factors were highly correlated ($r = 0.67$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained marginally satisfactory across models with increasing constraints.

The average reliabilities (Cronbach's alpha) across national samples were 0.87 for S_CITEFF (ranging from 0.80 to 0.92) and 0.79 for S_SCACT (ranging from 0.71 to 0.88) (see Table 12.24). These scales were based on the item parameters shown in Table 12.25.

Figure 12.13: Confirmatory factor analysis of items measuring students' citizenship self-efficacy and willingness to participate at school



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.079	0.092	0.090	0.095
CFI	0.97	0.97	0.96	0.94
TLI	0.96	0.96	0.96	0.95

Table 12.24: Reliabilities for scales measuring students' citizenship self-efficacy and willingness to participate at school

Country	S_CITEFF	S_SCACT
Brazil	0.88	0.80
Bulgaria	0.88	0.81
Chinese Taipei	0.92	0.81
Colombia	0.90	0.79
Croatia	0.86	0.78
Cyprus	0.89	0.85
Denmark	0.87	0.77
Estonia	0.88	0.83
France	0.86	0.73
Italy	0.84	0.77
Latvia	0.89	0.88
Lithuania	0.89	0.82
Malta	0.89	0.83
Netherlands	0.88	0.85
North Rhine-Westphalia*	0.84	0.75
Norway	0.91	0.80
Poland	0.84	0.71
Romania	0.80	0.72
Schleswig-Holstein*	0.83	0.72
Serbia	0.89	0.83
Slovak Republic	0.86	0.81
Slovenia	0.90	0.84
Spain	0.85	0.72
Sweden	0.91	0.84
Average	0.87	0.79
Minimum value	0.80	0.71
Maximum value	0.92	0.88

Table 12.25: Item parameters for scales measuring students' citizenship self-efficacy and willingness to participate at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_CITEFF	How well do you think you would do the following activities?				
IS4G29A	Argue your point of view about a controversial political or social issue	-0.47	-2.10	-0.17	2.27
IS4G29B	Stand as a candidate in a [school election]	0.16	-1.99	-0.14	2.14
IS4G29C	Organise a group of students in order to achieve changes at school	-0.02	-2.02	-0.13	2.15
IS4G29D	Follow a debate about a controversial issue	-0.10	-2.09	-0.20	2.28
IS4G29E	Write a letter or email to a newspaper giving your view on a current issue	0.10	-1.98	-0.10	2.08
IS4G29F	Speak in front of your class about a social or political issue	0.34	-1.78	-0.11	1.89
IS4G29G	Assess the credibility of information about political or social issues	-0.01	-2.00	-0.20	2.21
S_SCACT	If you were given the chance, how likely is it that you would participate in each activity?				
IS4G30A	Vote in a school election of [class representatives] or [school parliament/council]	-0.75	-1.14	-0.13	1.27
IS4G30B	Join a group of students campaigning for an issue you agree with	-0.15	-1.85	-0.09	1.94
IS4G30C	Become a candidate for [class representative] or [school parliament/council]	0.54	-1.50	0.25	1.25
IS4G30D	Take part in discussions in a [student assembly/gathering]	0.36	-1.60	0.03	1.57

Students' expected participation in activities to express their opinions

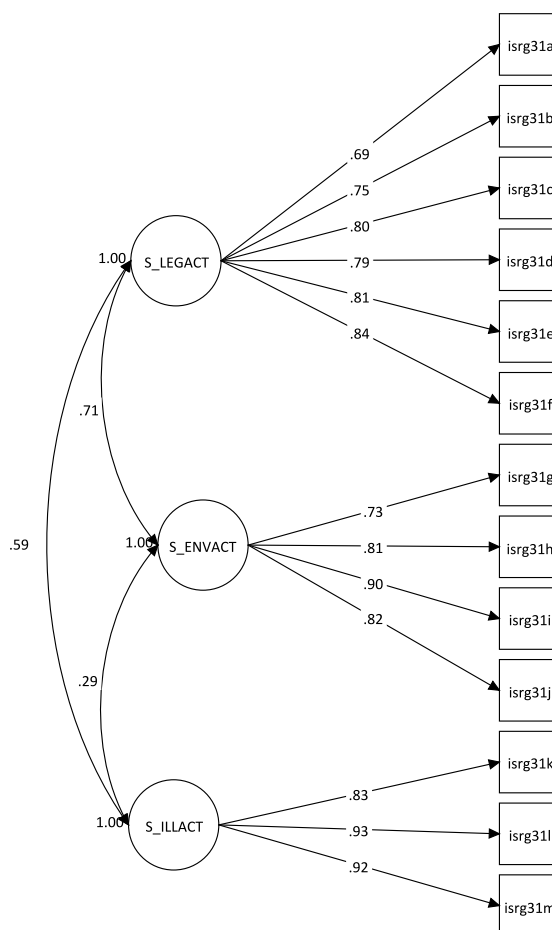
Question 31 of the student questionnaire asked students to indicate whether they would expect taking part in different activities to express their opinion in the future ("I would certainly do this," "I would probably do this," "I would probably not do this," "I would certainly not do this").

Three scales were derived from this item set. Five of the items were used to derive a scale reflecting *students' expected participation in legal activities* (S_LEGACT), four of the items were used to derive a scale of *students' expected participation in environmental protection activities* (S_ENVACT), and three items were used to derive a scale reflecting *students' expected participation in illegal protest activities* (S_ILLACT). Higher scale scores on all scales reflect higher levels of expectations for each type of engagement. S_LEGACT and S_ILLACT were equated to ICCS 2016.

We used confirmatory factor analyses assuming a three-dimensional model for these items (Figure 12.14). The model fit was marginally satisfactory for the pooled sample (according to the RMSEA), and all three latent factors were positively correlated, with the strongest correlation observed between S_LEGACT and S_ENVACT ($r = 0.71$), while the weakest correlation was found between S_ILLACT and S_ENVACT (0.29). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was equally marginally satisfactory across models with different constraints.

The average reliabilities (Cronbach's alpha) across national samples were 0.86 for S_LEGACT (ranging from 0.80 to 0.91), 0.83 for S_ENVACT (ranging from 0.71 to 0.88), and 0.86 for S_ILLACT (ranging from 0.80 to 0.93) (see Table 12.26). These scales were based on the item parameters shown in Table 12.27.

Figure 12.14: Confirmatory factor analysis of items measuring students' expected participation in activities to express their opinions



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.091	0.101	0.098	0.092
CFI	0.96	0.96	0.96	0.95
TLI	0.95	0.95	0.96	0.96

Table 12.26: Reliabilities for scales measuring students' expected participation in activities to express their opinions

Country	S_LEGACT	S_ENVACT	S_ILLACT
Brazil	0.88	0.78	0.89
Bulgaria	0.87	0.85	0.86
Chinese Taipei	0.91	0.84	0.93
Colombia	0.90	0.83	0.89
Croatia	0.83	0.81	0.84
Cyprus	0.89	0.85	0.86
Denmark	0.81	0.81	0.90
Estonia	0.85	0.82	0.86
France	0.84	0.82	0.84
Italy	0.84	0.81	0.82
Latvia	0.88	0.86	0.87
Lithuania	0.88	0.86	0.89
Malta	0.88	0.83	0.88
Netherlands	0.86	0.83	0.86
North Rhine-Westphalia*	0.82	0.82	0.86
Norway	0.91	0.88	0.90
Poland	0.80	0.83	0.83
Romania	0.81	0.71	0.80
Schleswig-Holstein*	0.82	0.82	0.85
Serbia	0.86	0.82	0.85
Slovak Republic	0.84	0.81	0.87
Slovenia	0.90	0.87	0.89
Spain	0.85	0.81	0.84
Sweden	0.87	0.86	0.89
Average	0.86	0.83	0.86
Minimum value	0.80	0.71	0.80
Maximum value	0.91	0.88	0.93

Table 12.27: Partial credit model item parameters for scales measuring students' expected participation in activities to express their opinions

Scale or item	Question/Item wording	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?				
IS4G31A	Talk to others about your views on political or social issues	-0.83	-2.09	-0.07	2.16
IS4G31B	Contact an [elected representative]	0.45	-2.50	0.29	2.21
IS4G31C	Take part in a peaceful march or rally	-0.08	-2.26	0.12	2.14
IS4G31D	Collect signatures for a petition	-0.05	-2.26	0.07	2.19
IS4G31E	Contribute to an online discussion about social or political issues	0.08	-2.33	0.20	2.14
IS4G31F	Organize an online campaign in support of a political or social issue	0.43	-2.42	0.31	2.11
S_ENVACT	Would you take part in any of the following activities to express your opinion in the future?				
IS4G31G	Refuse to buy products that are harmful for the environment	0.11	-1.90	-0.25	2.15
IS4G31H	Tell someone to stop causing damage to the environment	-0.31	-1.84	-0.28	2.13
IS4G31I	Participate in an organized protest to demand more action to protect our environment	0.44	-2.22	0.02	2.20
IS4G31J	Encourage other people to make personal efforts to help the environment (e.g. through saving water)	-0.24	-1.89	-0.39	2.28
S_ILLACT	Would you take part in any of the following activities to express your opinion in the future?				
IS4G31K	Spray-paint protest slogans on walls	-0.21	-2.58	0.36	2.22
IS4G31L	Stage a protest by blocking traffic	0.10	-2.57	0.34	2.23
IS4G31M	Occupy public buildings as a sign of protest	0.11	-2.46	0.41	2.05

Students' expected political participation

Question 32 of the student questionnaire asked students to indicate whether they would expect taking part in different types of political participation in the future ("I would certainly do this," "I would probably do this,"

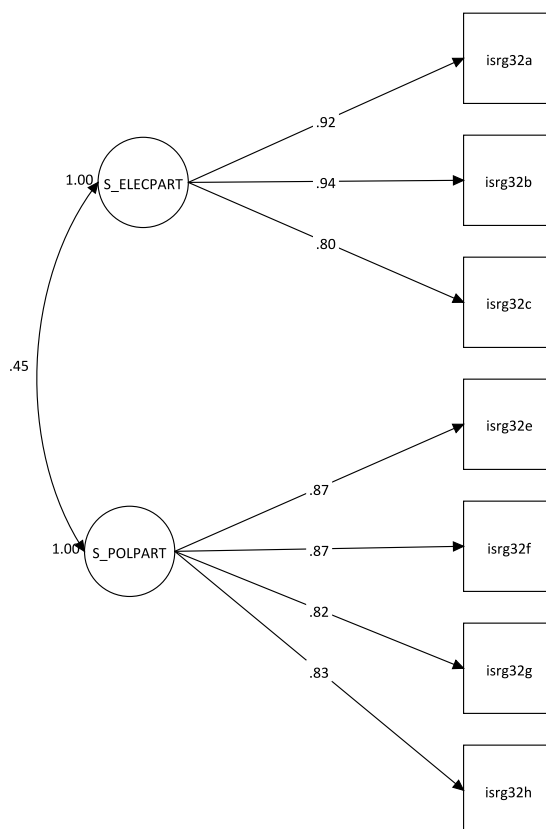
“I would probably not do this,” “I would certainly not do this”).

Two scales were derived from this item set. Three of the items in Question 32 were used to derive a scale reflecting *students’ expected electoral participation* (S_ELECPART), while four further items included in Question 32 were used to derive a scale of *students’ expected active political participation* (S_POLPART). Higher scale scores on both scales reflect higher levels of expected participation. Both scales were equated to ICCS 2009.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students’ expected political participation (Figure 12.15). The model fit was satisfactory for the pooled data set, and the two latent factors were moderately correlated ($r = 0.45$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was poor for more constrained models.¹²

The average reliabilities (Cronbach’s alpha) across national samples were 0.86 for S_ELECPART (ranging from 0.79 to 0.92) and 0.86 for S_POLPART (ranging from 0.74 to 0.91) (see Table 12.28). These scales were based on the item parameters shown in Table 12.29.

Figure 12.15: Confirmatory factor analysis of items measuring students’ expected political participation



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.076	*	0.107	0.111
CFI	0.99	*	0.99	0.98
TLI	0.99	*	0.99	0.98

* Model did not converge.

¹² The scalar model did not converge for this item set and could not be reported.

Table 12.28: Reliabilities for scales measuring students' expected political participation

Country	S_ELECPART	S_POLPART
Brazil	0.87	0.89
Bulgaria	0.86	0.91
Chinese Taipei	0.92	0.90
Colombia	0.88	0.90
Croatia	0.84	0.88
Cyprus	0.89	0.89
Denmark	0.83	0.76
Estonia	0.90	0.88
France	0.82	0.79
Italy	0.87	0.85
Latvia	0.91	0.90
Lithuania	0.88	0.88
Malta	0.86	0.88
Netherlands	0.87	0.91
North Rhine-Westphalia*	0.85	0.85
Norway	0.91	0.91
Poland	0.79	0.74
Romania	0.81	0.84
Schleswig-Holstein*	0.83	0.82
Serbia	0.87	0.90
Slovak Republic	0.86	0.87
Slovenia	0.90	0.89
Spain	0.89	0.84
Sweden	0.85	0.86
Average	0.86	0.86
Minimum value	0.79	0.74
Maximum value	0.92	0.91

Table 12.29: Item parameters for scales measuring students' expected political participation

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_ELECPART	When you are an adult, what do you think you will do?				
IS4G32A	Vote in [local elections]	0.01	-2.21	-0.49	2.70
IS4G32B	Vote in [national elections]	-0.07	-2.27	-0.41	2.68
IS4G32C	Get information about candidates before voting in an election	0.06	-2.32	-0.25	2.57
S_POLPART	When you are an adult, what do you think you will do?				
IS4G32E	Join a political party	0.08	-2.66	0.61	2.05
IS4G32F	Join a trade union	-0.06	-2.94	0.49	2.45
IS4G32G	Stand as a candidate in [local elections]	0.21	-2.45	0.52	1.92
IS4G32H	Join an organisation for a political or social cause	-0.24	-2.87	0.36	2.51

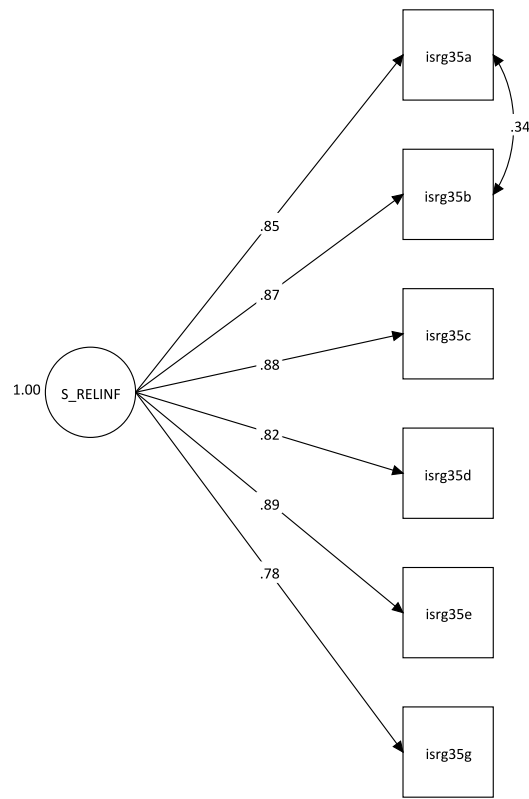
Students' attitudes toward religious influence in society

Question 35 of the student questionnaire was optional for countries, and it asked students to indicate their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with different statements about religion or activities outside of school. Six out of seven items in this question were used to derive a scale of *students' positive attitudes toward the influence of religion in society* (S_RELINF). Higher scale scores on both scales reflect more positive attitudes, and the scale was equated to ICCS 2009.

We used confirmatory factor analyses assuming a one-dimensional model for these items (Figure 12.16). The model fit was satisfactory for the pooled data set, and the model included a correlated error term for two items with similar content. When reviewing measurement invariance across multiple-group models with different constraints, the model fit was only marginally satisfactory for more constrained models, indicated some lack of measurement invariance across countries.

The average reliabilities (Cronbach's alpha) across national samples for this scale were 0.90 (ranging from 0.87 to 0.93) (see Table 12.30). The scale was based on the item parameters shown in Table 12.31.

Figure 12.16: Confirmatory factor analysis of items measuring students' positive attitudes toward religious influence in society



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.068	0.076	0.092	0.096
CFI	1.00	1.00	0.99	0.98
TLI	0.99	0.99	0.99	0.99

Table 12.30: Reliabilities for scale measuring students' positive attitudes toward religious influence in society

Country	S_RELINF
Brazil	0.87
Bulgaria	0.90
Chinese Taipei	0.89
Colombia	0.90
Croatia	0.89
Cyprus	0.90
Denmark	0.90
Estonia	-
France	-
Italy	-
Latvia	0.92
Lithuania	0.92
Malta	0.90
Netherlands	0.91
North Rhine-Westphalia*	0.92
Norway	0.92
Poland	0.90
Romania	0.88
Schleswig-Holstein*	0.92
Serbia	0.87
Slovak Republic	0.90
Slovenia	0.93
Spain	0.92
Sweden	0.92
Average	0.90
Minimum value	0.87
Maximum value	0.93

Table 12.31: Item parameters for scale measuring students' positive attitudes toward religious influence in society

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
S_RELINF	How much do you agree or disagree with the following statements about religion?				
IS4G35A	Religion is more important to me than what is happening in national politics.	-0.44	-2.25	0.28	1.97
IS4G35B	Religion helps me to decide what is right and what is wrong.	-0.40	-2.29	-0.12	2.41
IS4G35C	Religious leaders should have more power in society.	0.39	-2.74	0.32	2.42
IS4G35D	Religion should influence people's behaviour towards others.	-0.28	-2.33	-0.24	2.57
IS4G35E	Rules of life based on religion are more important than civil laws.	0.29	-2.69	0.32	2.36
IS4G35G	Religious people are better citizens.	0.44	-2.41	0.33	2.08

European student questionnaire

Students' European identity and perceptions of learning about Europe at school

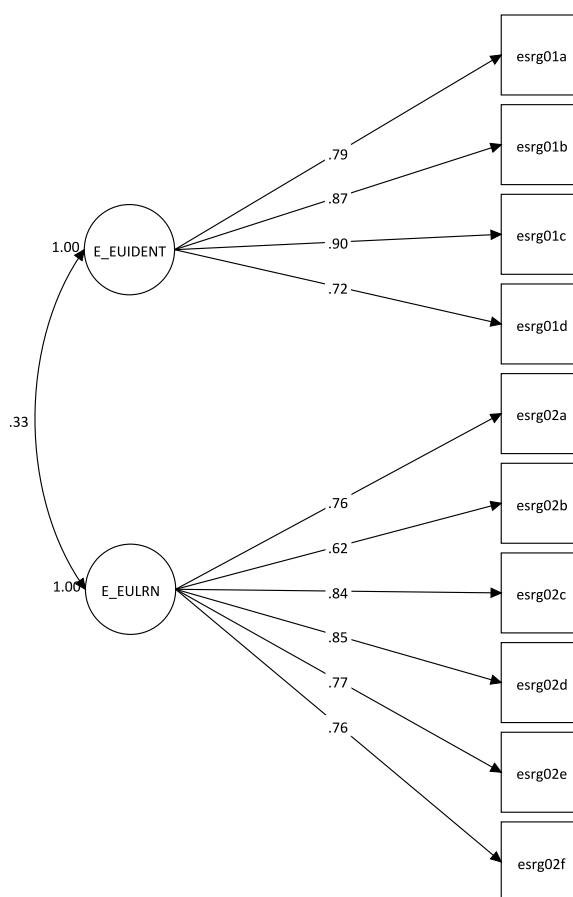
Question 1 of the European student questionnaire required students to indicate their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with statements about their European identity. In Question 2 of this instrument students reported about the extent that they had learned about European issues or topics at school using the response categories "to a large extent," "to a moderate extent," "to a small extent," or "not at all."

Two scales were derived from these two item sets. Four of the items in Question 1 were used to derive a scale reflecting *students' sense of European identity* (E_EUIDENT), while five of the items included in Question 2 were used to derive a scale of *students' reports on opportunities for learning about Europe at school* (E_EURLRN). Higher scale scores on E_EUIDENT indicate higher levels of European identity, while on E_EURLRN they reflect higher levels of having learned about European issues or topics. E_EUIDENT was equated to ICCS 2009, while, due to some item modifications from a similar question in the previous survey, E_EURLN was treated as a new scale for ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' European identity and learning about Europe (Figure 12.17). The model fit was satisfactory for the pooled dataset, and the two latent factors were moderately correlated ($r = 0.33$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit was somewhat poorer for the most constrained (scalar) model.

The average reliabilities (Cronbach's alpha) across national samples were 0.81 for E_EUIDENT (ranging from 0.72 to 0.86) and 0.82 for E_EURLN (ranging from 0.77 to 0.88) (see Table 12.32). These scales were based on the item parameters shown in Table 12.33.

Figure 12.17: Confirmatory factor analysis of items measuring students' European identity and perceptions of learning about Europe at school



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.060	0.069	0.066	0.076
CFI	0.98	0.98	0.98	0.96
TLI	0.98	0.98	0.98	0.97

Table 12.32: Reliabilities for scales measuring students' European identity and perceptions of learning about Europe at school

Country	E_EUIDENT	E_EULRN
Bulgaria	0.85	0.84
Croatia	0.84	0.82
Cyprus	0.85	0.88
Denmark	0.80	0.79
Estonia	0.84	0.83
France	0.80	0.82
Italy	0.79	0.78
Latvia	0.83	0.85
Lithuania	0.79	0.82
Malta	0.82	0.86
Netherlands	0.82	0.83
North Rhine-Westphalia*	0.78	0.82
Norway	0.86	0.84
Poland	0.79	0.82
Romania	0.72	0.77
Schleswig-Holstein*	0.76	0.81
Slovak Republic	0.76	0.81
Slovenia	0.83	0.82
Spain	0.82	0.84
Sweden	0.82	0.84
Average	0.81	0.82
Minimum value	0.72	0.77
Maximum value	0.86	0.88

Table 12.33: Item parameters for scales measuring students' European identity and perceptions of learning about Europe at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_EUIDENT	How much do you agree or disagree with the following statements?				
ES4G01A	I see myself as European.	-0.58	-1.28	-1.12	2.40
ES4G01B	I am proud to live in Europe.	-0.43	-1.98	-1.14	3.12
ES4G01C	I feel part of Europe.	0.05	-2.47	-0.70	3.17
ES4G01D	I see myself first as a citizen of Europe and then as a citizen of the world.	0.96	-2.65	-0.39	3.04
E_EULRN	At school, to what extent have you learned about the following issues or topics?				
ES4G02A	Political and economic systems of other European countries	0.44	-2.28	-0.38	2.66
ES4G02B	The history of Europe	-1.03	-1.86	-0.37	2.23
ES4G02C	Political and social issues in other European countries	0.47	-2.54	-0.21	2.75
ES4G02D	Political and economic cooperation between European countries	0.51	-2.35	-0.31	2.66
ES4G02E	The European Union	-0.39	-2.10	-0.27	2.37

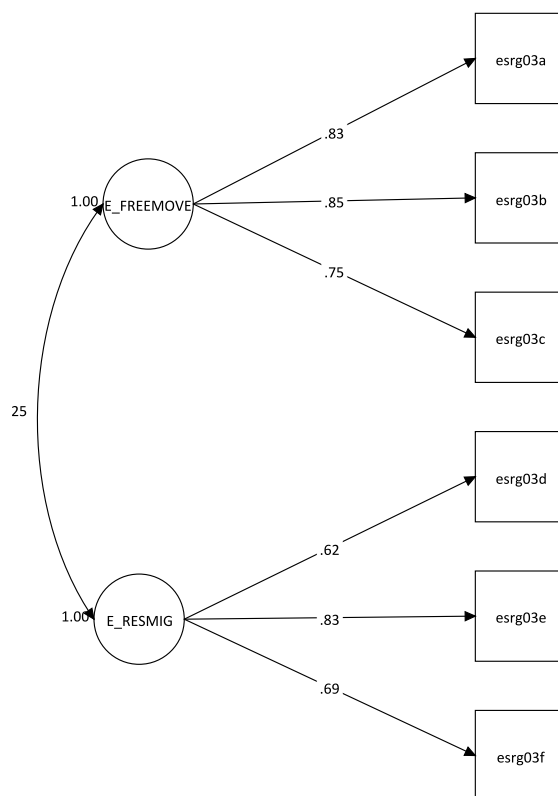
Students' attitudes toward freedom of movement in Europe

Question 3 of the European student questionnaire required students to indicate their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with positively and negatively worded statements about freedom of movement in Europe. Two scales were derived from these two item sets. The three positively worded items were used to derive a scale reflecting *students' endorsement of freedom of movement within Europe* (E_FREEMOVE), while the three negatively worded the items were used to derive a scale of *students' endorsement of restrictions of movement in Europe* (E_RESMIG). Higher scale scores on E_FREEMOVE indicate higher levels of endorsement of freedom of movement, while on E_RESMIG they reflect higher levels of support for restriction of movement within Europe. E_FREEMOVE was equated to ICCS 2016, while, due to some item modifications, E_RESMIG was treated as a new scale for ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' attitudes toward freedom of movement in Europe (Figure 12.17). The RMSEA fit index indicated a non-satisfactory model fit for the pooled dataset, while CFI and TLI indicated marginally satisfactory fit. The two latent factors were moderately correlated ($r = 0.25$), suggesting that many students expressed both positive attitudes as well as support for restricting freedom of movement. When reviewing measurement invariance across multiple-group models with different constraints, the model fit was not satisfactory according to the RMSEA, and only marginally satisfactory according to the other indices across multiple-group models with varying constraints.

The average reliabilities (Cronbach's alpha) across national samples were 0.77 for E_FREEMOVE (ranging from 0.69 to 0.82) and 0.68 for E_RES MIG (ranging from 0.56 to 0.78) (see Table 12.34). These scales were based on the item parameters shown in Table 12.35.

Figure 12.18: Confirmatory factor analysis of items measuring students' attitudes toward freedom of movement in Europe



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.119	0.128	0.110	0.093
CFI	0.96	0.96	0.96	0.94
TLI	0.92	0.92	0.94	0.96

Table 12.34: Reliabilities for scales measuring students' attitudes toward freedom of movement in Europe

Country	E_FREEMOVE	E_RES MIG
Bulgaria	0.79	0.74
Croatia	0.81	0.56
Cyprus	0.82	0.73
Denmark	0.76	0.62
Estonia	0.80	0.70
France	0.72	0.66
Italy	0.69	0.67
Latvia	0.78	0.70
Lithuania	0.75	0.69
Malta	0.80	0.70
Netherlands	0.80	0.70
North Rhine-Westphalia*	0.76	0.61
Norway	0.82	0.77
Poland	0.74	0.62
Romania	0.69	0.64
Schleswig-Holstein*	0.74	0.60
Slovak Republic	0.76	0.63
Slovenia	0.79	0.69
Spain	0.74	0.68
Sweden	0.81	0.78
Average	0.77	0.68
Minimum value	0.69	0.56
Maximum value	0.82	0.78

Table 12.35: Item parameters for scales measuring students' attitudes toward freedom of movement in Europe

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_FREEMOVE	How much do you agree or disagree with the following statements related to the freedom for European citizens to work in other European countries?				
ES4G03A	Allowing citizens of European countries to work anywhere in Europe is good for the European economy.	-0.16	-2.20	-1.47	3.67
ES4G03B	Citizens of European countries should be allowed to work anywhere in Europe.	-0.13	-2.58	-1.07	3.65
ES4G03C	Allowing citizens of European countries to work anywhere in Europe helps to reduce unemployment.	0.29	-2.80	-0.86	3.65
E_RES MIG	How much do you agree or disagree with the following statements related to the freedom for European citizens to work in other European countries?				
ES4G03D	The freedom for citizens of European countries to work anywhere in Europe should be limited.	0.63	-2.33	0.40	1.93
ES4G03E	The freedom for citizens of European countries to work in another European country should be regulated by agreements between individual countries.	-0.30	-2.42	-0.38	2.80
ES4G03F	Citizens of European countries seeking to work in another European country should apply for work permits like people from outside Europe.	-0.34	-2.30	-0.35	2.65

Students' support for European cooperation

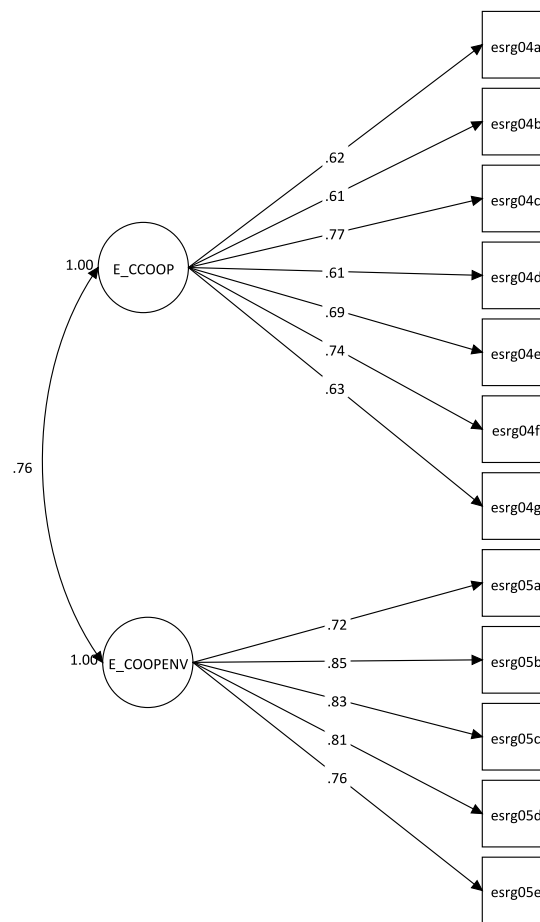
Question 4 of the European student questionnaire required students to indicate their level of agreement ("strongly agree," "agree," "disagree," or "strongly disagree") with statements about different types of cooperation in Europe. Question 5, using the same response categories, asked specifically about agreement with environmental cooperation across European countries.

Two scales were derived from these two item sets. All seven of the items in Question 4 were used to derive a scale reflecting *students' support for cooperation among European countries* (E_CCOOP), while all five items included in Question 5 were used to derive a scale of *students' endorsement of environmental cooperation in Europe* (E_COOPENV). Higher scale scores on both scales reflect more support for cooperation in Europe. Both scales were new for ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' support for cooperation in Europe (Figure 12.19). The model fit was satisfactory for the pooled sample, and the two latent factors were strongly correlated ($r = 0.76$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained satisfactory across multiple-group models with different constraints.

The average reliabilities (Cronbach's alpha) across national samples were 0.78 for E_CCOOP (ranging from 0.70 to 0.84) and 0.83 for E_COOPENV (ranging from 0.76 to 0.91) (see Table 12.36). These scales were based on the item parameters shown in Table 12.37.

Figure 12.19: Confirmatory factor analysis of items measuring students' support for European cooperation



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.055	0.065	0.056	0.063
CFI	0.97	0.97	0.97	0.96
TLI	0.97	0.96	0.97	0.97

Table 12.36: Reliabilities for scales measuring students' support for European cooperation

Country	E_CCOOP	E_COOPENV
Bulgaria	0.84	0.83
Croatia	0.80	0.83
Cyprus	0.83	0.85
Denmark	0.77	0.82
Estonia	0.78	0.82
France	0.75	0.81
Italy	0.70	0.78
Latvia	0.83	0.84
Lithuania	0.83	0.81
Malta	0.82	0.85
Netherlands	0.76	0.83
North Rhine-Westphalia*	0.74	0.81
Norway	0.78	0.89
Poland	0.74	0.82
Romania	0.72	0.76
Schleswig-Holstein*	0.71	0.81
Slovak Republic	0.77	0.82
Slovenia	0.83	0.84
Spain	0.75	0.83
Sweden	0.83	0.91
Average	0.78	0.83
Minimum value	0.70	0.76
Maximum value	0.84	0.91

Table 12.37: Item parameters for scales measuring students' support for European cooperation

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_CCOOP	How much do you agree or disagree with the following statements?				
ES4G04A	European countries should recognize all educational qualifications achieved in any other European country.	-0.29	-1.52	-0.90	2.42
ES4G04B	European countries should have a European army for international missions.	0.04	-1.98	-0.63	2.61
ES4G04C	European countries should adopt common rules to prevent and combat terrorism.	-0.37	-1.43	-0.91	2.34
ES4G04D	European countries should adopt the same regulations to combat illegal entry from non-European countries.	0.36	-1.82	-0.64	2.46
ES4G04E	European countries should have the same rules regarding the acceptance of people escaping persecution in their countries for reasons of nationality, ethnicity, religion, or political opinions.	0.08	-1.73	-0.66	2.39
ES4G04F	European countries should adopt common rules to reduce social and economic inequalities between rich and poor people.	-0.01	-1.59	-0.77	2.37
ES4G04G	European countries should have common rules to combat infectious diseases (e.g. [measles, COVID-19]).	0.19	-1.49	-0.41	1.90
E_COOPENV	How much do you agree or disagree with the following statements?				
ES4G05A	European countries should have the same rules to protect the environment.	-0.09	-2.27	-0.58	2.85
ES4G05B	European countries should adopt common protocols and rules to reduce climate change.	-0.06	-2.34	-0.86	3.20
ES4G05C	European countries should promote an economic growth sustainable for the environment.	-0.09	-2.19	-1.10	3.29
ES4G05D	European countries should promote the use of renewable energy sources.	0.01	-1.68	-1.16	2.84
ES4G05E	European countries should encourage the use of [clean technologies] in countries outside Europe.	0.23	-2.05	-0.93	2.98

Students' expectations for Europe's future

Question 7 of the student questionnaire required students to indicate the extent to which they rated the likelihood of different possible scenarios for Europe's future in a series of activities outside of school.

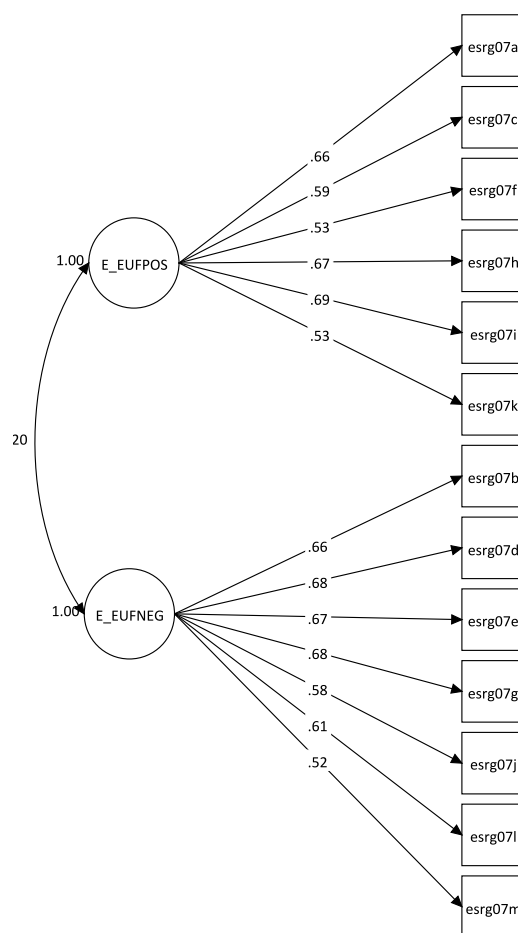
Students used the response categories “very likely,” “likely,” “unlikely,” and “very unlikely.”

Two scales were derived from these two item sets. Six of the items were used to derive a scale reflecting *students’ positive expectations for European future* (E_EUFPOS), while seven of the items included in this question were used to derive a scale of *students’ negative expectations for European future* (E_EUFNEG). Higher scale scores on both scales reflect more endorsement of either positive or negative scenarios for Europe’s future. Both scales were new to ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students’ expectations for Europe’s future (Figure 12.20). The model fit was marginally satisfactory for the pooled sample, and the two latent factors were positively (albeit weakly) correlated ($r = 0.20$). When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained marginally satisfactory across models.

The average reliabilities (Cronbach’s alpha) across national samples were 0.70 for E_EUFPOS (ranging from 0.61 to 0.80) and 0.68 for E_EUFNEG (ranging from 0.54 to 0.77) (see Table 12.38). These scales were based on the item parameters shown in Table 12.39.

Figure 12.20: Confirmatory factor analysis of items measuring students’ expectations for Europe’s future



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.062	0.076	0.073	0.077
CFI	0.92	0.89	0.89	0.84
TLI	0.90	0.87	0.88	0.87

Table 12.38: Reliabilities for scales measuring students' expectations for Europe's future

Country	E_EUFPOS	E_EUATNEG
Bulgaria	0.80	0.74
Croatia	0.71	0.63
Cyprus	0.78	0.76
Denmark	0.65	0.65
Estonia	0.69	0.72
France	0.70	0.65
Italy	0.66	0.66
Latvia	0.73	0.67
Lithuania	0.69	0.73
Malta	0.74	0.74
Netherlands	0.66	0.69
North Rhine-Westphalia*	0.64	0.58
Norway	0.72	0.77
Poland	0.66	0.62
Romania	0.69	0.60
Schleswig-Holstein*	0.61	0.54
Slovak Republic	0.70	0.64
Slovenia	0.77	0.74
Spain	0.69	0.65
Sweden	0.73	0.75
Average	0.70	0.68
Minimum value	0.61	0.54
Maximum value	0.80	0.77

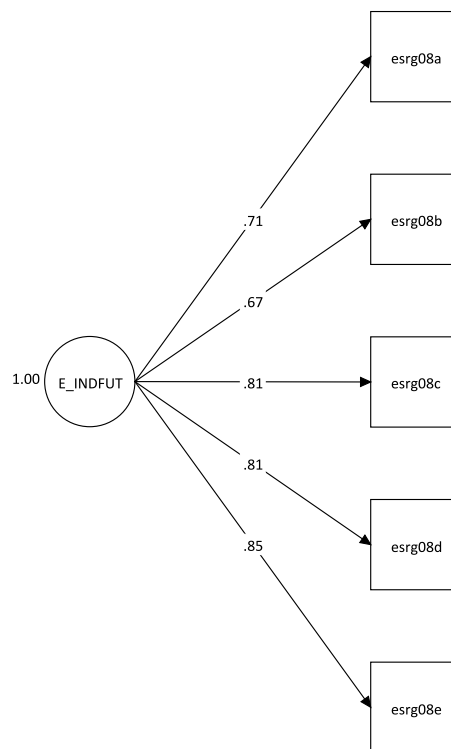
Table 12.39: Item parameters for scale measuring students' expectations for Europe's future

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_EUFPOS	What is Europe likely to look like in 10 years?				
ES4G07A	There will be stronger cooperation among European countries.	-0.61	-1.59	-0.71	2.30
ES4G07C	There will be peace across Europe.	0.52	-2.08	-0.04	2.12
ES4G07F	There will be less air and water pollution in Europe.	0.62	-1.97	-0.03	2.00
ES4G07H	Democracy will be strengthened across Europe.	-0.20	-2.17	-0.51	2.68
ES4G07I	Poor people will have more access to healthcare.	-0.14	-1.99	-0.49	2.49
ES4G07K	There will be more women among political leaders.	-0.18	-1.74	-0.42	2.15
E_EUFNEG	What is Europe likely to look like in 10 years?				
ES4G07B	There will be a rise in racism.	0.26	-2.29	0.43	1.86
ES4G07D	Terrorism will be more of a threat all across Europe.	0.23	-2.46	0.23	2.23
ES4G07E	There will be larger economic differences between rich and poor countries in Europe.	-0.50	-2.34	-0.14	2.48
ES4G07G	There will be a rise in poverty and unemployment in Europe.	-0.06	-2.46	0.16	2.30
ES4G07J	Politics will be increasingly influenced by small groups of rich people.	-0.18	-2.08	-0.12	2.20
ES4G07L	There will be a rise in religious intolerance.	0.16	-2.36	0.10	2.27

Students' expectations for their own future

Question 8 of the student questionnaire required respondents to indicate the extent to which they rated the likelihood of different statements about their own future, using the response categories “very likely,” “likely,” “unlikely,” and “very unlikely.” All of the five items were used to derive a scale reflecting *students' expectations for their own individual future* (E_INDFUT), where higher scale scores on both scales reflect more positive expectations. The scale was equated to ICCS 2016 to allow comparisons of scale scores across the two last cycles.

Figure 12.21: Confirmatory factor analysis of items measuring students' expectations for their own future



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.052	0.062	0.044	0.057
CFI	1.00	1.00	1.00	0.99
TLI	0.99	0.99	1.00	0.99

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess students' expectations for their own future (Figure 12.21). The model fit was satisfactory for the pooled sample, and, when reviewing measurement invariance across multiple-group models with different constraints, the model fit remained satisfactory across models.

The average reliability (Cronbach's alpha) across national samples for E_INDFUT was 0.81 (ranging from 0.76 to 0.86) (see Table 12.40). The scale was based on the item parameters shown in Table 12.41.

Table 12.40 Reliabilities for scale measuring students' expectations for their own future

Country	E_IND FUT
Bulgaria	0.85
Croatia	0.80
Cyprus	0.86
Denmark	0.79
Estonia	0.85
France	0.81
Italy	0.78
Latvia	0.85
Lithuania	0.82
Malta	0.82
Netherlands	0.80
North Rhine-Westphalia*	0.76
Norway	0.85
Poland	0.79
Romania	0.78
Schleswig-Holstein*	0.77
Slovak Republic	0.80
Slovenia	0.80
Spain	0.82
Sweden	0.83
Average	0.81
Minimum value	0.76
Maximum value	0.86

Table 12.41 Item parameters for scale measuring students' expectations for their own future

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_IND FUT	How likely do you think it is that your future will look like this?				
ES4G08A	I will find a steady job.	-0.51	-1.71	-1.13	2.83
ES4G08B	My financial situation will be better than that of my parents.	0.48	-3.20	-0.37	3.57
ES4G08C	I will find a job I like.	-0.10	-2.26	-0.56	2.81
ES4G08D	I will have the opportunity to travel abroad for leisure.	0.20	-1.94	-0.71	2.66
ES4G08E	I will earn enough money to start a family.	-0.07	-1.40	-1.16	2.56

Students' reports on ethical consumerism and sustainable behaviors

Question 10 of the European student questionnaire asked students how often ("often," "sometimes," "rarely," or "never") they or their parents had bought only certain products or refused to buy products for ethical reasons. Using the same response categories, Question 11 gathered data about students' sustainable behaviors (such as reducing water use or avoiding plastic packaging).

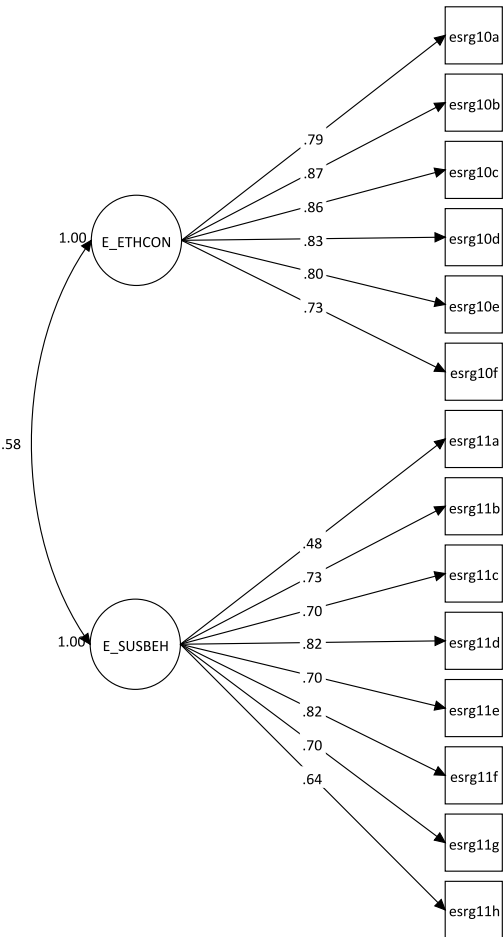
Two scales were derived from these two item sets. All six items in Question 10 were used to derive a scale reflecting *students' reports on political consumerism behaviors* (E_ETHCON), while all eight of the items included in Question 11 were used to derive a scale of *students' reports on their sustainable behaviors* (E_SUSBEH). Higher scale scores on both scales reflect more frequent engagement in each type of activities. Both scales were new for ICCS 2022.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' reports on ethical consumerism and sustainable (Figure 12.22). The model fit was only marginally satisfactory for the pooled sample, and the two latent factors were highly and positively correlated ($r = 0.58$).

When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained marginally satisfactory across models.

The average reliabilities (Cronbach’s alpha) across national samples were 0.87 for E_ETHCON (ranging from 0.78 to 0.92) and 0.84 for E_SUSBEH (ranging from 0.79 to 0.88) (see Table 12.42). These scales were based on the item parameters shown in Table 12.43.

Figure 12.22: Confirmatory factor analysis of items measuring students' reports on ethical consumerism and sustainable behaviors



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.098	0.104	0.100	0.096
CFI	0.93	0.94	0.93	0.92
TLI	0.92	0.93	0.93	0.94

Table 12.42 Reliabilities for scales measuring students' reports on ethical consumerism and sustainable behaviors

Country	E_ETHCON	E_SUSBEH
Bulgaria	0.84	0.87
Croatia	0.88	0.81
Cyprus	0.87	0.87
Denmark	0.88	0.83
Estonia	0.88	0.87
France	0.89	0.85
Italy	0.87	0.81
Latvia	0.90	0.85
Lithuania	0.85	0.84
Malta	0.87	0.84
Netherlands	0.91	0.88
North Rhine-Westphalia*	0.85	0.82
Norway	0.92	0.87
Poland	0.89	0.80
Romania	0.78	0.79
Schleswig-Holstein*	0.87	0.84
Slovak Republic	0.86	0.83
Slovenia	0.91	0.87
Spain	0.87	0.81
Sweden	0.90	0.87
Average	0.87	0.84
Minimum value	0.78	0.79
Maximum value	0.92	0.88

Table 12.43 Item parameters for scales measuring students' reports on ethical consumerism and sustainable behaviors

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_ETHCON	During the last twelve months, how often have you done or have you asked your [parents or guardians] to do the following things?				
ES4G10A	Refuse to buy goods produced by companies using child labor	0.29	-0.82	-0.39	1.21
ES4G10B	Refuse to buy goods whose production has a negative impact on the environment	-0.08	-1.39	-0.38	1.78
ES4G10C	Refuse to buy goods produced by a company violating social rights of their employees	0.20	-1.26	-0.25	1.51
ES4G10D	Buy only goods that can be recycled afterwards	-0.26	-1.46	-0.25	1.72
ES4G10E	Buy [green products]	-0.60	-1.47	-0.31	1.78
ES4G10F	Get information whether companies are [socially responsible] before buying their products	0.45	-1.26	-0.19	1.45
E_SUSBEH	During the last twelve months, how often have you done each of the actions listed below?				
ES4G11A	Purchase used instead of new clothing	0.96	-0.51	-0.24	0.75
ES4G11B	Reduce water use (e.g. when brushing your teeth, having a shower, washing dishes)	0.01	-1.04	-0.24	1.28
ES4G11C	Reduce the use of electricity (e.g. switching off the lights when leaving a room, turning down the heat when it is not too cold)	-0.47	-0.93	-0.15	1.08
ES4G11D	Avoid buying products with plastic packaging (e.g. school supplies, groceries)	0.46	-1.28	-0.15	1.43
ES4G11E	Reuse old items in good condition instead of buying new ones	-0.25	-1.18	-0.18	1.36
ES4G11F	Limit the use of plastic items (e.g. disposable plastic glasses, water bottles, plastic shopping bags)	0.05	-1.18	-0.15	1.33
ES4G11G	Reduce food waste (e.g. avoiding buying more food than necessary, eating leftovers)	-0.40	-1.04	-0.23	1.28
ES4G11H	Repair rather than replacing items you have (e.g. fix your bike instead of buying a new one, mending a backpack instead of buying a new one)	-0.35	-1.18	-0.19	1.37

Students' attitudes toward the EU

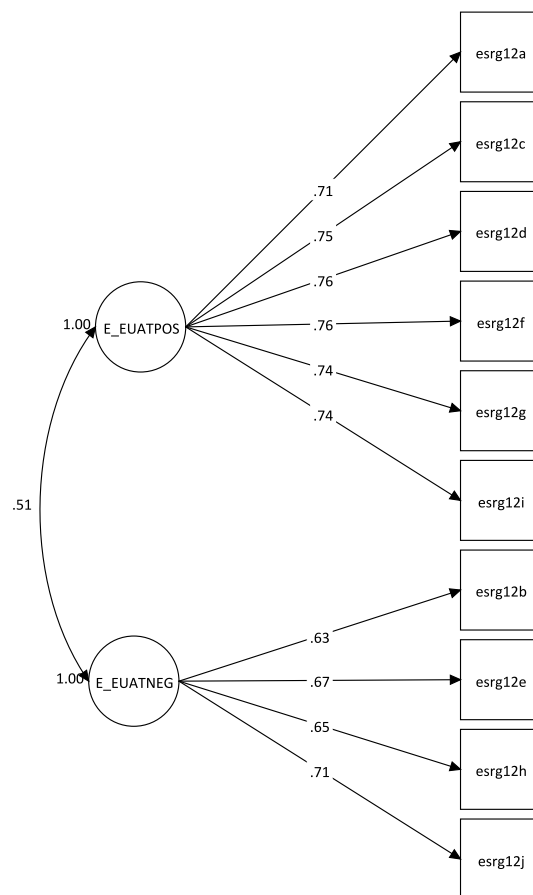
Question 12 of the European student questionnaire asked students to indicate their level of agreement

("strongly agree," "agree," "disagree," or "strongly disagree") with statements about the European Union. Two scales were derived from this item set. Six positively worded items were used to derive a scale reflecting *students' positive attitudes toward European Union* (E_EUATPOS), while four negatively worded items were used to derive a scale of *students' negative attitudes toward European Union* (E_EUATNEG). Higher scale scores reflect more support for either positive or negative statements toward the EU.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess students' attitudes toward the EU (Figure 12.23). The model fit was satisfactory for the pooled sample, and the two latent factors were positively correlated ($r = 0.51$), suggesting that many respondents tended to express both positive and negative perceptions of the EU. When reviewing measurement invariance across multiple-group models with different constraints, the model fit remained satisfactory across multiple-group models with different constraints.

The average reliabilities (Cronbach's alpha) across national samples were 0.82 for E_EUATPOS (ranging from 0.71 to 0.87) and 0.68 for E_EUATNEG (ranging from 0.54 to 0.77) (see Table 12.44). These scales were based on the item parameters shown in Table 12.45.

Figure 12.23: Confirmatory factor analysis of items measuring students' attitudes toward the EU



Multiple-group models				
	Pooled sample	Configural	Metric	Scalar
RMSEA	0.051	0.065	0.058	0.062
CFI	0.98	0.97	0.97	0.96
TLI	0.97	0.96	0.97	0.97

Table 12.44 Reliabilities for scales measuring students' attitudes toward the EU

Country	E_EUATPOS	E_EUATNEG
Bulgaria	0.86	0.74
Croatia	0.84	0.63
Cyprus	0.87	0.76
Denmark	0.76	0.65
Estonia	0.85	0.72
France	0.78	0.65
Italy	0.80	0.66
Latvia	0.83	0.67
Lithuania	0.85	0.73
Malta	0.86	0.74
Netherlands	0.81	0.69
North Rhine-Westphalia*	0.75	0.58
Norway	0.87	0.77
Poland	0.81	0.62
Romania	0.75	0.60
Schleswig-Holstein*	0.71	0.54
Slovak Republic	0.83	0.64
Slovenia	0.85	0.74
Spain	0.81	0.65
Sweden	0.85	0.75
Average	0.82	0.68
Minimum value	0.71	0.54
Maximum value	0.87	0.77

Table 12.45 Item parameters for scales measuring students' attitudes toward the EU

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
E_EUATPOS	How much do you agree or disagree with the following statements?				
ES4G12A	[EU] promotes respect for human rights all over Europe.	-0.41	-2.08	-1.03	3.11
ES4G12C	[EU] takes care of the environment.	0.19	-2.87	-0.57	3.44
ES4G12D	[EU] is good for the economy of individual countries.	0.19	-2.49	-0.99	3.48
ES4G12F	[EU] makes Europe a safe place to live.	0.01	-2.48	-0.80	3.28
ES4G12G	[EU] is good because countries share a common set of rules and laws.	-0.03	-2.44	-0.94	3.38
ES4G12I	[EU] promotes freedom of speech.	0.05	-2.30	-0.84	3.15
E_EUATNEG	How much do you agree or disagree with the following statements?				
ES4G12B	[EU] institutions cost too much money.	-0.13	-3.16	-0.04	3.20
ES4G12E	[EU] policies are too strongly influenced by the richest member states.	-0.27	-2.69	-0.10	2.79
ES4G12H	[EU] is run mainly by unelected bureaucrats.	0.41	-2.85	0.16	2.69
ES4G12J	The adoption of [EU] policies takes too long to be effective.	0.00	-2.56	-0.21	2.77

Teacher questionnaire

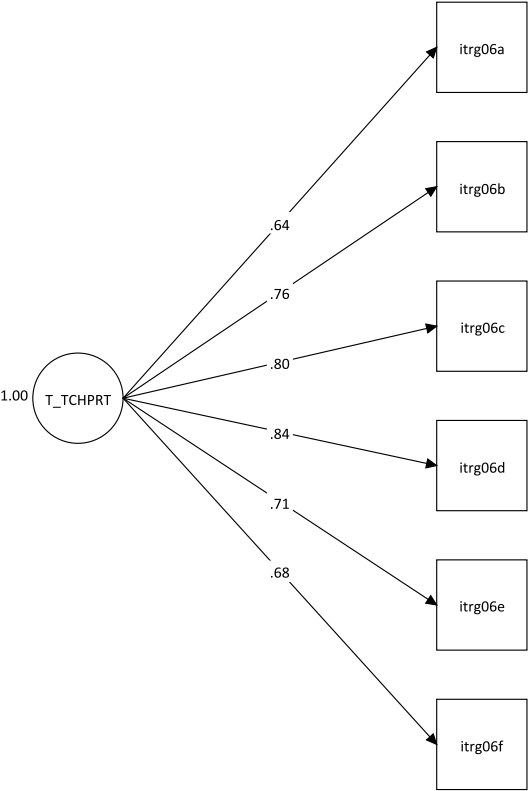
Teachers' perceptions of teacher participation at school

Question 6 of the teacher questionnaire required teachers to give their perspective of the number of teachers that participated in a list of activities at school ("all or nearly all", "most of them", "some of them", or "none or hardly any"). One scale was derived from this item set. The six items were used to derive a scale reflecting *teachers' perception of teacher participation at school* (T_TCHPRT). Higher scale scores reflect higher numbers of teachers participating.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teacher's perception of teacher participation at school (Figure 12.24). The model fit was satisfactory for the pooled sample.

The average reliability (Cronbach's alpha) across national samples was 0.81 for T_TCHPRT (ranging from 0.69 to 0.88) (see Table 12.46). The scale was based on the item parameters shown in Table 12.47.

Figure 12.24 Confirmatory factor analysis of items measuring teachers' perceptions of teacher participation at school



	Pooled sample
RMSEA	0.064
CFI	0.98
TLI	0.97

Table 12.46: Reliabilities for scale measuring teachers' perceptions of teacher participation at school

Country	T_TCHPRT
Brazil	0.88
Bulgaria	0.84
Chinese Taipei	0.81
Colombia	0.83
Croatia	0.82
Cyprus	0.85
Denmark	0.81
Estonia	0.78
France	0.81
Italy	0.84
Latvia	0.81
Lithuania	0.78
Malta	0.81
Netherlands	0.69
North Rhine-Westphalia*	0.75
Norway	0.78
Poland	0.83
Romania	0.84
Serbia	0.83
Slovak Republic	0.84
Slovenia	0.81
Spain	0.84
Sweden	0.77
Average	0.81
Minimum value	0.69
Maximum value	0.88

Table 12.47: Item parameters for scale measuring teachers' perceptions of teacher participation at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_TCHPRT	Teachers' perception of teacher participation at school				
IT4G06A	Working together when devising teaching activities	0.20	-2.97	0.42	2.55
IT4G06B	Helping each other to solve conflicts between students in the school	-0.76	-3.24	0.26	2.97
IT4G06C	Taking on tasks and responsibilities in addition to teaching (such as organising school projects, school trips etc.)	0.32	-3.55	0.57	2.97
IT4G06D	Actively taking part in school [development/improvement activities]	0.12	-3.43	0.48	2.95
IT4G06E	Engaging in [tutoring and counselling activities]	0.70	-3.16	0.73	2.43
IT4G06F	Actively cooperating with [parents/guardians]	-0.58	-2.64	-0.05	2.69

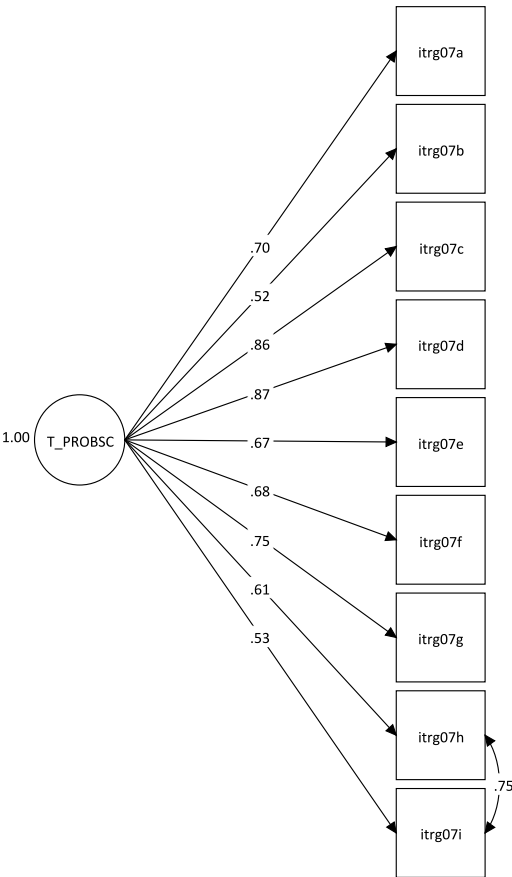
Teachers' perceptions of social problems at school

Question 7 of the teacher questionnaire required teachers to indicate perception of the frequency of social problems occurring among students at school ("never", "sometimes", "often" or "very often"). One scale was derived from this item set. The nine items were used to derive a scale reflecting *teachers' perceptions of social problems at school* (T_PROBSC). Higher scale scores reflect a higher number of social problems.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teacher's perceptions of social problems at school (Figure 12.25). The model fit was satisfactory for the pooled sample.

The average reliability (Cronbach's alpha) across national samples was 0.79 for T_PROBSC (ranging from 0.71 to 0.86) (see Table 12.48). The scale was based on the item parameters shown in Table 12.49.

Figure 12.25 Confirmatory factor analysis of items measuring teachers' perceptions of social problems at school



	Pooled sample
RMSEA	0.064
CFI	0.96
TLI	0.94

Table 12.48: Reliabilities for scale measuring teachers' perceptions of social problems at school

Country	T_PROBSC
Brazil	0.86
Bulgaria	0.81
Chinese Taipei	0.79
Colombia	0.81
Croatia	0.81
Cyprus	0.84
Denmark	0.83
Estonia	0.80
France	0.81
Italy	0.71
Latvia	0.77
Lithuania	0.77
Malta	0.85
Netherlands	0.74
North Rhine-Westphalia*	0.82
Norway	0.74
Poland	0.78
Romania	0.77
Serbia	0.77
Slovak Republic	0.82
Slovenia	0.78
Spain	0.81
Sweden	0.78
Average	0.79
Minimum value	0.71
Maximum value	0.86

Table 12.49: Item parameters for scale measuring teachers' perceptions of social problems at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_PROBSC	Teachers' perceptions of social problems at school				
IT4G07A	Vandalism	-0.74	-3.09	1.20	1.88
IT4G07B	Truancy	-1.63	-3.60	1.26	2.34
IT4G07C	Ethnic intolerance	0.17	-2.69	1.13	1.56
IT4G07D	Religious intolerance	0.84	-2.28	0.90	1.39
IT4G07E	[Bullying]	-1.19	-3.69	1.21	2.48
IT4G07F	Violence	-0.42	-3.35	1.32	2.03
IT4G07G	Sexual harassment	0.97	-2.20	1.21	1.00
IT4G07H	Drug abuse	1.09	-1.84	1.51	0.33
IT4G07I	Alcohol abuse	0.92	-1.94	1.03	0.92

Teachers' reports of student activities in the community

Question 8 of the teacher questionnaire required teachers to indicate whether they and their students participated ("yes" or "no") in a list of activities in cooperation with external groups/organizations. One scale was derived from this item set. The ten items were used to derive a scale reflecting *teachers' reports of student activities in the community* (T_STDCOM). Higher scale scores reflect a higher number of student activities.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teachers' reports of student activities in the community (Figure 12.26). The model fit was satisfactory for the pooled sample.

The average reliability (Cronbach's alpha) across national samples was 0.79 for T_STDCOM (ranging from 0.72 to 0.86) (see Table 12.50). The scale was based on the item parameters shown in Table 12.51.

Figure 12.26: Confirmatory factor analysis of items measuring teachers' reports of student activities in the community

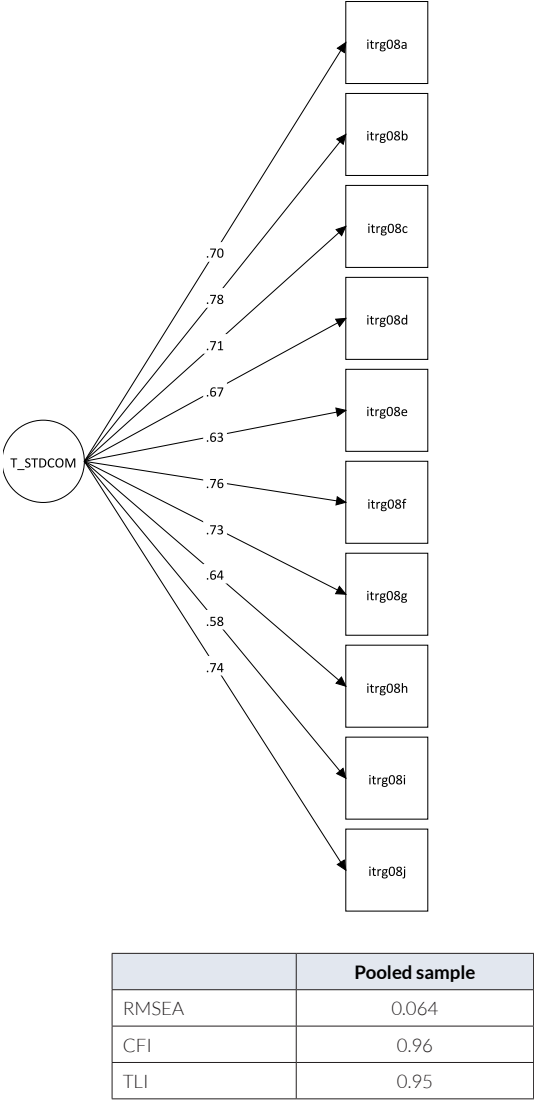


Table 12.50: Reliabilities for scale measuring teachers' reports of student activities in the community

Country	T_STDCOM
Brazil	0.82
Bulgaria	0.81
Chinese Taipei	0.82
Colombia	0.82
Croatia	0.82
Cyprus	0.77
Denmark	0.76
Estonia	0.80
France	0.74
Italy	0.76
Latvia	0.78
Lithuania	0.81
Malta	0.86
Netherlands	0.72
North Rhine-Westphalia*	0.78
Norway	0.74
Poland	0.80
Romania	0.83
Serbia	0.78
Slovak Republic	0.77
Slovenia	0.81
Spain	0.81
Sweden	0.77
Average	0.79
Minimum value	0.72
Maximum value	0.86

Table 12.51: Item parameters for scale measuring teachers' reports of student activities in the community

Scale or item	Question/item wording	Delta
T_STDCOM	Teachers reports of student activities in the community	
IT4G08A	Activities related to environmental sustainability (e.g. [energy and water saving, recycling])	-0.85
IT4G08B	Activities related to human rights	-0.29
IT4G08C	Activities for underprivileged people or groups	0.05
IT4G08D	Cultural activities (e.g. theatre, music)	-0.80
IT4G08E	Multicultural and intercultural activities within the [local community] (e.g. [promotion and celebration of cultural diversity, food street market])	0.70
IT4G08F	Activities to raise people's awareness of social issues, such as [poverty, gender equality, domestic violence against women, sexual violence against women, violence against children]	-0.38
IT4G08G	Activities aimed at protecting the cultural and historical heritage in the [local community]	0.19
IT4G08H	Visits to political institutions (e.g. [Parliament house, Prime Minister's/President's official residence])	2.68
IT4G08I	Sports events	-0.84
IT4G08J	Activities to raise people's awareness of global issues (e.g. climate change, world poverty, international conflicts, child labour)	-0.48

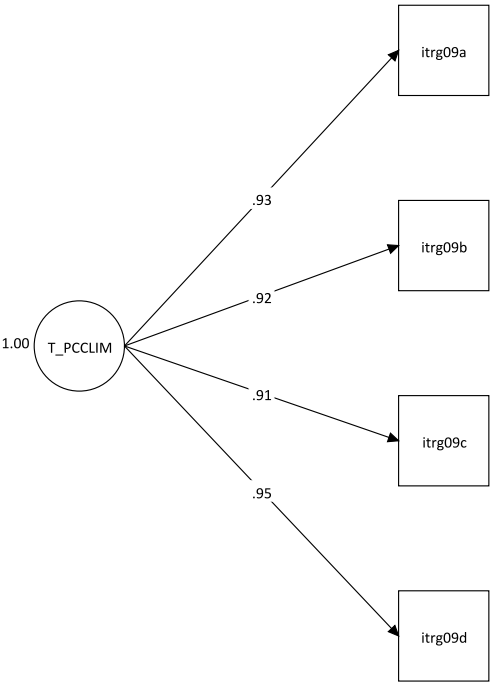
Teachers' perceptions of classroom climate

Question 9 of the teacher questionnaire required teachers to indicate their perception of the classroom climate that exists between students ("all or nearly all", "most of them", "some of them", or "none or hardly any"). One scale was derived from this item set. The four items were used to derive a scale reflecting *teachers' perceptions of classroom climate* (T_PCCLIM). Higher scale scores reflect a higher classroom climate.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teachers' perceptions of classroom climate (Figure 12.27). The RMSEA index indicated poor model fit, while both CFI and TLI suggested a close fit.

The average reliability (Cronbach's alpha) across national samples was 0.88 for T_PCCLIM (ranging from 0.84 to 0.92) (see Table 12.52). The scale was based on the item parameters shown in Table 12.53.

Figure 12.27: Confirmatory factor analysis of items measuring teachers' perceptions of classroom climate



	Pooled sample
RMSEA	0.119
CFI	1.00
TLI	0.99

Table 12.52: Reliabilities for scale measuring teachers' perceptions of classroom climate

Country	T_PCCLIM
Brazil	0.91
Bulgaria	0.87
Chinese Taipei	0.92
Colombia	0.86
Croatia	0.88
Cyprus	0.87
Denmark	0.90
Estonia	0.89
France	0.87
Italy	0.89
Latvia	0.87
Lithuania	0.87
Malta	0.91
Netherlands	0.85
North Rhine-Westphalia*	0.84
Norway	0.88
Poland	0.90
Romania	0.90
Serbia	0.90
Slovak Republic	0.88
Slovenia	0.88
Spain	0.91
Sweden	0.88
Average	0.88
Minimum value	0.84
Maximum value	0.92

Table 12.53: Item parameters for scale measuring teachers' perceptions of classroom climate

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)
T_PCCLIM	Teachers' perceptions of classroom climate			
IT4G09A	get on well with their classmates?	-0.52	-5.06	5.06
IT4G09B	are well intergrated in the class?	-0.46	-4.85	4.85
IT4G09C	respect their classmates even if they have different opinions?	1.07	-4.19	4.19
IT4G09D	have a good relationship with other students?	-0.08	-4.90	4.90

Teachers' reports on student involvement in activities

Question 10 of the teacher questionnaire required teachers to indicate the level of involvement ("to a large extent", "to a moderate extent", "to a small extent" or "not at all") that their students have in a list of activities. One scale was derived from this item set. The six items were used to derive a scale reflecting *teachers' reports on student involvement in activities* (T_STDINV). Higher scale scores reflect a higher level of student involvement.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teachers' reports on student involvement in activities (Figure 12.28). The model fit for the pooled data was satisfactory.

The average reliability (Cronbach's alpha) across national samples was 0.80 for T_STDINV (ranging from 0.64 to 0.87) (see Table 12.54). The scale was based on the item parameters shown in Table 12.55.

Figure 12.28: Confirmatory factor analysis of items measuring teachers' reports on student involvement in activities

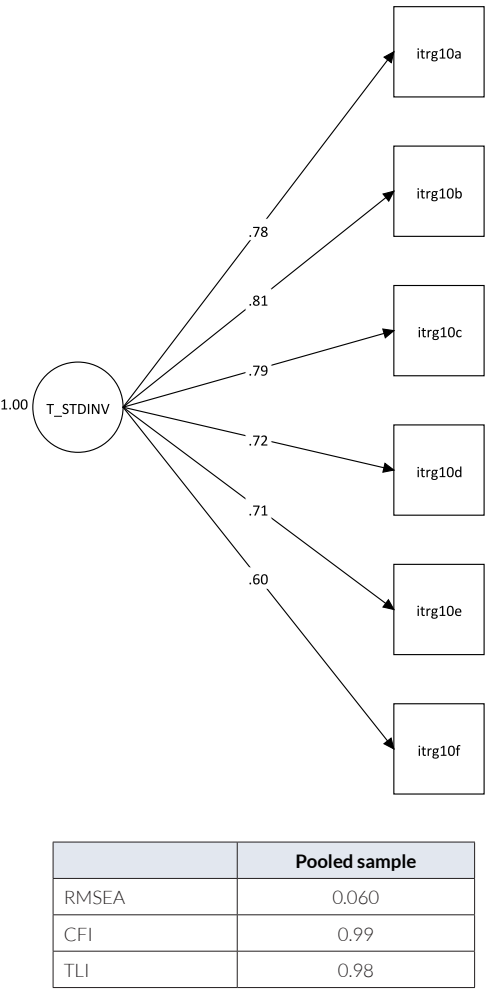


Table 12.54: Reliabilities for scale measuring teachers' reports on student involvement in activities

Country	T_STDINV
Brazil	0.87
Bulgaria	0.82
Chinese Taipei	0.87
Colombia	0.80
Croatia	0.82
Cyprus	0.86
Denmark	0.64
Estonia	0.81
France	0.82
Italy	0.81
Latvia	0.83
Lithuania	0.83
Malta	0.87
Netherlands	0.73
North Rhine-Westphalia*	0.76
Norway	0.67
Poland	0.85
Romania	0.84
Serbia	0.82
Slovak Republic	0.83
Slovenia	0.81
Spain	0.78
Sweden	0.65
Average	0.80
Minimum value	0.64
Maximum value	0.87

Table 12.55: Item parameters for scale measuring teachers' reports on student involvement in activities

Scale or item	Question/Item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_STDINV	Teachers' reports on student involvement in activities				
IT4G10A	Taking part in decisions related to teaching content	-0.15	-2.16	-0.19	2.34
IT4G10B	Contributing to classroom activities planning	-0.42	-2.37	-0.11	2.48
IT4G10C	Participating in establishing assessment criteria	0.43	-1.85	-0.13	1.97
IT4G10D	Participating in school self-evaluation processes	0.19	-1.95	-0.18	2.13
IT4G10E	Contributing to the choice of the teaching materials (e.g. textbooks, educational software, multimedia resources ...)	1.02	-1.73	-0.04	1.77
IT4G10F	Taking part in establishing classroom rules	-1.07	-1.40	-0.18	1.58

Teachers' reports on activities dealing with diversity

Question 11 of the teacher questionnaire required teachers to indicate the level ("to a large extent", "to a moderate extent", "to a small extent" or "not at all") that activities dealing with diversity are addressed. One scale was derived from this item set. The six items were used to derive a scale reflecting *teachers' reports on activities dealing with diversity* (T_DIVACT). Higher scale scores reflect a higher level of activities dealing with diversity.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess teachers' reports on activities dealing with diversity (Figure 12.29). The model fit for the pooled sample was acceptable.

The average reliability (Cronbach's alpha) across national samples was 0.86 for T_DIVACT (ranging from 0.82 to 0.89) (see Table 12.56). The scale was based on the item parameters shown in Table 12.57.

Figure 12.29: Confirmatory factor analysis of items measuring teachers' reports on activities dealing with diversity

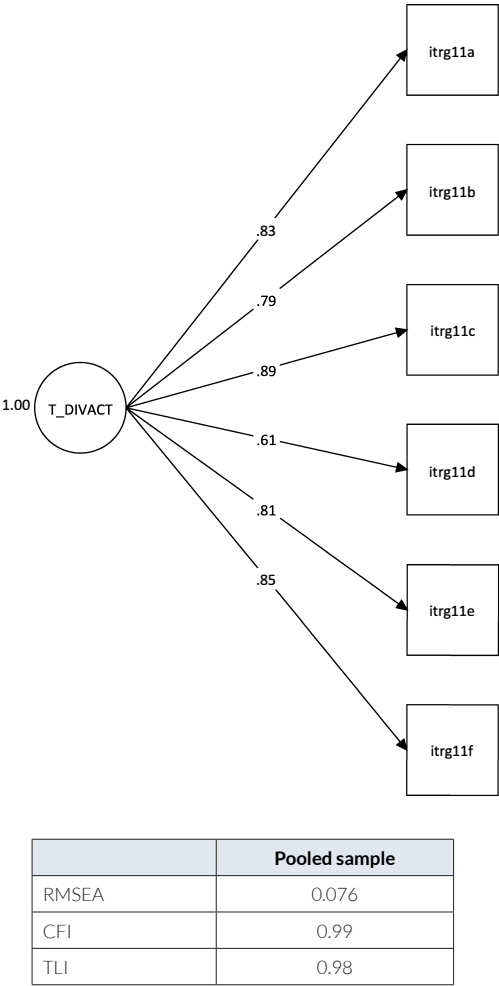


Table 12.56: Reliabilities for scale measuring teachers' reports on activities dealing with diversity

Country	T_DIVACT
Brazil	0.84
Bulgaria	0.83
Chinese Taipei	0.88
Colombia	0.85
Croatia	0.87
Cyprus	0.88
Denmark	0.83
Estonia	0.87
France	0.84
Italy	0.83
Latvia	0.87
Lithuania	0.87
Malta	0.89
Netherlands	0.82
North Rhine-Westphalia*	0.87
Norway	0.83
Poland	0.87
Romania	0.85
Serbia	0.83
Slovak Republic	0.87
Slovenia	0.89
Spain	0.88
Sweden	0.83
Average	0.86
Minimum value	0.82
Maximum value	0.89

Table 12.57: Item parameters for scale measuring teachers' reports on activities dealing with diversity

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_DIVACT	Teachers' reports on activities dealing with diversity				
IT4G11A	I discuss cultural differences with students.	0.11	-2.25	-0.18	2.43
IT4G11B	I encourage students to understand different points of view in class discussions.	-1.52	-1.88	-0.44	2.33
IT4G11C	I ask students to explore different cultural perspectives.	0.43	-2.01	-0.22	2.23
IT4G11D	I encourage students from different backgrounds to work together (e.g. in group works, peer learning activities.).	-0.69	-2.03	-0.22	2.25
IT4G11E	I involve students in discussions on gender issues (e.g. gender equity, gender stereotypes and gender diversity).	0.72	-1.81	-0.13	1.94
IT4G11F	I ask students to explore different social and economic perspectives.	0.95	-2.15	-0.15	2.30

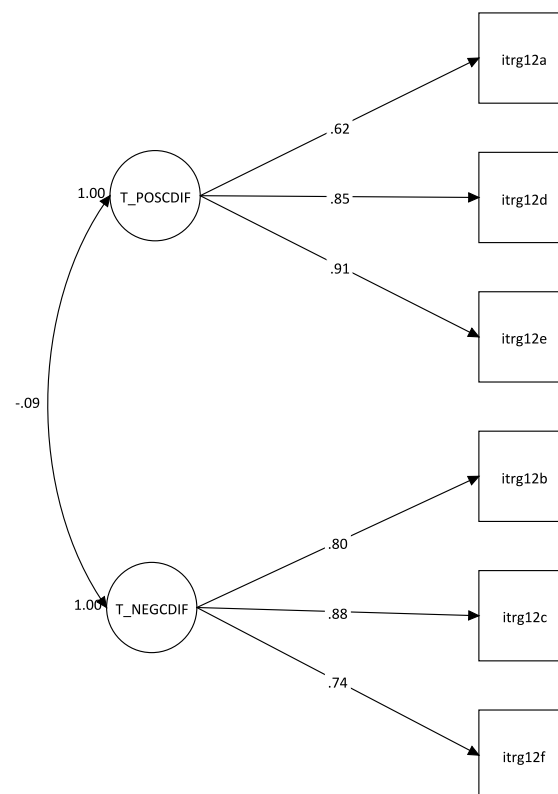
Teachers' opinions about the influence of cultural and ethnic differences on teaching activities

Question 12 of the teacher questionnaire required teachers to indicate their opinion ("strongly agree", "agree", "disagree" or "strongly disagree") on the influence of cultural and ethnic differences among students with regards to teaching and learning. Two scales were derived from this item set. Three positively worded items were used to derive a scale reflecting *Teachers' positive opinions about the influence of cultural and ethnic differences on teaching activities* (T_POSCDIF), while three negatively worded items were used to derive a scale of *Teachers' negative opinions about the influence of cultural and ethnic differences on teaching activities* (T_NEGCDIF). Higher scale scores reflect a stronger opinion (either positive or negative) with regards to the influence of cultural and ethnic differences on teaching activities

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess teachers' opinions on influence of cultural and ethnic differences on teaching (Figure 12.30). The model fit for the pooled data was acceptable and the two latent factors were slightly negatively correlated ($r = -0.09$).

The average reliabilities (Cronbach's alpha) across national samples were 0.73 for T_POSCDIF (ranging from 0.61 to 0.80) and 0.77 for T_NEGCDIF (ranging from 0.67 to 0.84) (see Table 12.58). These scales were based on the item parameters shown in Table 12.59.

Figure 12.30: Confirmatory factor analysis of items measuring teachers' opinions on the influence of cultural and ethnic diversity on teaching activities



	Pooled sample
RMSEA	0.073
CFI	0.98
TLI	0.97

Table 12.58: Reliabilities for scales measuring teachers' opinions on the influence of cultural and ethnic diversity on teaching activities

Country	T_POSCDIF	T_NEGCDIF
Brazil	0.79	0.83
Bulgaria	0.73	0.84
Chinese Taipei	0.73	0.82
Colombia	0.75	0.81
Croatia	0.73	0.81
Cyprus	0.76	0.80
Denmark	0.77	0.70
Estonia	0.74	0.77
France	0.75	0.76
Italy	0.72	0.75
Latvia	0.64	0.72
Lithuania	0.69	0.83
Malta	0.79	0.77
Netherlands	0.63	0.67
North Rhine-Westphalia*	0.72	0.75
Norway	0.75	0.72
Poland	0.73	0.74
Romania	0.80	0.82
Serbia	0.73	0.78
Slovak Republic	0.61	0.75
Slovenia	0.65	0.77
Spain	0.80	0.74
Sweden	0.78	0.72
Average	0.73	0.77
Minimum value	0.61	0.67
Maximum value	0.80	0.84

Table 12.59: Partial credit model item parameters for scales measuring teachers' opinions on the influence of cultural and ethnic diversity on teaching activities

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_POSCDIF	Teachers' positive opinions about the influence of cultural and ethnic differences on teaching activities				
IT4G12A	Cultural and ethnic differences among students are an important resource for teaching.	-0.12	-2.66	-0.72	3.38
IT4G12D	Cultural and ethnic differences among students strengthen students' sense of empathy.	0.20	-3.36	-0.71	4.07
IT4G12E	Cultural and ethnic differences among students promote students' civic-mindedness.	-0.08	-3.20	-0.81	4.02
T_NEGCDIF	Teachers' negative opinions about the influence of cultural and ethnic differences on teaching activities				
IT4G12B	Cultural and ethnic differences among students make it difficult to deal with controversial issues during classes.	-0.61	-4.45	0.37	4.08
IT4G12C	Cultural and ethnic differences among students make teaching activities more difficult.	-0.01	-4.15	0.47	3.68
IT4G12F	Cultural and ethnic differences among students make it difficult to have a good classroom climate.	0.62	-3.89	0.72	3.16

Teachers' opinions about the influence of socioeconomic differences on teaching activities

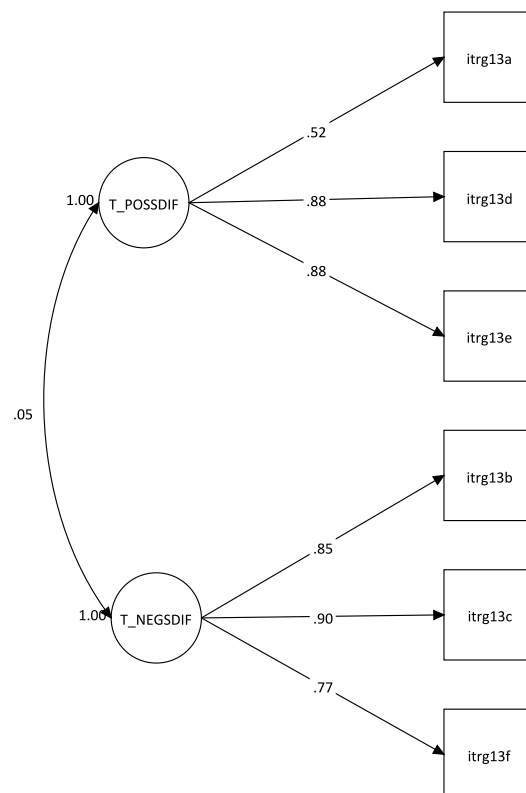
Question 13 of the teacher questionnaire required teachers to indicate their opinion ("Strongly agree", "Agree", "Disagree" or "Strongly disagree") on the influence of socioeconomic differences among students with regards to teaching and learning. Two scales were derived from this item set. Three positively worded items were used to derive a scale reflecting *Teachers' positive opinions about the influence of socioeconomic differences on teaching activities* (T_POSDIF), while three negatively worded items were used to derive a scale of *Teachers' negative opinions about the influence of socioeconomic differences on teaching activities* (T_NEGSDIF). Higher scale scores reflect a stronger opinion (either positive or negative) with regards to the influence of socioeconomic differences on teaching activities.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess teachers' opinions on the influence of socioeconomic differences on teaching (Figure 12.31). The RMSEA

index indicated poor model fit while CFI and TLI indicated only marginally acceptable fit. The two latent factors were only weakly correlated ($r = 0.05$).

The average reliabilities (Cronbach's alpha) across national samples were 0.72 for T_POSSDIF (ranging from 0.49 to 0.85) and 0.79 for T_NEGSDIF (ranging from 0.70 to 0.86) (see Table 12.60). Both scales were derived based on the item parameters shown in Table 12.61.

Figure 12.31: Confirmatory factor analysis of items measuring teachers' opinions on the influence of socioeconomic differences on teaching activities



	Pooled sample
RMSEA	0.141
CFI	0.95
TLI	0.91

Table 12.60: Reliabilities for scales measuring teachers' opinions on the influence of socioeconomic differences on teaching activities

Country	T_POSSDIF	T_NEGSDIF
Brazil	0.77	0.83
Bulgaria	0.68	0.86
Chinese Taipei	0.70	0.85
Colombia	0.72	0.84
Croatia	0.70	0.86
Cyprus	0.72	0.84
Denmark	0.85	0.70
Estonia	0.73	0.82
France	0.77	0.77
Italy	0.79	0.77
Latvia	0.63	0.76
Lithuania	0.68	0.83
Malta	0.79	0.79
Netherlands	0.55	0.70
North Rhine-Westphalia*	0.74	0.74
Norway	0.78	0.76
Poland	0.70	0.77
Romania	0.79	0.86
Serbia	0.64	0.81
Slovak Republic	0.49	0.79
Slovenia	0.64	0.80
Spain	0.77	0.77
Sweden	0.82	0.76
Average	0.72	0.79
Minimum value	0.49	0.70
Maximum value	0.85	0.86

Table 12.61: Partial credit model item parameters for scales measuring teachers' opinions on the influence of socioeconomic differences on teaching activities

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_POSSDIF	Teachers' positive opinions about the influence of socioeconomic differences on teaching activities				
IT4G13A	Social and economic differences among students are an important resource for teaching.	0.44	-3.08	-0.18	3.26
IT4G13D	Social and economic differences among students strengthen students' sense of empathy.	-0.13	-3.56	-0.37	3.92
IT4G13E	Social and economic differences among students promote students' civic-mindedness.	-0.31	-3.41	-0.43	3.84
T_NEGSDIF	Teachers' negative opinions about the influence of socioeconomic differences on teaching activities				
IT4G13B	Social and economic differences among students make it difficult to deal with controversial issues during classes.	-0.33	-5.35	0.43	4.92
IT4G13C	Social and economic differences among students make teaching activities more difficult.	-0.09	-4.99	0.46	4.52
IT4G13F	Social and economic differences among students make it difficult to have a good classroom climate.	0.42	-4.73	0.62	4.11

Teachers' reports on activities related to digital technologies

Question 15 of the teacher questionnaire required teachers to indicate whether they and their students have undertaken ("Never", "Once", "Once a month" or "More than once a month") a list of activities relating to digital technology. One scale was derived from this item set. The four items were used to derive a scale reflecting *teachers' reports on activities related to digital technologies* (T_ACTDIG). Higher scale scores reflect a higher level of activities relating to digital technologies.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teachers' reports on activities related to digital technologies (Figure 12.32). The model fit for the pooled data was good.

The average reliability (Cronbach's alpha) across national samples was 0.79 for T_ACTDIG (ranging from 0.72 to 0.85) (see Table 12.62). The scale was based on the item parameters shown in Table 12.63.

Figure 12.32: Confirmatory factor analysis of items measuring teachers' reports on activities related to digital technologies

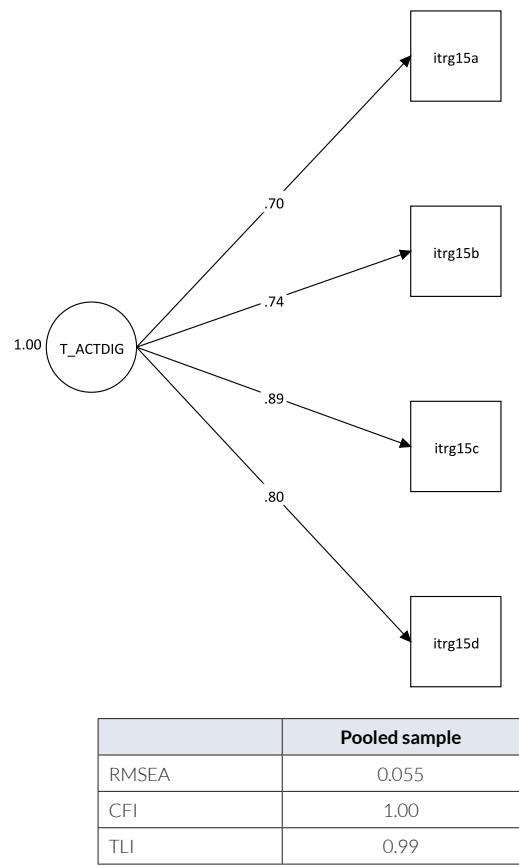


Table 12.62: Reliabilities for scale measuring teachers' reports on activities related to digital technologies

Country	T_ACTDIG
Brazil	0.83
Bulgaria	0.84
Chinese Taipei	0.77
Colombia	0.79
Croatia	0.82
Cyprus	0.82
Denmark	0.72
Estonia	0.80
France	0.76
Italy	0.76
Latvia	0.78
Lithuania	0.81
Malta	0.81
Netherlands	0.72
North Rhine-Westphalia*	0.79
Norway	0.78
Poland	0.80
Romania	0.85
Serbia	0.76
Slovak Republic	0.79
Slovenia	0.79
Spain	0.72
Sweden	0.77
Average	0.79
Minimum value	0.72
Maximum value	0.85

Table 12.63: Item parameters for scale measuring teachers' reports on activities related to digital technologies

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_ACTDIG	Teachers' reports on activities related to digital technologies				
IT4G15A	Activities on information and data literacy (e.g. searching and managing data, evaluating online content)	-0.57	-1.05	-0.02	1.07
IT4G15B	Activities aimed at encouraging students to contribute to online discussions on civic and social issues	1.12	-0.63	-0.15	0.79
IT4G15C	Activities aimed at improving students' skills to find information about civic and social issues on the internet	-0.06	-0.90	-0.03	0.93
IT4G15D	Activities aimed at developing students' awareness of issues related to online responsibility (e.g. safe and responsible internet use to avoid [cyber-bullying])	-0.50	-1.25	0.28	0.97

Teachers' perceptions of the importance of good citizenship

Question 17 of the teacher questionnaire required teachers to indicate their perception on the importance ("very important", "quite important", "not very important" or "not important at all") of certain behaviors with regards to their students becoming good adult citizens. Three scales were derived from this item set. Four items were used to derive a scale reflecting *Teachers' perception of the importance of conventional citizenship* (T_CITCON), four items were used to derive a scale of *Teachers' perception of the importance of social movement related citizenship* (T_CITSOC), while four items were used to derive a scale of *Teachers' perceptions of the importance of global citizenship* (T_GLOBCIT). Higher scale scores reflect a stronger opinion with regards to the importance of certain behaviors leading to students becoming good adult citizens.

We used confirmatory factor analyses assuming a three-dimensional model for these items to assess teachers' perceptions on the importance of good citizenship (Figure 12.33). The model fit for the pooled data was good and there were moderate correlations between T_CITCON and T_CITSOC ($r = 0.36$) as well as between T_CITCON and T_GLOBCIT ($r = 0.34$), while we observed a strong correlation ($r = 0.82$) between T_CITSOC and T_GLOBCIT.

The average reliabilities (Cronbach's alpha) across national samples were 0.61 for T_CITCON (ranging from 0.55 to 0.67), 0.74 for T_CITSOC (ranging from 0.63 to 0.83) and 0.74 for T_GLOBCIT (ranging from 0.66 to 0.81) (see Table 12.64). These scales were based on the item parameters shown in Table 12.65.

Figure 12.33: Confirmatory factor analysis of items measuring teachers' perceptions of the importance of good citizenship

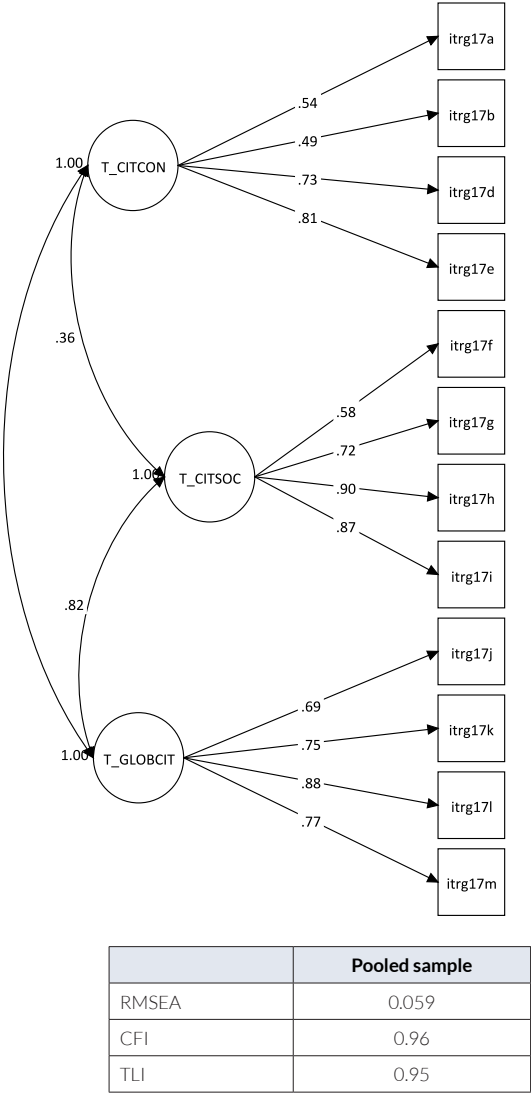


Table 12.64: Reliabilities for scales measuring teachers' perceptions of the importance of good citizenship

Country	T_CITCON	T_CITSOC	T_GLOBCIT
Brazil	0.65	0.73	0.75
Bulgaria	0.65	0.67	0.71
Chinese Taipei	0.60	0.71	0.75
Colombia	0.63	0.63	0.71
Croatia	0.62	0.78	0.78
Cyprus	0.62	0.75	0.78
Denmark	0.60	0.80	0.68
Estonia	0.59	0.71	0.74
France	0.57	0.74	0.73
Italy	0.60	0.74	0.73
Latvia	0.60	0.73	0.66
Lithuania	0.59	0.72	0.75
Malta	0.66	0.83	0.81
Netherlands	0.59	0.74	0.67
North Rhine-Westphalia*	0.58	0.81	0.74
Norway	0.58	0.78	0.73
Poland	0.55	0.73	0.74
Romania	0.67	0.68	0.73
Serbia	0.62	0.72	0.75
Slovak Republic	0.61	0.72	0.74
Slovenia	0.61	0.74	0.76
Spain	0.63	0.77	0.78
Sweden	0.62	0.83	0.75
Average	0.61	0.74	0.74
Minimum value	0.55	0.63	0.66
Maximum value	0.67	0.83	0.81

Table 12.65: Partial credit model item parameters for scales measuring teachers' perceptions of the importance of good citizenship

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_CITCON	<i>Teachers' perception of the importance of conventional citizenship</i>				
IT4G17A	Voting in every national election	-1.63	-1.17	-0.39	1.56
IT4G17B	Joining a political party	2.20	-2.41	0.93	1.48
IT4G17D	Following political issues in the newspaper, on the radio, on TV or on the Internet	-1.03	-1.94	-0.34	2.28
IT4G17E	Engaging in political discussions	0.47	-2.55	0.15	2.40
T_CITSOC	<i>Teachers' perception of the importance of social movement related citizenship</i>				
IT4G17F	Participating in peaceful protests against laws believed to be unjust	1.97	-2.78	-0.19	2.97
IT4G17G	Participating in activities to benefit people in the [local community]	-0.52	-2.30	-0.92	3.22
IT4G17H	Taking part in activities promoting human rights	-0.39	-2.63	-0.51	3.14
IT4G17I	Taking part in activities to protect the environment	-1.06	-2.35	-0.77	3.13
T_GLOBCIT	<i>Teachers' perceptions of the importance of global citizenship</i>				
IT4G17J	Showing interest in different cultures and languages	0.05	-3.14	-0.49	3.63
IT4G17K	Making changes to one's personal lifestyle in order to become more environmental friendly	-0.52	-2.27	-0.85	3.11
IT4G17L	Supporting initiatives that promote equal opportunities for all people across the world	-0.06	-2.54	-0.67	3.21
IT4G17M	Helping people in less developed countries	0.53	-3.08	-0.60	3.68

Teachers' reports on civic-related activities in class

Question 20 of the teacher questionnaire required teachers of subjects related to civic and citizenship education to indicate how often ("never", "sometimes", "often" or "very often") students take part in a list of civic-related activities in class. One scale was derived from this item set. Eight items were used to derive a scale reflecting *Teachers' reports on civic-related activities in class* (T_CIVCLAS). Higher scale scores reflect a

higher level of civic-related activities taking place.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teachers' reports on civic-related activities in class (Figure 12.34). The model fit for the pooled data was good for the model based on pooled data.

The average reliability (Cronbach's alpha) across national samples was 0.79 for T_CIVCLAS (ranging from 0.66 to 0.86) (see Table 12.66). These scales were based on the item parameters shown in Table 12.67

Figure 12.34: Confirmatory factor analysis of items measuring teachers' reports on civic-related activities in class

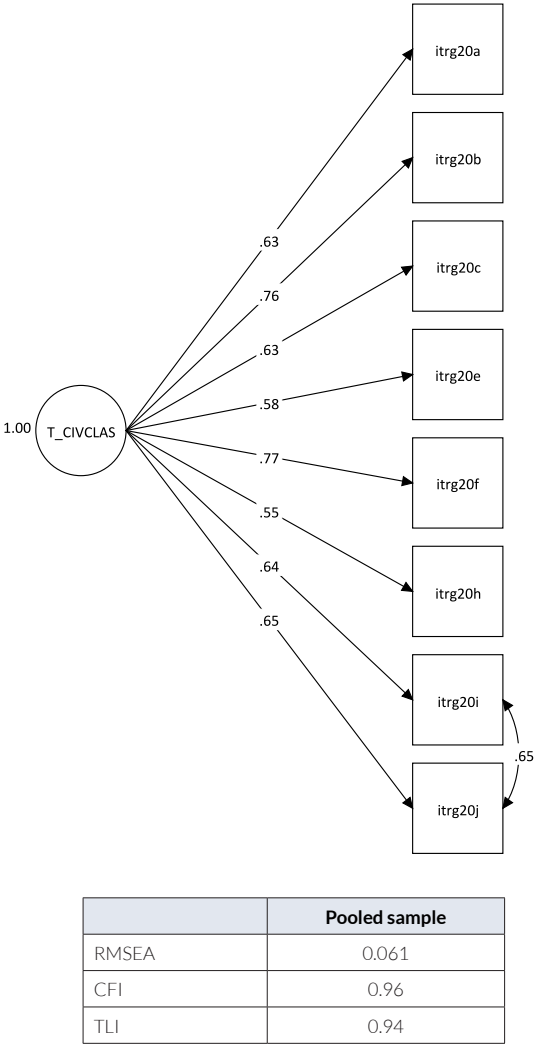


Table 12.66: Reliabilities for scale measuring teachers' reports on civic-related activities in class

Country	T_CIVCLAS
Brazil	0.85
Bulgaria	0.84
Chinese Taipei	0.80
Colombia	0.84
Croatia	0.86
Cyprus	0.80
Denmark	0.66
Estonia	0.77
France	0.81
Italy	0.81
Latvia	0.79
Lithuania	0.81
Malta	0.71
Netherlands	0.76
North Rhine-Westphalia*	0.80
Norway	0.70
Poland	0.81
Romania	0.85
Serbia	0.86
Slovak Republic	0.83
Slovenia	0.79
Spain	0.80
Sweden	0.67
Average	0.79
Minimum value	0.66
Maximum value	0.86

Table 12.67: Item parameters for scale measuring teachers' reports on civic-related activities in class

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_CIVCLAS	Teachers' reports on civic-related activities in class				
IT4G20A	Students work on projects that involve gathering information outside school (e.g. interviews in the neighbourhood, small scale surveys).	1.09	-2.04	0.90	1.14
IT4G20B	Students work in small groups on different topics/issues.	-0.26	-2.21	0.48	1.73
IT4G20C	Students participate in role plays.	0.61	-1.90	0.51	1.39
IT4G20E	Students discuss current issues.	-1.24	-2.12	0.32	1.80
IT4G20F	Students research and/or analyse information gathered from multiple Web sources (e.g. wikis, online newspapers).	-0.27	-2.12	0.39	1.73
IT4G20H	Students propose topics/issues for the following lessons.	0.80	-2.28	0.93	1.35
IT4G20I	Students make presentations using digital technologies (e.g. [PowerPoint/Prezi, videos, multimedia]).	-0.19	-1.85	0.48	1.38
IT4G20J	Students use digital technologies for project or class work.	-0.45	-1.93	0.53	1.40

Teachers' reports on activities related to global issues

Question 21 of the teacher questionnaire required teachers to indicate ("Yes" or "No") whether a list of activities relating to global issues was carried out with students. One scale was derived from this item set. Five items were used to derive a scale reflecting *Teachers' reports on activities related to global issues* (T_ACTGLOB). Higher scale scores reflect a higher level of activities relating to global issues being carried out with students.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess teachers' reports on activities relating to global issues. (Figure 12.35). The model fit for the pooled data was good.

The average reliability (Cronbach’s alpha) across national samples was 0.60 for T_ACTGLOB (ranging from 0.22 to 0.77) (see Table 12.68). These scales were based on the item parameters shown in Table 12.69

Figure 12.35: Confirmatory factor analysis of items measuring teachers’ reports on activities related to global issues

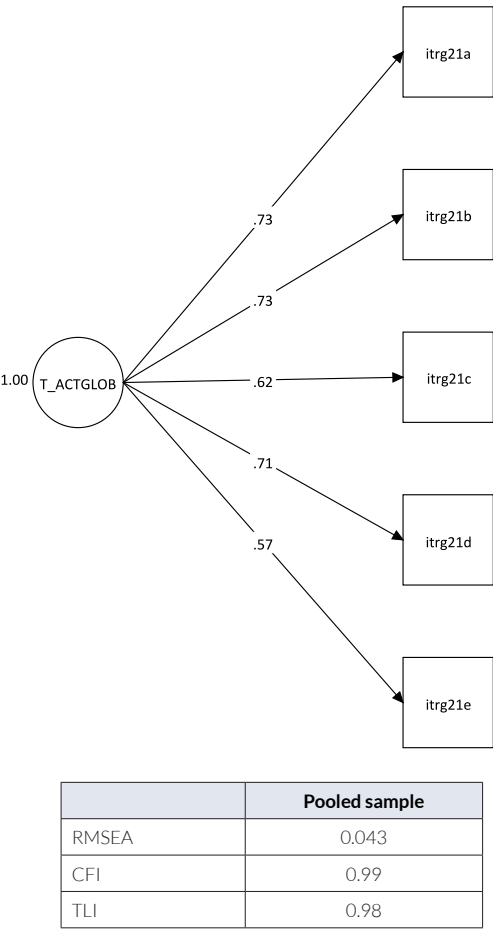


Table 12.68: Reliabilities for scale measuring teachers' reports on activities related to global issues

Country	T_ACTGLOB
Brazil	0.73
Bulgaria	0.65
Chinese Taipei	0.75
Colombia	0.77
Croatia	0.69
Cyprus	0.66
Denmark	0.22
Estonia	0.61
France	0.50
Italy	0.56
Latvia	0.50
Lithuania	0.61
Malta	0.69
Netherlands	0.56
North Rhine-Westphalia*	0.55
Norway	0.52
Poland	0.58
Romania	0.77
Serbia	0.51
Slovak Republic	0.59
Slovenia	0.73
Spain	0.63
Sweden	0.51
Average	0.60
Minimum value	0.22
Maximum value	0.77

Table 12.69: Item parameters for scale measuring teachers' reports on activities related to global issues

Scale or item	Question/item wording	Delta
T_ACTGLOB	Teachers' reports on activities related to global issues	
IT4G21A	Activities to raise students' awareness of important global issues (e.g. collecting, analysing and evaluating information in reports from NGOs or international organisations, in newspapers or online)	0.26
IT4G21B	Activities to raise students' awareness of the relations between local and global issues (e.g. activities on social and economic interconnections, on the global economy, on the roots of migration)	-0.35
IT4G21C	Activities to inform students about [Fair trade] (e.g. activities focused on where the food comes from and on the people involved in producing it)	0.80
IT4G21D	Activities to make students aware about collective engagement to achieve improvements worldwide (e.g. [climate change protests, environment clean-up movements, social justice movements])	-0.30
IT4G21E	Activities to analyse how diverse identities influence the ability to live together	-0.41

Teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics

Question 22 of the teacher questionnaire required teachers to indicate how prepared ("very well prepared," "quite well prepared," "not very well prepared," or "not prepared at all") they were to teach different topics and skills related to civic and citizenship education (CCE). Two scales were derived from this item set. Thirteen items were used to derive a scale reflecting *Teachers' preparedness for teaching CCE topics and skills* (T_PRPCCE) and thirteen items were used to derive a scale reflecting *Teachers' perceptions of students' opportunities to learn about civic topics and skills* (T_OPPLRN). Higher scale scores reflect a higher level of preparedness for teaching and students' opportunities to learn topics and skills related to civic and citizenship education.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess teachers'

reports on teaching topics related to civic and citizenship (Figure 12.36). The model fit for the pooled sample was good and the two latent factors were strongly correlated ($r = 0.50$).

The average reliabilities (Cronbach's alpha) across national samples were 0.91 for T_PRPCCE (ranging from 0.82 to 0.95) and 0.93 for (ranging from 0.88 to 0.95) for T_OPPLRN (see Table 12.70). These scales were based on the item parameters shown in Table 12.71.

Figure 12.36: Confirmatory factor analysis of items measuring teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics

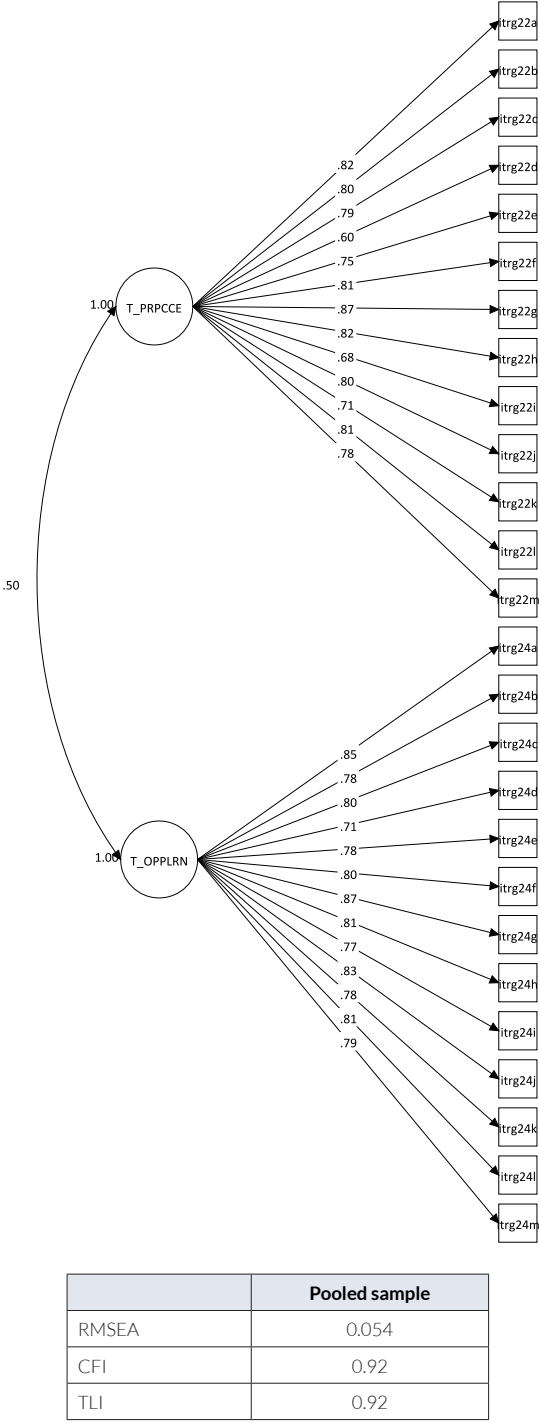


Table 12.70: Reliabilities for scales measuring teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics

Country	T_PRPCCE	T_OPPLRN
Brazil	0.94	0.95
Bulgaria	0.95	0.95
Chinese Taipei	0.90	0.89
Colombia	0.92	0.95
Croatia	0.93	0.94
Cyprus	0.92	0.93
Denmark	0.85	0.88
Estonia	0.90	0.94
France	0.92	0.92
Italy	0.90	0.91
Latvia	0.88	0.94
Lithuania	0.93	0.94
Malta	0.90	0.94
Netherlands	0.82	0.91
North Rhine-Westphalia*	0.88	0.91
Norway	0.89	0.92
Poland	0.90	0.94
Romania	0.94	0.95
Serbia	0.92	0.91
Slovak Republic	0.91	0.92
Slovenia	0.92	0.93
Spain	0.91	0.93
Sweden	0.89	0.88
Average	0.91	0.93
Minimum value	0.82	0.88
Maximum value	0.95	0.95

Table 12.71: Partial credit model item parameters for scales measuring teachers' reports on their preparedness for teaching and student's opportunities to learn CCE topics

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
T_PRPCCE	<i>Teachers' preparedness for teaching CCE topics and skills</i>				
IT4G22A	Human rights	-0.35	-3.06	-0.52	3.58
IT4G22B	Voting and elections	0.38	-2.58	-0.15	2.73
IT4G22C	The global community and international organisations	0.80	-3.20	0.06	3.14
IT4G22D	The environment and environmental sustainability	-0.03	-3.20	-0.26	3.46
IT4G22E	Emigration and immigration	0.58	-3.08	0.08	3.00
IT4G22F	Equal opportunities for men and women	-0.30	-3.08	-0.26	3.34
IT4G22G	Citizens' rights and responsibilities	-0.55	-2.89	-0.36	3.25
IT4G22H	The constitution and political systems	0.83	-2.39	-0.05	2.44
IT4G22I	Responsible Internet use (e.g. privacy, source reliability, social media)	-0.31	-3.06	-0.20	3.26
IT4G22J	Critical and independent thinking	-0.59	-2.84	-0.47	3.31
IT4G22K	Conflict resolution	-0.46	-3.48	-0.02	3.50
IT4G22L	Global issues (such as world poverty, international conflicts, child labour, social justice)	-0.09	-3.02	-0.21	3.23
IT4G22M	Diversity and inclusiveness	0.07	-3.16	-0.12	3.28
T_OPPLRN	<i>Teachers' perceptions of students' opportunities to learn about civic topics and skills</i>				
IT4G24A	Human rights	-0.88	-3.42	0.15	3.27
IT4G24B	Voting and elections	0.75	-2.78	0.18	2.60
IT4G24C	The global community and international organisations	0.95	-3.07	-0.06	3.12
IT4G24D	The environment and environmental sustainability	-0.47	-2.67	-0.14	2.81
IT4G24E	Emigration and immigration	0.94	-2.80	-0.12	2.92
IT4G24F	Equal opportunities for men and women	0.07	-2.88	-0.07	2.95
IT4G24G	Citizens' rights and responsibilities	-0.57	-2.79	-0.10	2.89

IT4G24H	The constitution and political systems	0.65	-2.54	-0.10	2.64
IT4G24I	Responsible Internet use (e.g. privacy, source reliability, social media)	-0.81	-2.35	-0.31	2.66
IT4G24J	Critical and independent thinking	-0.46	-3.10	-0.04	3.14
IT4G24K	Conflict resolution	-0.46	-3.30	0.10	3.20
IT4G24L	Global issues (such as world poverty, international conflicts, child labour, social justice)	0.22	-3.02	-0.10	3.12
IT4G24M	Diversity and inclusiveness	0.06	-3.15	0.03	3.12

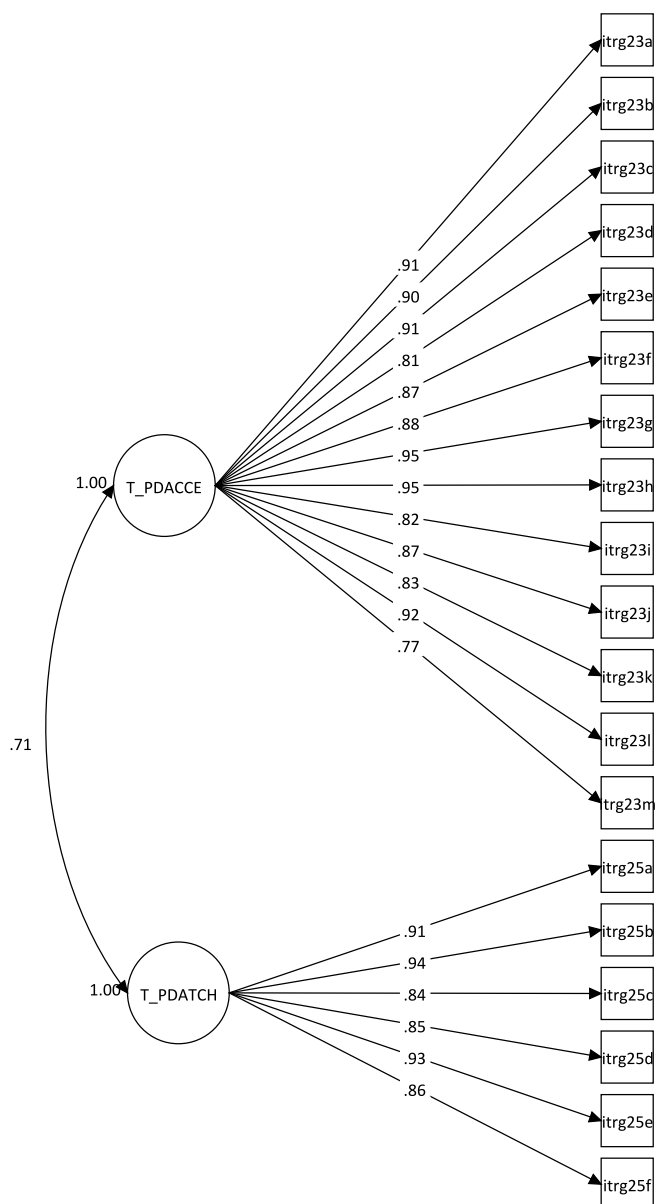
Teachers' reports on their PD activities

Question 23 of the teacher questionnaire required teachers to indicate whether they attended ("yes, during pre-service training," "yes, during in-service training," "yes, during both pre- and in-service training," or "no") any teacher training courses addressing a list of topics and skills. Two scales were derived from this item set. Thirteen items were used to derive a scale reflecting *teachers' reports on their professional development (PD) activities* (T_PDACE) and thirteen items were used to derive a scale reflecting *teachers' PD activities for teaching methods* (T_PDATCH). Higher scale scores reflect a higher level of attendance for the teacher training courses.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess teachers' reports on their PD activities (Figure 12.37). The model fit for the pooled sample was good and the two latent factors were strongly correlated ($r = 0.71$).

The average reliabilities (Cronbach's alpha) across national samples were 0.92 for T_PDACE (ranging from 0.87 to 0.96) and 0.83 for T_PDATCH (ranging from 0.70 to 0.92) (see Table 12.72). These scales were based on the item parameters shown in Table 12.73.

Figure 12.37: Confirmatory factor analysis of items measuring teachers' reports on their PD activities



	Pooled sample
RMSEA	0.059
CFI	0.99
TLI	0.98

Table 12.72: Reliabilities for scales measuring teachers' reports on their PD activities

Country	T_PDACE	T_PDATCH
Brazil	0.96	0.87
Bulgaria	0.93	0.86
Chinese Taipei	0.91	0.81
Colombia	0.95	0.92
Croatia	0.94	0.92
Cyprus	0.93	0.88
Denmark	0.90	0.80
Estonia	0.92	0.81
France	0.91	0.82
Italy	0.90	0.84
Latvia	0.92	0.83
Lithuania	0.90	0.84
Malta	0.90	0.85
Netherlands	0.88	0.70
North Rhine-Westphalia*	0.91	0.84
Norway	0.95	0.87
Poland	0.93	0.77
Romania	0.96	0.88
Serbia	0.87	0.81
Slovak Republic	0.93	0.87
Slovenia	0.93	0.85
Spain	0.91	0.81
Sweden	0.87	0.71
Average	0.92	0.83
Minimum value	0.87	0.70
Maximum value	0.96	0.92

Table 12.73: Partial credit model item parameters for scales measuring teachers' reports on their PD activities

Scale or item	Question/item wording	Delta
T_PDACE	Teachers' PD activities for CCE topics and skills	
IT4G23A	Human rights	-0.26
IT4G23B	Voting and elections	1.23
IT4G23C	The global community and international organisations	0.88
IT4G23D	The environment and environmental sustainability	-0.06
IT4G23E	Emigration and immigration	1.12
IT4G23F	Equal opportunities for men and women	0.51
IT4G23G	Citizens' rights and responsibilities	-0.24
IT4G23H	The constitution and political systems	0.60
IT4G23I	Responsible Internet use (e.g. privacy, source reliability, social media)	-1.39
IT4G23J	Critical and independent thinking	-0.58
IT4G23K	Conflict resolution	-1.33
IT4G23L	Global issues (such as world poverty, international conflicts, child labour, social justice)	0.23
IT4G23M	Diversity and inclusiveness	-0.71
T_PDATCH	Teachers' PD activities for teaching methods	
IT4G25A	Pair and group work	-0.83
IT4G25B	Classroom discussion	-0.08
IT4G25C	Role play	0.40
IT4G25D	Research work	0.90
IT4G25E	Problem solving	-0.28
IT4G25F	Project work	-0.11

School questionnaire

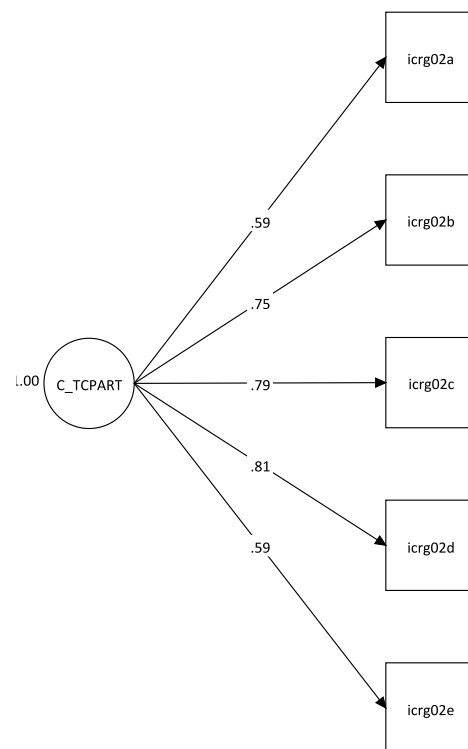
Schools' principals' perceptions of teacher participation at school

Question 2 of the school questionnaire required principals to give their perception as to the level ("all or nearly all," "most of them," "some of them," or "none or hardly any") of teacher participation in school governance. One scale was derived from this item set. Five items were used to derive a scale reflecting *school principals' perceptions of teacher participation at school* (C_TCPART). Higher scale scores reflect a higher level of teacher participation at school.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess schools' principals' perceptions of teacher participation at school (Figure 12.38). The model fit for the pooled sample was satisfactory.

The average reliabilities (Cronbach's alpha) across national samples was 0.72 for C_TCPART (ranging from 0.49 to 0.85) (see Table 12.74). These scales were based on the item parameters shown in Table 12.75.

Figure 12.38: Confirmatory factor analysis of items measuring school principals' perceptions of teacher participation at school



	Pooled sample
RMSEA	0.041
CFI	0.99
TLI	0.99

Table 12.74: Reliabilities for scale measuring school principals' perceptions of teacher participation at school

Country	C_TCPART
Brazil	0.85
Bulgaria	-
Chinese Taipei	0.78
Colombia	0.80
Croatia	0.77
Cyprus	0.53
Denmark	0.61
Estonia	0.71
France	0.75
Italy	0.77
Latvia	0.75
Lithuania	0.85
Malta	0.75
Netherlands	0.49
North Rhine-Westphalia*	0.65
Norway	0.70
Poland	0.66
Romania	0.85
Schleswig-Holstein*	0.64
Serbia	0.75
Slovak Republic	0.66
Slovenia	0.73
Spain	0.82
Sweden	0.75
Average	0.72
Minimum value	0.49
Maximum value	0.85

Table 12.75: Item parameters for scale measuring school principals' perceptions of teacher participation at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
C_TCPART	In your opinion, how many teachers participate as follows at this school?				
IC4G02A	Making useful suggestions for improving school governance	0.78	-4.06	0.63	3.43
IC4G02B	Supporting good discipline throughout the school	-1.41	-3.58	-0.03	3.61
IC4G02C	Actively taking part in school [development/improvement activities]	-0.83	-4.86	0.72	4.15
IC4G02D	Encouraging students' active participation in school life	-0.54	-3.42	-0.10	3.53
IC4G02E	Being willing to be members of the [school council, school governing board] as teacher representatives	2.01	-3.78	0.96	2.82

Schools' principals' perceptions of student and parent involvement at school

Question 3 of the school questionnaire required principals to give their perception as to the level ("to a large extent," "to a moderate extent," "to a small extent," or "not at all") of student and parent involvement at school with regards to a list of statements. Two scales were derived from this item set. Four items were used to derive a scale reflecting *perceptions of student involvement at school* (C_STDINV), and five items were used to derive a scale reflecting *perceptions of parent involvement at school* (C_PARINV). Higher scale scores reflect a higher level of involvement at school.

We used confirmatory factor analyses assuming a two-dimensional model for these items to assess schools' principals' perceptions of student and parent involvement at school (Figure 12.39). The model fit for the

pooled sample was acceptable and the two latent factors were strongly correlated ($r = 0.78$).

The average reliabilities (Cronbach's alpha) across national samples were 0.75 for C_STDINV (ranging from 0.60 to 0.85) and 0.77 for C_PARINV (ranging from 0.57 to 0.88) (see Table 12.76). These scales were based on the item parameters shown in Table 12.77.

Figure 12.39: Confirmatory factor analysis of items school principals' perceptions of student and parent involvement at school

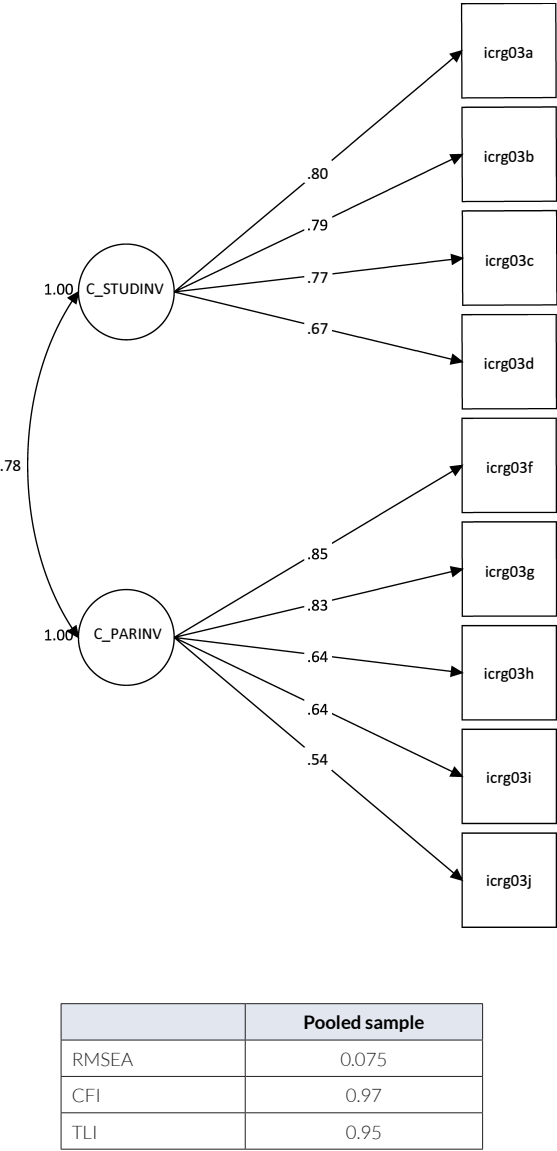


Table 12.76: Reliabilities for scales measuring school principals' perceptions of student and parent involvement at school

Country	C_STDINV	C_PARINV
Brazil	0.79	0.78
Bulgaria	0.79	0.86
Chinese Taipei	0.78	0.87
Colombia	0.78	0.75
Croatia	0.83	0.75
Cyprus	0.85	0.88
Denmark	0.61	0.75
Estonia	0.79	0.74
France	0.78	0.75
Italy	0.76	0.76
Latvia	0.72	0.85
Lithuania	0.81	0.77
Malta	0.78	0.58
Netherlands	0.78	0.72
North Rhine-Westphalia*	0.68	0.77
Norway	0.66	0.68
Poland	0.69	0.80
Romania	0.75	0.84
Schleswig-Holstein*	0.60	0.75
Serbia	0.79	0.78
Slovak Republic	0.76	0.85
Slovenia	0.77	0.80
Spain	0.78	0.83
Sweden	0.63	0.57
Average	0.75	0.77
Minimum value	0.60	0.57
Maximum value	0.85	0.88

Table 12.77: Partial credit model item parameters for scales measuring school principals' perceptions of student and parent involvement at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
C_STDINV	In your opinion, to what extent do the following statements apply to this school, in the current school year?				
IC4G03A	Students are involved in designing [school educational plan].	0.78	-3.02	0.03	2.99
IC4G03B	Students are involved in the definition of school rules and regulations.	-0.47	-2.71	-0.08	2.79
IC4G03C	Students are encouraged to contribute to decisions related to teaching content	0.46	-2.75	-0.19	2.94
IC4G03D	Students are encouraged to contribute to classroom activities planning.	-0.77	-2.72	-0.03	2.75
C_PARINV	In your opinion, to what extent do the following statements apply to this school, in the current school year?				
IC4G03F	[Parents or guardians] are involved in designing [school educational plan].	0.50	-2.30	-0.12	2.42
IC4G03G	[Parents or guardians] are involved in the development of the school rules and regulations.	0.03	-2.03	-0.13	2.16
IC4G03H	[Parents or guardians] are involved in school self-evaluation processes.	-0.32	-2.14	-0.12	2.26
IC4G03I	[Parents or guardians] are involved in decisions related to the school annual budget and financial planning.	0.97	-1.62	-0.22	1.83
IC4G03J	[Parents or guardians] are invited to participate in school projects and initiatives.	-1.19	-2.44	0.22	2.22

Schools' principals' perceptions of school activities in the community

Question 8 of the school questionnaire required principals to give their perception as to how many students took part ("all or nearly all," "most of them," "some of them," "none or hardly any," or "Not offered at school") in

a list of activities in the community. One scale was derived from this item set. Ten items were used to derive a scale reflecting *school principals perceptions of student activities in the community* (C_STDCOM). Higher scale scores reflect a higher level of student participation in activities.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess schools' principals' perceptions of school activities in the community (Figure 12.40). The model fit for the pooled sample was acceptable.

The average reliability (Cronbach's alpha) across national samples was 0.82 for C_STDCOM (ranging from 0.65 to 0.92) (see Table 12.78). The scale was based on the item parameters shown in Table 12.79.

Figure 12.40: Confirmatory factor analysis of items school principals' perceptions of school activities in the community

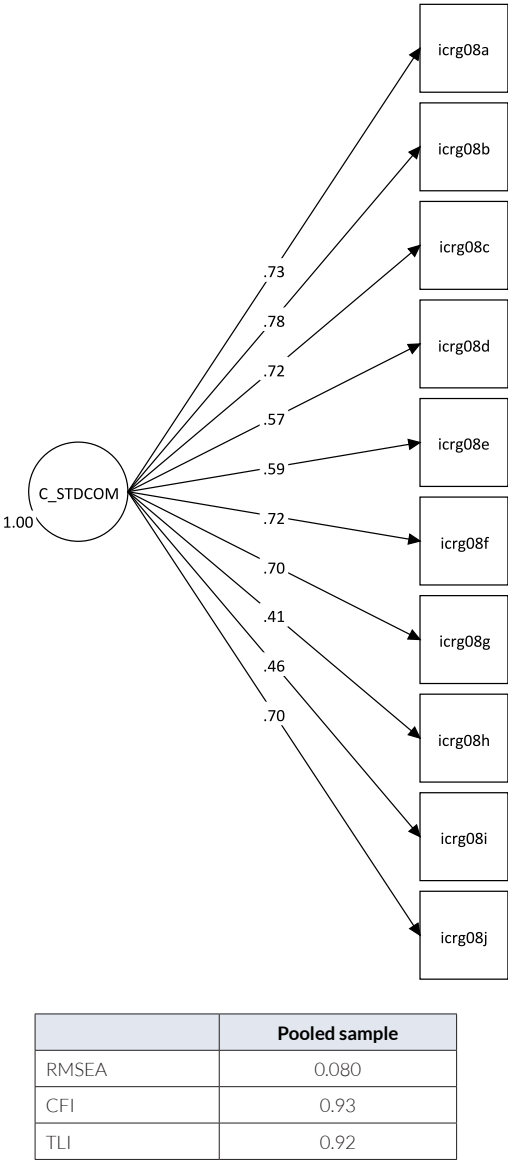


Table 12.78: Reliabilities for scales measuring school principals' perceptions of school activities in the community

Country	C_STDCOM
Brazil	0.92
Bulgaria	0.83
Chinese Taipei	0.87
Colombia	0.83
Croatia	0.85
Cyprus	0.89
Denmark	0.80
Estonia	0.83
France	0.82
Italy	0.84
Latvia	0.73
Lithuania	0.87
Malta	0.87
Netherlands	0.71
North Rhine-Westphalia*	0.81
Norway	0.81
Poland	0.77
Romania	0.88
Schleswig-Holstein*	0.65
Serbia	0.86
Slovak Republic	0.89
Slovenia	0.81
Spain	0.83
Sweden	0.71
Average	0.82
Minimum value	0.65
Maximum value	0.92

Table 12.79: Partial credit model item parameters for scale measuring school principals' perceptions of school activities in the community

Scale or item	Question/item wording	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?				
IC4G08A	Activities related to environmental sustainability (e.g. [energy and water saving, recycling])	-0.58	-0.92	0.14	0.78
IC4G08B	Activities related to human rights	-0.15	-0.90	0.09	0.81
IC4G08C	Activities for underprivileged people or groups	0.17	-0.89	0.35	0.55
IC4G08D	Cultural activities (e.g. theatre, music)	-0.78	-1.00	0.12	0.88
IC4G08E	Multicultural and intercultural activities within the [local community] (e.g. [promotion and celebration of cultural diversity, food street market])	0.56	-0.60	0.14	0.45
IC4G08F	Activities to raise people's awareness of social issues, such as [poverty, gender equality, domestic violence against women, sexual violence against women, violence against children]	-0.15	-0.68	0.00	0.68
IC4G08G	Activities aimed at protecting the cultural and historical heritage within the [local community]	0.34	-0.67	-0.11	0.77
IC4G08H	Visits to political institutions (e.g. [Parliament house, Prime Minister's/President's official residence])	1.40	-0.35	0.33	0.02

IC4G08I	Sports events	-0.74	-0.70	0.07	0.63
IC4G08J	Activities to raise people's awareness of global issues (e.g. [climate change, world poverty, international conflicts, child labour])	-0.07	-0.74	0.05	0.69

Schools' principals' reports on environmentally friendly practices at school

Question 10 of the school questionnaire required principals to give their perception as to the extent ("to a large extent," "to a moderate extent," "to a small extent," "not at all," or "Not applicable") that environment-friendly practices are followed at school. One scale was derived from this item set. Nine items were used to derive a scale reflecting *school principals' perceptions of environment-friendly practices being followed* (C_ENPRAC). Higher scale scores reflect a higher level of environmentally friendly practices being followed at school.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess schools' principals' reports on environment friendly practices at school (Figure 12.41). The model fit for the pooled sample was acceptable.

The average reliability (Cronbach's alpha) across national samples were 0.73 for C_ENPRAC (ranging from 0.57 to 0.85) (see Table 12.80). These scales were based on the item parameters shown in Table 12.81.

Figure 12.41: Confirmatory factor analysis of items measuring school principals' reports on environment-friendly practices at school

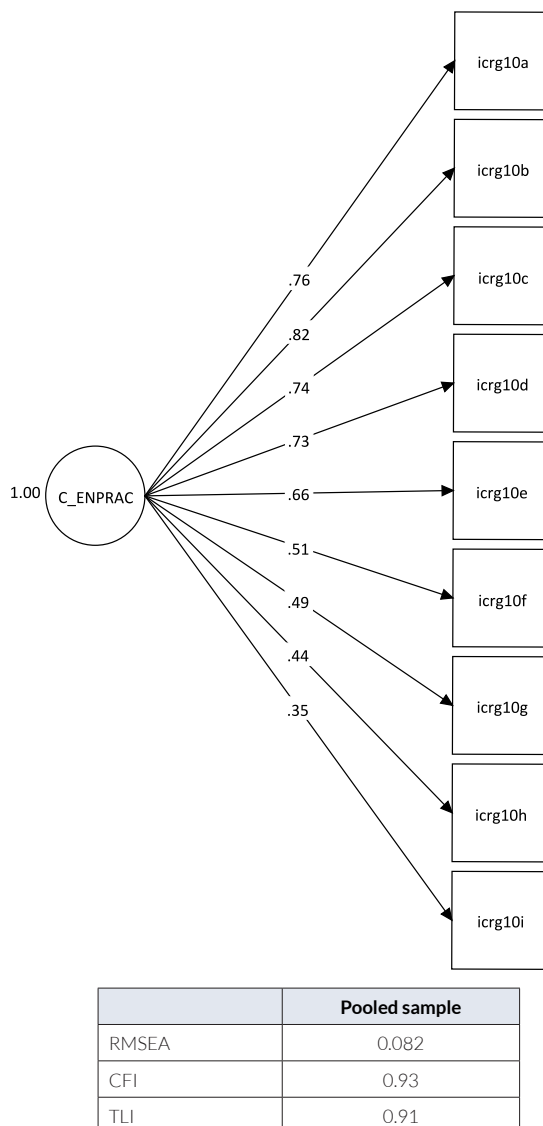


Table 12.80: Reliabilities for scale measuring school principals' reports on environment-friendly practices at school

Country	C_ENPRAC
Brazil	0.74
Bulgaria	0.61
Chinese Taipei	0.80
Colombia	0.77
Croatia	0.77
Cyprus	0.80
Denmark	0.64
Estonia	0.71
France	0.80
Italy	0.72
Latvia	0.69
Lithuania	0.73
Malta	0.69
Netherlands	0.57
North Rhine-Westphalia*	0.79
Norway	0.77
Poland	0.73
Romania	0.79
Schleswig-Holstein*	0.77
Serbia	0.85
Slovak Republic	0.66
Slovenia	0.68
Spain	0.75
Sweden	0.77
Average	0.73
Minimum value	0.57
Maximum value	0.85

Table 12.81: Item parameters for scale measuring school principals' reports on environment-friendly practices at school

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
C_ENPRAC	To what extent are the following practices implemented at this school?				
IC4G10A	Differential waste collection	-0.83	-0.51	-0.18	0.68
IC4G10B	Waste reduction (e.g. [encouraging waste-free lunches, limiting the use of plastic disposable products])	-0.46	-0.76	-0.21	0.97
IC4G10C	Purchasing of environmentally friendly items (e.g. [recycled paper for printing, biodegradable cutlery and dishes])	-0.12	-0.85	-0.18	1.04
IC4G10D	Energy-saving practices	-0.75	-1.12	-0.25	1.37
IC4G10E	Activities to encourage students' environmental-friendly behaviours (e.g. posters, leaflets)	-0.79	-1.13	-0.03	1.16
IC4G10F	Use of [fair trade] products (e.g. [tea or coffee in the staffroom, canteen food])	0.69	-0.37	-0.19	0.56
IC4G10G	Use of local food for meals in the canteen	0.37	0.29	-0.47	0.19
IC4G10H	Re-allocation of intact and non-consumed foods to charities or those in need	1.23	0.27	-0.32	0.05
IC4G10I	Educational [school gardens]	0.66	0.38	-0.40	0.02

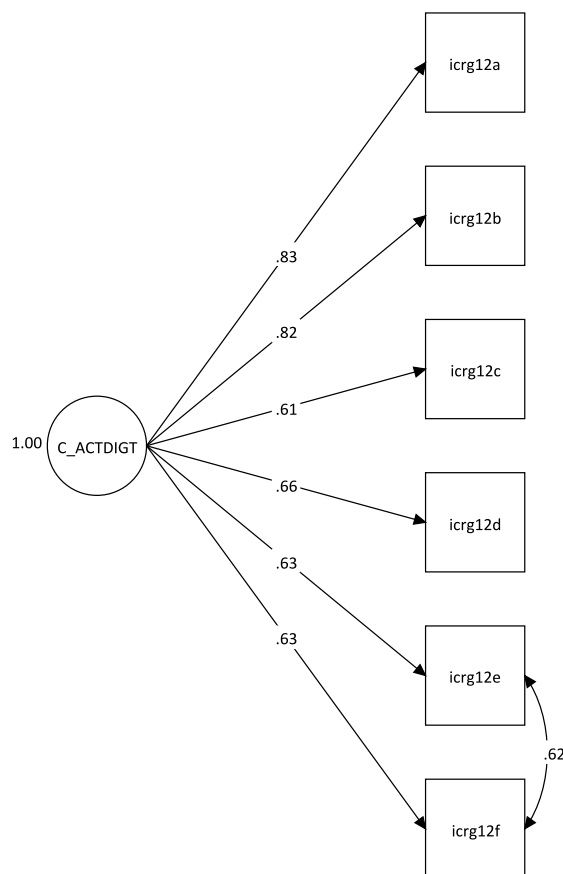
Schools' principals' reports on training activities related to the use of digital technologies

Question 12 of the school questionnaire required principals to state whether their school has undertaken training activities relating to the use of digital technologies ("yes, for teachers and students," "yes, only for teachers," "yes, only for students," or "no, this is not offered"). One scale was derived from this item set. Six items were used to derive a scale reflecting *school principals' reports on training activities related to the use of digital technologies* (C_ACTDIGT). Higher scale scores reflect a higher level of training activities taking place.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess schools' principals' reports on training activities related to the use of digital technologies (Figure 12.42). The model fit for the pooled sample was good.

The average reliability (Cronbach's alpha) across national samples were 0.73 for C_ACTDIGT (ranging from 0.55 to 0.86) (see Table 12.82). These scales were based on the item parameters shown in Table 12.83.

Figure 12.42: Confirmatory factor analysis of items measuring school principals' reports on training activities related to the use of digital technologies



	Pooled sample
RMSEA	0.054
CFI	0.99
TLI	0.98

Table 12.82: Reliabilities for scale measuring school principals' reports on training activities related to the use of digital technologies

Country	C_ACTDIGT
Brazil	0.84
Bulgaria	0.75
Chinese Taipei	0.55
Colombia	0.80
Croatia	0.68
Cyprus	0.76
Denmark	0.83
Estonia	0.83
France	0.84
Italy	0.75
Latvia	0.73
Lithuania	0.81
Malta	0.71
Netherlands	0.72
North Rhine-Westphalia*	0.79
Norway	0.65
Poland	0.72
Romania	0.86
Schleswig-Holstein*	0.56
Serbia	0.76
Slovak Republic	0.62
Slovenia	0.66
Spain	0.78
Sweden	0.68
Average	0.73
Minimum value	0.55
Maximum value	0.86

Table 12.83: Item parameters for scale measuring school principals' reports on training activities related to the use of digital technologies

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)
C_ACTDIGT	During the current school year, has your school undertaken any of the following training activities related to the use of digital technologies?			
IC4G12A	Information and data literacy (e.g. searching, managing and evaluating data, information and digital content)	-0.33	-0.46	0.46
IC4G12B	Communication and collaboration through digital technologies (e.g. interacting and sharing)	-0.78	-0.12	0.12
IC4G12C	Engaging in citizenship through digital technologies (e.g. posting and sharing or commenting on civic or social issues, starting or participating in an online campaign)	1.52	0.22	-0.22
IC4G12D	Management of digital identity (e.g. personal accountability, self-branding)	1.09	-0.23	0.23
IC4G12E	Online security and self-protection (e.g. protecting devices and personal data and privacy)	-0.52	-0.60	0.60
IC4G12F	Online responsibility (e.g. safe and responsible internet use to prevent and contrast [cyber-bullying])	-0.97	-1.02	1.02

Schools' principals' reports on the availability of resources in local community

Question 13 of the school questionnaire required principals to state whether a list of resources was available

("yes" or "no") in the immediate area where the school is located. Eleven items were used to derive a scale reflecting *school principals' reports on the availability of resources in the local community* (C_AVRESCOM). Higher scale scores reflect a higher level of resource availability.

We used confirmatory factor analyses assuming a one-dimensional model for these items to assess schools' principals' reports on the availability of resources in local community (Figure 12.43). The model fit for the pooled sample was good.

The average reliability (Cronbach's alpha) across national samples was 0.79 for C_AVRESCOM (ranging from 0.70 to 0.88) (see Table 12.84). These scales were based on the item parameters shown in Table 12.85.

Figure 12.43: Confirmatory factor analysis of items measuring school principals' reports on the availability of resources in local community

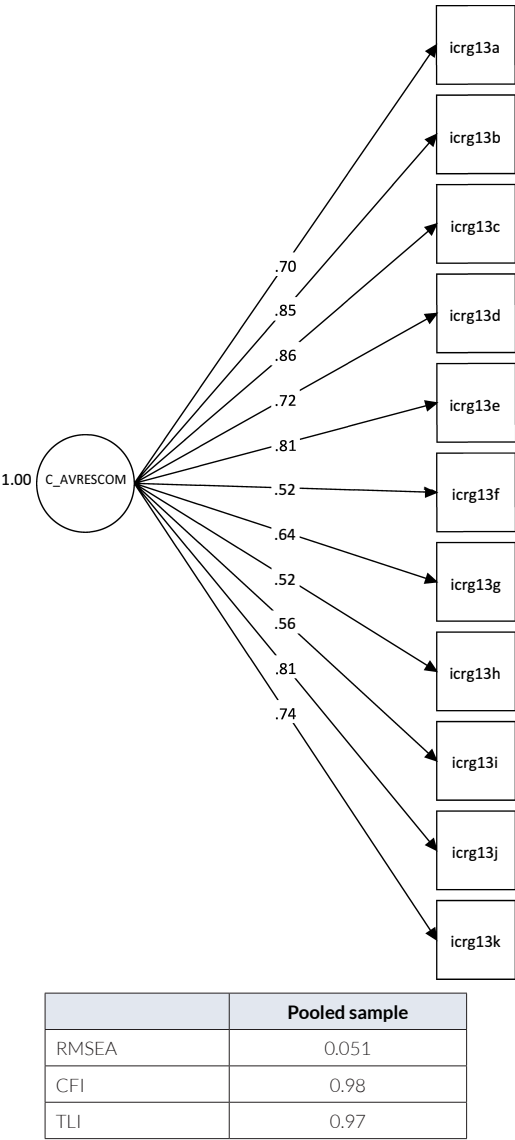


Table 12.84: Reliabilities for scale measuring school principals' reports on the availability of resources in local community

Country	C_AVRESCOM
Brazil	0.86
Bulgaria	0.85
Chinese Taipei	0.82
Colombia	0.84
Croatia	0.83
Cyprus	0.82
Denmark	0.79
Estonia	0.77
France	0.78
Italy	0.72
Latvia	0.75
Lithuania	0.70
Malta	0.77
Netherlands	0.73
North Rhine-Westphalia*	0.83
Norway	0.75
Poland	0.83
Romania	0.74
Schleswig-Holstein*	0.76
Serbia	0.88
Slovak Republic	0.82
Slovenia	0.83
Spain	0.79
Sweden	0.77
Average	0.79
Minimum value	0.70
Maximum value	0.88

Table 12.85: Item parameters for scale measuring school principals' reports on the availability of resources in local community

Scale or item	Question/item wording	Delta
C_AVRESCOM	Are the following resources available in the immediate area where the school is located?	
IC4G13A	Public library	-0.98
IC4G13B	Cinema	1.90
IC4G13C	Theatre or Concert Hall	1.55
IC4G13D	[Language school]	2.30
IC4G13E	Museum or Art Gallery	1.27
IC4G13F	Playground	-2.14
IC4G13G	Public garden or Park	-1.31
IC4G13H	Religious centre (e.g. [church, mosque, synagogue])	-2.07
IC4G13I	Sports facilities (e.g. swimming pool, tennis courts, basketball court, [football] field)	-1.96
IC4G13J	Music schools	0.96
IC4G13K	Cultural clubs or centres	0.49

Schools' principals' perceptions of social tensions in the community

Question 14 of the school questionnaire required principals to state their perception as to whether a list of issues are a source of social tensions ("To a large extent," "To a moderate extent," "To a small extent," or "Not at all") within the community where the school is located. Three scales were derived from this item set. Three items were used to derive a scale reflecting *principal's perceptions of ethnic tensions in the community* (C_COMETN), three items were used to derive a scale reflecting *principal's perceptions of poverty-related tension in the community* (C_COMPOV) and six items were used to derive a scale reflecting *principal's perceptions of crime-related tension in the community* (C_COMCRI). Higher scale scores reflect a higher perception as to the source of social tensions.

We used confirmatory factor analyses assuming a three-dimensional model for these items to assess schools' principals' perceptions of social tensions in the community (Figure 12.44). The model fit for the pooled sample was acceptable and the three latent factors were quite highly correlated for C_COMETN and C_COMPOV ($r = 0.60$), C_COMETN and C_COMCRI ($r = 0.74$), as well as C_COMPOV and C_COMCRI ($r = 0.62$).

The average reliabilities (Cronbach's alpha) across national samples were 0.77 for C_COMETN (ranging from 0.49 to 0.91), 0.83 for and 0.83 for C_COMPOV (ranging from 0.62 to 0.93) and 0.85 for C_COMCRI (ranging from 0.71 to 0.94) (see Table 12.86). These scales were based on the item parameters shown in Table 12.87.

Figure 12.44: Confirmatory factor analysis of items school principals' perceptions of social tensions in the community

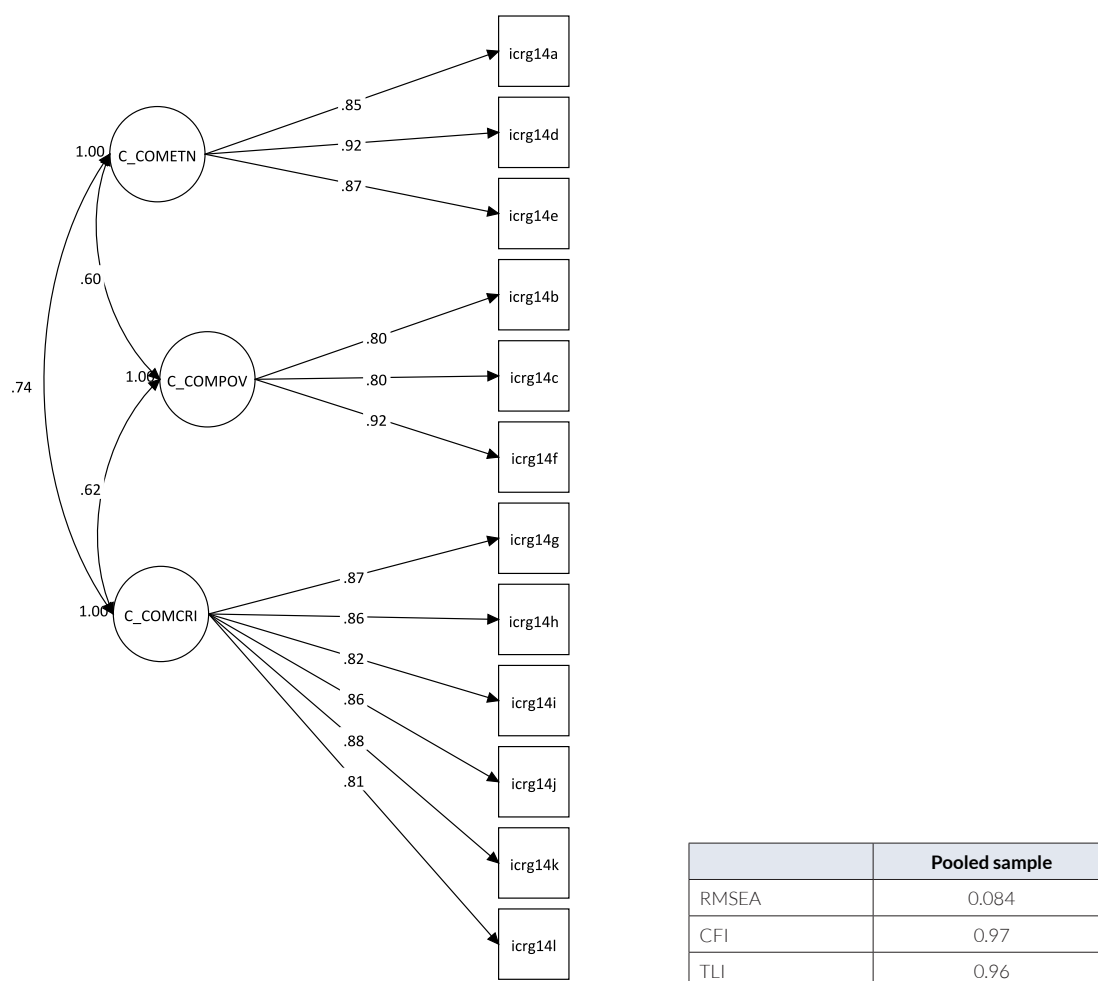


Table 12.86: Reliabilities for scales measuring school principals' perceptions of social tensions in the community

Country	C_COMETN	C_COMPOV	C_COMCRI
Brazil	0.74	0.86	0.93
Bulgaria	0.70	0.83	0.86
Chinese Taipei	0.91	0.84	0.92
Colombia	0.79	0.89	0.94
Croatia	0.73	0.62	0.79
Cyprus	0.89	0.86	0.87
Denmark	0.83	0.80	0.87
Estonia	0.69	0.81	0.83
France	0.82	0.84	0.86
Italy	0.84	0.82	0.87
Latvia	0.68	0.84	0.75
Lithuania	0.80	0.86	0.85
Malta	0.80	0.88	0.89
Netherlands	0.86	0.87	0.86
North Rhine-Westphalia*	0.86	0.93	0.89
Norway	0.83	0.89	0.87
Poland	0.71	0.70	0.79
Romania	0.49	0.75	0.71
Schleswig-Holstein*	0.61	0.89	0.87
Serbia	0.52	0.71	0.84
Slovak Republic	0.78	0.89	0.89
Slovenia	0.85	0.81	0.84
Spain	0.85	0.84	0.88
Sweden	0.85	0.89	0.87
Average	0.77	0.83	0.85
Minimum value	0.49	0.62	0.71
Maximum value	0.91	0.93	0.94

Table 12.87: Partial credit model item parameters for scales measuring school principals' perceptions of social tensions in the community

Scale or item	Question/item wording	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?				
IC4G14A	Intolerance towards immigrants	-0.53	-3.46	0.40	3.06
IC4G14D	Religious intolerance	0.27	-3.77	0.56	3.21
IC4G14E	Ethnic conflicts	0.25	-3.29	0.62	2.67
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?				
IC4G14B	Poor quality of housing	0.48	-3.44	0.38	3.06
IC4G14C	Unemployment	-0.95	-3.40	0.48	2.91
IC4G14F	Extensive poverty	0.46	-2.98	0.37	2.60
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?				
IC4G14G	Organised crime	0.77	-2.61	0.41	2.20
IC4G14H	Youth gangs	0.51	-2.83	0.22	2.61
IC4G14I	[Petty crime]	-0.66	-4.04	0.64	3.39
IC4G14J	[Sexual harassment]	0.84	-3.33	0.88	2.45
IC4G14K	Drug abuse	-0.45	-3.40	0.49	2.91
IC4G14L	Alcohol abuse	-1.01	-3.62	0.12	3.50

Latin American student questionnaire

As in previous cycles, ICCS 2022 included a regional questionnaire for participating countries from the Latin American region. However, while during the first study phase the study counted on the participation of five to six countries, only two countries (Brazil and Colombia) finally implemented the main survey using a reduced version of the instrument only including material from the previous cycles. Because data from

only two countries do not provide sufficient information for cross-country analyses, scaling reviews and re-calibrations, the international study team decided to derive only four scales that were derived using the same IRT item parameters and transformations as in the previous survey (ICCS 2016). This makes it possible to compare results from Colombia for these four scales with those from the previous cycles.

The following scales were constructed for ICCS 2022 and are included in the corresponding databases for Brazil and Colombia:

- *Students' endorsement of authoritarian government practices* (L_AUTGOV): This scale was derived using seven items included in Question 2 and scores were equated to ICCS 2009.
- *Students' endorsement of corrupt practices in government* (L_ATTCCORR): This scale was derived based on five items included in Question 3 and scores were equated to ICCS 2009.
- *Students' acceptance of disobeying the law* (L_DISLAW): This scale was derived using ten items included in Question 2 and scores were equated to ICCS 2009.
- *Students' positive attitudes toward homosexuals* (L_ATTHS): This scale was derived based on five items included in Question 5 and scores were equated to ICCS 2016.

All these scales were constructed based on their respective ICCS 2016 item parameters (Table 12.86), and subsequently transformed using the linear transformations displayed in Table 12.87. IRT scores (θ_n) based on the ICCS 2016 item parameters were set on their reporting metric (θ'_n) established in ICCS 2009 or 2016 using the following equation:

$$\theta'_n = A + B \times \theta_n$$

Table 12.88: Item parameters for Latin American questionnaire scales

Scale or item	Question/item wording	Delta	Tau (1)	Tau (2)	Tau (3)
L_AUTGOV	<i>How much do you agree or disagree with the following statements about the government and its leaders?</i>				
LS4G02A	It is better for government leaders to make decisions without consulting anybody.	0.49	-1.25	0.82	0.43
LS4G02B	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
LS4G02C	People whose opinions are different than those of the government must be considered its enemies.	0.54	-1.56	0.78	0.78
LS4G02D	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
LS4G02E	The government should close communication media that are critical.	0.30	-1.69	0.40	1.29
LS4G02F	Concentration of power in one person guarantees order.	-0.85	-1.73	0.06	1.67
LS4G02G	If the president does not agree with <Congress>, he/she should <dissolve> it.	-0.19	-1.80	0.33	1.47
L_ATTCORR	<i>How much do you agree or disagree with the following statements about the public service and government?</i>				
LS4G03A	It is acceptable for a civil servant to accept bribes if his salary is too low.	0.36	-1.69	0.81	0.88
LS4G03B	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
LS4G03C	Good candidates grant personal benefits to voters in return for their votes.	-0.08	-1.72	0.20	1.52
LS4G03D	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
LS4G03E	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
L_DISLAW	<i>How much do you agree or disagree with the following statements about situations where the law is disobeyed?</i>				
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
L_ATTHS	<i>How much do you agree or disagree with the following statements with regard to homosexual orientations?</i>				
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

Table 12.89: Linear transformations applied to set Latin American questionnaire scale scores to reporting metric

Scale	A	B
L_AUTGOV	53.56	7.80
L_ATT CORR	53.96	5.67
L_DISLAW	49.44	7.68
L_ATT HS	45.92	4.48

12.5 Reviewing psychometric differences across delivery modes

As for cognitive data, the delivery of student questionnaires on paper in additional student samples (*bridge study*) in 11 participating countries (that had opted for computer-based delivery and participated in the previous cycle) allowed an analysis of psychometric item characteristics across the two delivery modes used in ICCS 2022. In a first step we compared IRT scaling parameters from separate calibrations on paper using bridging data with the final ones that were derived from the combined international student sample. Table 12.90 shows the correlations between two sets of the combined item-step parameters (delta parametrization) as well as between Thurstonian thresholds for each of the 30 scales derived from the international student questionnaire.

Table 12.90: Correlations between IRT parameters from computer- and paper-based

Scale	Item-step parameters	Thurstonian thresholds
S_POLDISC	0.999	1.000
S_ENGDM	0.997	0.999
S_COMPART	0.994	0.993
S_SCHPART	0.997	0.996
S_OPDISC	0.998	0.999
S_CIVLRN	0.998	0.999
S_STUTREL	0.999	0.999
S_INTACT	1.000	1.000
S_INFDEC	0.999	0.999
S_SYSSAT	1.000	1.000
S_SYSCRT	1.000	0.999
S_DEMTHRT	0.995	0.997
S_CITCON	0.998	0.997
S_CITSOC	0.999	1.000
S_GLOBCIT	0.999	0.999
S_ENREST	0.998	0.998
S_IMMPOS	0.999	1.000
S_GENEQL	0.987	0.995
S_ATTENV	0.997	0.997
S_ETHRGHT	0.998	0.999
S_INTRUST	0.999	0.999
S_ENVCON	0.989	0.998
S_CITEFF	0.999	0.999
S_SCACT	0.999	0.999
S_LEGACT	0.999	0.999
S_ENVACT	0.997	0.998
S_ILLACT	0.992	0.998
S_ELECPART	0.999	0.997
S_POLPART	0.998	0.999
S_RELINF	0.999	0.999

For all 30 scales we found almost perfect correlations between IRT parameters from computer- and paper-based calibrations. These findings confirmed a very high level of measurement invariance across delivery modes for the parameters of the respective scaling models. This result provided a good basis for using the same item parameters (based on the calibration of the combined national core samples) for deriving the scales for both paper-based and computer-based student questionnaire data.

We further reviewed differences in outcomes between average scales scores across bridging study and main survey samples in the 11 countries that had data from both delivery modes (see Tables A.1 to A.4 in Appendix A). While for most scales and countries we found no statistically significant differences, for some we recorded relatively minor differences for the following indices:

- S_CITCON: Students' beliefs of the importance of conventional citizenship activities tended to be somewhat higher among those responding the questionnaire on computer with an average difference of more than one score point, these differences were significant in nine out of 11 countries.
- S_GENEQL: Students' endorsement of gender equality was somewhat higher among those completing the paper version with an average difference of -0.6 score points and significant differences in four countries (Chinese Taipei, Colombia, Latvia, and Norway).
- S_CITEFF: Students' citizenship self-efficacy tended to be rated higher when responding on a computer, with an average difference of 1.4 score points and significant differences in seven out of 11 countries.
- S_LEGACT: Students' expected participation in legal activities to express opinions tended to be higher when completing the questionnaire on a computer, with an average difference of one score point and significant differences in six out of 11 countries.
- S_ILLACT: Students were significantly more likely to expect participation in illegal protest when responding on computer in seven out of 11 countries with an average difference of 1.3 score points.
- S_POLPART: Students' expectations to engage in active political participation tended to be higher when provided on computer in all but two countries with an average difference of 1.3 score points.

The observed differences in student responses to the underlying questionnaire items of these scales suggest that respondents may have reacted differently to the delivery mode in terms of their openness to express opinions. Previous research has found that respondents, possibly also due to privacy concerns, may express beliefs more openly in computer-based environments than on paper (see, for example, Feigelson & Dwight, 2000; Hart & Goldstein, 1985); however, other research has also claimed that computer-based questionnaire may slightly increase socially desirable responses (Finegan & Allen, 1994). The findings from these comparative analyses should be taken into consideration when interpreting cross-nationally comparative results from ICCS 2022 questionnaire scales, especially for those for which we highlighted more systematic differences.

References

- Adams, R. J., Wu, M. L., Cloney, D., Berezner, A., & Wilson, M. (2020). *ACER ConQuest: Generalised Item Response Modelling Software* (Version 5.29) [Computer software]. Australian Council for Educational Research. <https://www.acer.org/au/conquest>
- Bentler, P.M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606. <https://doi.org/10.1037/0033-2909.88.3.588>
- Bollen, K. A., & Long, S. J., (Eds.). (1993). *Testing structural equation models*. Sage Publications.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford 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. <https://doi.org/10.1007/BF02310555>
- Damiani, V. Losito, B., Agrusti, G., Schulz, W. (2024). *Young citizens' views and engagement in a changing Europe. IEA international civic and citizenship education study 2022 European report*. Springer.
- 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. <https://doi.org/10.1037/1082-989X.4.3.272>
- Feigelson, M. E., & Dwight, S. A. (2000). Can asking questions by computer improve the candidness of responding? A meta-analytic perspective. *Consulting Psychology Journal: Practice and Research*, 52(4), 248–255. <https://doi.org/10.1037/1061-4087.52.4.248>
- Finegan, J.E., & Allen, N. J. (1994). Computerized and written questionnaires: Are they equivalent? *Computers in Human Behavior*, 10(4), 483–496. [https://doi.org/10.1016/0747-5632\(94\)90042-6](https://doi.org/10.1016/0747-5632(94)90042-6)
- 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. <https://doi.org/10.1037/1082-989X.9.4.466>
- 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), 1–56. [https://doi.org/10.1016/0049-089X\(92\)90017-B](https://doi.org/10.1016/0049-089X(92)90017-B)
- Hart, R. R., & Goldstein, M. A. (1985). Computer-assisted psychological assessment. *Computers in Human Services*, 1(3), 69–75. https://doi.org/10.1300/J407v01n03_05
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3-4), 117–144. <https://doi.org/10.1080/03610739208253916>
- 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. <https://doi.org/10.1080/10705519909540118>
- International Labour Organization. (2007). *International Standard Classification of Occupations: ISCO-2008*. International Labour Office.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183–202. <https://doi.org/10.1007/BF02289343>
- Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions*. SAGE.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking* (2nd edition). Springer Science and Business Media, Inc. <https://doi.org/10.1007/978-1-4757-4310-4>

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. https://doi.org/10.1207/s15327906mbr3201_3

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. Springer. <https://doi.org/10.1007/978-3-319-73960-1>

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. <https://doi.org/10.1037/1082-989X.1.2.130>

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>

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). Springer. https://doi.org/10.1007/978-1-4757-2691-6_6

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). Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edition). McGraw-Hill.

OECD (2023). *PISA 2022 Technical Report*. Paris: OECD Publishing.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.

Schulz, W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study. In: *IERI Monograph Series Volume 2*, 113–135.

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*. Springer. <https://doi.org/10.1007/978-3-319-73963-2>

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). 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). International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., & Friedman, T. (2018). Scaling procedures for ICCS questionnaire items. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.), *ICCS 2016 technical report* (pp. 139–243). International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., & Friedman, T. (2020). Scaling procedures for ICILS 2018 questionnaire items. In J. Fraillon, W. Schulz, T. Friedman, & S. Meyer (Eds.), *ICILS 2018 technical report* (pp. 159–219). International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., Damiani, V., Friedman, T. (2024). *Education for citizenship in times of global challenge. IEA International Civic and Citizenship Education Study 2022 international report*. Springer.

Tucker, L., & MacCallum, R. (1997). Exploratory factor analysis. Unpublished manuscript.

UNESCO. (2006). International standard classification of education: ISCED 1997. UNESCO-UIS.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427-450. <https://doi.org/10.1007/BF02294627>

Wu, H., Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>

Chapter 13:

The reporting of ICCS 2022 results

Wolfram Schulz

13.1 Overview

This chapter describes the procedures used for reporting results in the International Civic and Citizenship Education Study (ICCS) 2022 publications (Schulz, et al., 2023; Damiani et al., 2024). 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 provided information about significance tests for differences between country and subsample means or percentages, as well as between results from ICCS 2022 and the two previous cycles in 2016 and 2009 for countries that participated in this and the preceding survey(s).

The chapter also includes descriptions of the multilevel (hierarchical) models used to estimate variation in students' civic knowledge, and the multiple regression analyses developed to explain variation in questionnaire scale scores reflecting students' expected electoral and active political participation.

13.2 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 6 for further details). During the second stage, intact classrooms were randomly sampled within schools. Cluster sampling techniques permit efficient and economic data collections. 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.

Large-scale assessments typically apply replication techniques that provide appropriate tools for the estimation of sampling variance on population estimates (Gonzalez & Foy, 2000; Wolter, 1985). As in previous cycles (see Schulz, 2011 & 2018), ICCS 2022 data were analyzed using 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, typically school—into pseudo-strata. Because assignment of schools to these “sampling zones” needs to be consistent with the sampling frame from which they were sampled, it is necessary to construct *sampling zones* within explicit strata. When faced with occurrences of an odd number of schools within an explicit stratum or the sampling frame, the ICCS 2022 sampling team randomly divided the remaining school into two halves, thereby forming a sampling zone of two “quasi” schools. If such a school consisted of 2 sampled classrooms, each classroom was treated as a “quasi” school.

Each of the countries participating in ICCS 2022 had up to 75 sampling zones. In countries with more than 150 schools, in order to keep the total number to 75, additional zones were collapsed into the first 75 zones. In Cyprus, where typically two classrooms were sampled from all schools with target grade students, the sampling team defined classrooms as PSUs for 62 schools, while randomly allocating students from the remaining schools to different sampling zones. In Malta a census of all schools was conducted, with most schools drawn randomly for the core survey while others were allocated to a sub-sample for the bridging study. Here, schools were allocated in the usual way resulting in 15 sampling zones for the national core school sample.

Within each of the sampling zones, one half (typically a school or classroom) was randomly assigned a value of 2 and the other school a value of 0. Replicated weights were computed for each of the sampling zones. This meant that one of the paired schools had a contribution of zero, the second a double contribution, and

all other schools remained the same. The replicate weights procedure is achieved by multiplying student or teacher weights with the jackknife indicators once only within 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 twice its original 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 13.1).

Table 13.1: 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, there are 75 replicate weights regardless of the number of sampling zones. In countries with fewer sampling zones, the remaining replicate weights are equal to the original sampling weight and therefore do 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 2022 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.¹ Alternatively, analysts can use software packages with inbuilt procedures for replication methods such as SAS, STATA, MPLUS, or other more specialized software, such as for example WESVAR® (Westat, 2007).

13.3 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 2022 test items). Therefore, population statistics for ICCS 2022 civic knowledge scores should always be estimated using all five plausible values. While each plausible value constitutes an unbiased estimate of the population value, only employing all five plausible values in the database provides a way of estimating the uncertainty of measurement in combination with the sampling variance as well as replicating the results published in the ICCS 2022 reports.

If θ is the population average of civic knowledge (at the level of the combined sample, individual countries, or sub-groups) 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 one to estimate the amount of error associated with the measurement of civic knowledge. The measurement variance or imputation variance IV_p is computed as:

$$IV_p = \frac{1}{P-1} \sum_{p=1}^P (\mu_{\theta}^p - \mu_{\theta})^2 .$$

Here, 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:

$$TV_{\mu} = SV_{\mu} + \left(1 + \frac{1}{P}\right) IV_{\mu} .$$

The estimate of the final standard error SE_{μ} is equal to:

$$SE_{\mu} = \sqrt{TV_{\mu}} .$$

¹ The IDB Analyzer is a plug-in for the Statistical Package for the Social Sciences (SPSS) that allows the user to combine and analyze data from IEA's large-scale assessments such as TIMSS, PIRLS, and SITES. The application can be downloaded at <http://www.iea.nl/index.php/data-tools/tools#section-308>

For the reporting of civic knowledge, we calculated the average scale scores and their sampling and overall standard errors for each country or benchmarking entity (Table 13.2). 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 13.2: 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
Bulgaria	456	4.61	4.64	3113
Chinese Taipei	583	2.26	2.30	3852
Colombia	452	3.75	3.76	4193
Croatia ¹	531	2.39	2.60	2759
Cyprus	459	2.39	2.48	3136
Estonia	545	5.45	5.50	2685
France	508	3.07	3.27	3533
Italy	523	3.52	3.57	2945
Latvia ¹	490	2.77	2.84	2876
Lithuania	509	3.89	3.98	3560
Malta	490	7.39	7.43	2720
Netherlands‡	508	4.04	4.09	2609
Norway (9) ¹	529	2.73	2.79	5673
Poland	554	2.49	2.52	4434
Romania	470	9.07	9.11	2768
Serbia	464	3.18	3.39	2640
Slovak Republic	501	3.19	3.27	3202
Slovenia	504	2.25	2.29	3466
Spain	510	3.15	3.26	3487
Sweden ¹	565	3.46	3.49	3263
Countries not meeting sample participation requirements				
Brazil	457	3.22	3.27	5885
Denmark	556	3.47	3.54	4769
German benchmarking participant meeting sample participation requirements				
North Rhine-Westphalia	524	2.44	2.62	3266
German benchmarking participant not meeting sample participation requirements				
Schleswig-Holstein	544	4.36	4.43	1487

Notes:

(9) Country deviated from international defined population and surveyed adjacent upper grade

‡ Nearly met guidelines for sampling participation rates only after replacement schools were included

¹ National defined population covers 90% to 95% of national target population

13.4 Reporting of differences

In our reporting of ICCS 2022 results, we conducted a variety of significance tests² for:

- Differences in population estimates between countries (pair-wise comparisons);
- Differences between a country and the ICCS 2022 average;
- Differences in population estimates between subgroups within countries;
- Differences between population estimates in ICCS 2022 and 2016 (estimation of changes over time) for countries that participated in both surveys; and
- Differences between population estimates in ICCS 2022 and 2009 (estimation of changes over time) for countries that participated in both surveys.

In the ICCS 2022 reports, we considered differences between two score averages (or percentages) *a* and *b* significant ($p < 0.05$) when the absolute value of the test statistic *t* was greater than the critical value, 1.96. We calculated *t* by dividing the difference by its standard error, $SE_{diff,ob}$:

² When reporting on statistical significance of difference in ICCS 2022 reports, we did not apply any methods that adjust for multiple comparisons (see, for example, Streiner, 2015).

$$t = \frac{(a - b)}{SE_{dif\_ab}} .$$

In the case of differences between score averages from independent samples (evident, for example, with respect to comparisons of country averages), the standard error of the difference SE_{dif_ab} can be computed as:

$$SE_{dif\_ab} = \sqrt{SE_a^2 + SE_b^2} .$$

Here, SE_a and SE_b are the standard errors of the means from the two independent samples a and b .

When comparing subsamples within countries that are not independent samples (e.g., gender groups), it became necessary to derive the standard error on the difference between statistics for subgroups of interest 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:

$$SE_{dif\_ab} = \sqrt{\left[\sum_{p=1}^P \left(\sum_{i=1}^{75} ((a_p^i - b_p^i) - (a_p - b_p))^2 \right) / P \right] + \left[\left(1 + \frac{1}{P} \right) \frac{\sum_{p=1}^P ((a_p - b_p) - (\bar{a}_p - \bar{b}_p))^2}{P - 1} \right]}$$

Here, a_p and b_p represent the weighted subgroup averages in groups a and b for each of the P plausible values, a_p^i and b_p^i are the subgroup averages within replicate samples for each of the P plausible values, and $\bar{a}_p$ and $\bar{b}_p$ are the means of the two weighted subgroup averages across the P plausible values.

When comparing the country means c with the overall ICCS 2022 average i , we had to account for the fact that each country meeting IEA guidelines for sample participation had contributed to the international standard error. To obtain the standard error SE_{dif_ic} of the difference between the overall ICCS 2022 average and an individual country average we calculated it 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 2022 averages of the questionnaire scales throughout the ICCS 2022 reports.

While the above formula was sufficient for the questionnaire scale scores, it was necessary to also take the imputation component of standard errors for countries into account when comparing the civic knowledge score averages of a country with the overall ICCS 2022 average. The imputation variance component of standard errors $SE_{i\_dif\_ic}^2$ was given as:

$$SE_{i\_dif\_ic}^2 = \sqrt{\left(1 + \frac{1}{P} \right) var(d_1, \dots, d_p, \dots, d_5)} ,$$

where d_p is the difference between the overall ICCS 2022 and the country mean for the plausible value p .³

We computed the final standard error (SE_{a,dif_ic}^2) of the difference between ICCS 2022 country civic knowledge scores and the ICCS 2022 average as:

$$SE_{a,dif\_ic} = \sqrt{SE_{dif\_ic}^2 + SE_{i,dif\_ic}^2}.$$

As estimates based on the student survey in the benchmarking entity North Rhine-Westphalia, which had met IEA sampling participation standards, were not included in the ICCS 2022 average, it was sufficient to apply the formula for independent samples to estimate determine the statistical difference by calculating the standard error for the difference as:

$$SE_{dif\_ic} = \sqrt{SE_{bp}^2 + SE_{\mu\_ICCS22}^2}$$

Here, SE_{bp} is the standard error for the result from the benchmarking participant, and SE_{μ_ICCS22} is the standard error for the ICCS 2022 average.

The ICCS 2022 international report also included comparisons of student survey results between ICCS 2016 and the previous cycles in 2016 and 2009. 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. The equating errors were estimated based on the variation of parameters for link items across the ICCS 2022 and 2016 cycles (see Chapters 11 and 12), and it needs to be acknowledged that these are approximations of its true value, which may vary across different points of the corresponding scale.

When testing the difference of a statistic between ICCS 2022 and 2016, we computed the standard error of the difference as follows:

$$SE(\mu_{ICCS22} - \mu_{ICCS16}) = \sqrt{SE_{(ICCS22)}^2 + SE_{(ICCS16)}^2 + EqErr_{2022\_2016}^2}$$

Here, μ can be any statistic in units on the equated ICCS scale (mean, percentile, gender difference, but not percentages) and $SE_{(ICCS22)}$ and $SE_{(ICCS16)}$ are the respective standard errors of this statistic from the two surveys. $EqErr_{2022\_2016}$ denotes the equating error that reflects an estimate of the uncertainty in the link between the adjacent surveys, which was equal to 2.6495 score points for the civic knowledge scale and ranged from 0.038 to 0.413 score points across the international and regional equated questionnaire scales (see Chapter 11 and 12 for the calculation of the respective equating errors). In Chapter 12, the equating errors for questionnaire scales are listed according to their respective reporting metrics (see Table 12.2).

For many ICCS 2022 countries, it was also possible to compare results with the first ICCS cycle in 2009. When testing the difference of a statistic between the first and third ICCS survey, we computed the standard error of the difference as follows:

$$SE(\mu_{ICCS22} - \mu_{ICCS09}) = \sqrt{SE_{(ICCS22)}^2 + SE_{(ICCS09)}^2 + EqErr_{2022\_2016}^2 + EqErr_{2016\_2009}^2}$$

In this equation, $EqErr_{2022\_2016}$ denotes the equating error estimated for the link between ICCS 2022 and ICCS 2016, while $EqErr_{2016\_2009}$ denotes the equating error for the link between ICCS 2016 and ICCS 2009, each reflecting the uncertainty in the link between consecutive assessments. For 2016 and 2009 this was equal to 3.0861 score points for the civic knowledge scale (see Gebhardt & Schulz, 2018) and ranged from 0.019 to 0.726 score points across the international and regional equated questionnaire scales (see Schulz & Friedman, 2018). Table 13.3 shows the respective equating errors between ICCS 2016 and 2022 as well as between 2009 and 2016 for civic knowledge and (international and European) student questionnaire scales.

³ Please note this formula was incorrectly displayed in the ICCS 2016 technical report (Schulz, 2018).

Table 13.3: Estimates of equating errors (scale scores)

Scale	Equating Error 2016/2022	Equating Error 2009/2016
Civic knowledge (PVS)	2.649	3.086
S_CITCON	0.240	0.411
S_CITEFF	0.202	0.261
S_CITSOC	0.169	0.428
S_CIVLRN	0.240	-
S_ELECPART	0.211	0.019
S_ETHRGHT	0.318	0.250
S_GENEQL	0.329	0.423
S_IMMRGHT	0.574	0.667
S_INTACT	0.413	-
S_INTRUST	0.155	0.291
S_ILLACT	0.133	-
S_LEGACT	0.168	-
S_OPDISC	0.212	0.340
S_POLDISC	0.127	0.438
S_POLPART	0.061	0.351
S_RELINF	0.115	0.115
S_SCACT	0.173	-
S_STUTREL	0.113	0.454
E_EUIDENT	0.177	0.726
E_FREEMOVE	0.040	-
E_INDFUT	0.319	-

The equating error could not be applied directly to report the significance of differences between percentages at or above Level B or at or below Level D between ICCS 2022 and ICCS 2016 or ICCS 2009, respectively. Therefore, we applied the following replication method to estimate the equating error for percentages at or above a certain level of proficiency (e.g., students with civic knowledge test scores corresponding to Levels A and B or to Levels D and below).

To estimate the standard error of the percentage at or above the cut-point that defines the threshold between Levels C and B (479) or between Levels Level D and C (395), based on the data from the ICCS 2022 survey, 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 (2.6895). 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_o is the percentage of students at or above Level B, or at or below Level D (Table 13.3). We used 1000 replicate samples (n) to compute these estimates. It is important to note that this procedure could not be applied to countries that did not participate in ICCS 2016 (Cyprus, Slovak Republic, and Spain), given that it is only based on ICCS 2022 data in combination with the equating error between ICCS 2016 and 2022 in civic knowledge score points. Table 13.4 includes the estimated equating errors for percentages for each country and the pooled ICCS samples.

Table 13.4: Estimates of equating error for percentages

Country	At or above Level B		At or above Level D	
	2022/2016	2016/2009	2022/2016	2016/2009
Bulgaria	0.81	0.83	0.78	0.56
Chinese Taipei	0.51	0.70	0.17	0.19
Colombia	1.05	1.45	1.25	0.84
Croatia	0.90	1.16	0.35	0.36
Cyprus	0.79		0.90	
Estonia	0.77	1.08	0.31	0.25
Italy	0.83	1.12	0.52	0.40
Latvia	1.39	1.39	0.53	0.73
Lithuania	1.00	1.23	0.51	0.60
Malta	0.97	0.95	0.63	0.62
Netherlands	0.79	1.07	0.44	0.53
Norway	0.88	0.61	0.34	0.34
Poland	0.71		0.23	
Slovak Republic	0.93		0.82	
Slovenia	1.02	1.11	0.52	0.43
Spain	0.93		0.43	
Sweden	0.81	0.62	0.27	0.25
ICCS (pooled sample)	0.90	0.98	0.57	0.52

When comparing percentages of students at or above Level B (or for those at or below Level D) between ICCS 2022 and ICCS 2016, within each participating country, using the calculated equating error based on the ICCS 2022 data ($EqErr_{(p\_cnt22)}$), the standard errors for the differences between percentages (at or above Level B, or at or below Level D) were calculated as:

$$SE(\rho_{ICCS22} - \rho_{ICCS16}) = \sqrt{SE(\rho_{ICCS22})^2 + SE(\rho_{ICCS16})^2 + EqErr_{(p\_cnt22)}^2}$$

When comparing percentages of students at or above Level B (or for those at or below Level D) between ICCS 2022 and ICCS 2009, for each participating country that participated in all three cycles, using the calculated equating error based on the ICCS 2022 data ($EqErr_{(p\_cnt22)}$) as well as those calculated based on ICCS 2016 data ($EqErr_{(p\_cnt16)}$), the standard errors for the differences between percentages (at or above Level B, or at or below Level D) were derived as:

$$SE(\rho_{ICCS22} - \rho_{ICCS09}) = \sqrt{SE(\rho_{ICCS22})^2 + SE(\rho_{ICCS09})^2 + EqErr_{(p\_cnt22)}^2 + EqErr_{(p\_cnt16)}^2}$$

When comparing percentages of students at or above Level B (or for those at or below Level D) between ICCS 2022 and ICCS 2009 for countries that had only participated in these two comparison cycles, it became necessary to use a different formula. Here, for each of these countries (Cyprus, Poland, Slovak Republic, and Spain), we used the estimated equating error based on the respective national ICCS 2022 data ($EqErr_{(p\_cnt22)}$) as well as the estimated error for the combined ICCS 2016 data from all countries ($EqErr_{(p\_ICCS16)}$) to obtain the standard errors for the differences between percentages (at or above Level B, or at or below Level D) as follows:

$$SE(\rho_{ICCS22} - \rho_{ICCS09}) = \sqrt{SE(\rho_{ICCS22})^2 + SE(\rho_{ICCS09})^2 + EqErr_{(p\_cnt22)}^2 + EqErr_{(p\_ICCS16)}^2}$$

13.5 Hierarchical linear modeling

To review which factors were 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 these models students were nested within intact classrooms that had been randomly selected within schools. As in the previous international report for ICCS 2016, we used predictor 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, classrooms) can be estimated as

$$Y_{ij} = \alpha_j + X_{ij}^n \beta_{ij} + X_j^m \beta_j + U_{0j} + \varepsilon_{ij} ,$$

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_{ij} , and X_j^m is a school- (or classroom-) level variable with its corresponding vector of regression coefficients β_j . U_{0j} is the residual term at the level of the cluster (school or classroom), 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

$$EV_j = (1 - \frac{U_{0j}^{null}}{U_{0j}^{null}}) * 100 ,$$

and we computed the explained variance at the student level EV_{ij} as

$$EV_{ij} = (1 - \frac{\varepsilon_{ij}^{null}}{\varepsilon_{ij}^{null}}) * 100 .$$

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 et al., 2010).

We used the software package MPlus (Version 7; see Muthén & Muthén, 2012) to estimate all hierarchical models. We excluded data from Malta from the analyses due to their low number of only 29 schools. Results from Brazil, Denmark and the benchmarking participant Schleswig-Holstein (Germany) 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 (weighted) of students on average across ICCS 2022 countries meeting sample participation requirements had valid data for all variables included in the model. In the benchmarking entity North Rhine-Westphalia only 84 percent of the weighted sample had valid data for inclusion in the analyses, while in Denmark, where

the national sample did not meet IEA sample participation guidelines due to higher levels of non-response, this percentage was lowest with 80 percent.

We calculated the respective unweighted and weighted percentages of students with valid data for all variables in the model (Table 13.5).

Table 13.5: ICCS 2022 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
Bulgaria	3113	2717	87	86
Chinese Taipei	3852	3761	98	98
Colombia	4193	3807	91	89
Croatia ¹	2759	2639	96	96
Cyprus	3136	2828	90	90
Estonia	2685	2621	98	98
France	3533	3280	93	92
Italy	2945	2868	97	97
Latvia ¹	2876	2712	94	94
Lithuania	3560	3400	96	96
Malta	2720	2544	94	93
Netherlands‡	2609	2293	88	87
Norway (9) ¹	5673	5123	90	90
Poland	4434	4285	97	97
Romania	2768	2612	94	95
Serbia	2640	2402	91	91
Slovak Republic	3202	3038	95	95
Slovenia	3466	3298	95	95
Spain	3487	3237	93	93
Sweden ¹	3263	3083	94	95
Countries not meeting sample participation requirements				
Brazil	5885	5271	90	90
Denmark	4769	3764	79	80
German benchmarking participant meeting sample participation requirements				
North Rhine-Westphalia	3266	2704	83	84
German benchmarking participant not meeting sample participation requirements				
Schleswig-Holstein	1487	1360	91	91

Notes:

(9) Country deviated from international defined population and surveyed adjacent upper grade

‡ Nearly met guidelines for sampling participation rates only after replacement schools were included

¹ National defined population covers 90% to 95% of national target population

In most countries, one intact classroom per school was sampled, which did not allow disentangling classroom- and school-level variance. In Cyprus, two or more 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 considered when interpreting the results of the multilevel analyses of ICCS 2022 data, which are presented in Chapter 7 of the international report (Schulz et al., 2024).

13.6 Multiple regression modeling

As in ICCS 2016, the international report for ICCS 2022 also included results from single-level multiple regression models to explain variation in the questionnaire scale scores reflecting students' expected electoral and active political participation, respectively (see Chapter 7 in Schulz et al., 2024). Predictor variables were student background variables, variables related to information sources, experience with civic participation, students' perceptions of the political system and its institution, and dispositions toward civic engagement.

Multiple regression models (see, for example, Pedhazur, 1997) 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 the number of regression coefficients β :

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

ICCS 2022 reported the unstandardized regression coefficients and the variance explained by the model to show the effects for each predictor as well as 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).

For these analyses, a listwise exclusion of missing responses was applied, that is, only students with valid data for all variables were included. Across the participating countries, on average 88 percent of students in the sample were included in the analyses of both dependent variables (Table 13.6). National (weighted) average percentages of students with valid data for all variables ranged from 81 percent in Bulgaria to 97 or 98 percent (respectively) in Chinese Taipei. The lowest percentages were recorded for the benchmarking participant North Rhine-Westphalia (80%) and Denmark (75%), which had not met IEA sample participation standards.

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 set 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 ($r_{u,j}^2$) as:

$$r_{u,j}^2 = r_n^2 - r_{n-m}^2 \quad ,$$

Here, r_n^2 is the R square for the model and r_{n-m}^2 is the R square for the regression model without the m variables in predictor block j .

We then expressed the unique contribution of predictor set j to the explained variance in the predictor variable Y in percentages:

$$UVC_j = r_{u,j}^2 \times 100 \quad .$$

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:

$$JVC_j = (r_n^2 \times 100) - \sum_{k=1}^k UVC_j$$

Table 13.6: ICCS 2022 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 analysis	Unweighted percentage of students in analysis	Weighted percentage of students in analysis
Bulgaria	3113	2519	81	80	2516	81	79
Chinese Taipei	3852	3695	96	96	3722	97	97
Colombia	4193	3525	84	82	3526	84	82
Croatia [‡]	2759	2576	93	94	2572	93	93
Cyprus	3136	2544	81	81	2571	82	82
Estonia	2685	2523	94	94	2533	94	94
France	3533	2951	84	83	2955	84	83
Italy	2945	2731	93	90	2737	93	91
Latvia [‡]	2876	2521	88	88	2527	88	88
Lithuania	3560	3241	91	92	3262	92	93
Malta	2720	2347	86	86	2356	87	86
Netherlands [‡]	2609	2229	85	85	2225	85	84
Norway (9) ¹	5673	4723	83	83	4710	83	83
Poland	4434	4156	94	94	4149	94	94
Romania	2768	2570	93	93	2566	93	93
Serbia	2640	2270	86	86	2277	86	87
Slovak Republic	3202	2888	90	89	2905	91	90
Slovenia	3466	3192	92	92	3200	92	93
Spain	3487	2967	85	84	2978	85	85
Sweden ¹	3263	2899	89	89	2904	89	89
Countries not meeting sample participation requirements							
Brazil	5885	4803	82	82	4805	82	82
Denmark	4769	3542	74	75	3539	74	75
German benchmarking participant meeting sample participation requirements							
North Rhine-Westphalia	3266	2523	77	80	2529	77	80
German benchmarking participant not meeting sample participation requirements							
Schleswig-Holstein	1487	1285	86	85	1293	87	86

Notes:

(9) Country deviated from international defined population and surveyed adjacent upper grade

[‡] Nearly met guidelines for sampling participation rates only after replacement schools were included¹ National defined population covers 90% to 95% of national target population

References

- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Damiani, V. Losito, B., Agrusti, G., Schulz, W. (2024). *Young citizens' views and engagement in a changing Europe. IEA international civic and citizenship education study 2022 European report*. International Association for the Evaluation of Educational Achievement (IEA).
- Gebhardt, E. & Schulz, W. (2018). Scaling procedures for ICCS 2016 test items. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.). *ICCS 2016 technical report* (pp. 117–137). International Association for the Evaluation of Educational Achievement (IEA).
- 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*. Boston College.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables. User's guide. (Version 7)*. Muthén & Muthén.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (Third edition). Harcourt Brace College Publishers
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage 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 2016 technical report* (pp. 261–280). International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W. (2018). The reporting of ICCS 2016 results. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.). *ICCS 2016 technical report* (pp. 245–256). International Association for the Evaluation of Educational Achievement (IEA).
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., Friedman, T., & Damiani, V. (2024). *Education for citizenship in times of global challenge. IEA International Civic and Citizenship Education Study 2022 international report*. Springer.
- Schulz, W. & Friedman, T. (2018). Scaling procedures for ICCS 2016 questionnaire items. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.), *ICCS 2016 technical report* (pp. 139–243). International Association for the Evaluation of Educational Achievement (IEA).
- Strainer, D. L. (2015). Best (but oft-forgotten) practices: the multiple problems of multiplicity—whether and how to correct for many statistical tests. *The American Journal of Clinical Nutrition*, 102(4), 721–728. <https://doi.org/10.3945/ajcn.115.113548>
- Westat (2007). *WesVar® 4.3: User's guide* [computer software]. Rockville, MD: Author.
- Wolter, K. M. (1985). *Introduction to variance estimation*. Springer

Appendix A

Organizations and individuals involved in ICCS 2022

International Study Center

The international study center is located at the Australian Council for Educational Research (ACER). ACER were responsible for designing and implementing the study in close cooperation with LPS (Laboratorio di Pedagogia Sperimentale at the Roma Tre University, Rome) and LUMSA University of Rome, and IEA.

Staff at ACER

Wolfram Schulz, *international study director*

Tim Friedman, *project coordinator*

John Ainley, *project researcher*

Dulce Lay, *data analyst*

Greg Macaskill, *data analyst*

Judy Nixon, *test development*

Laila Helou, *project researcher*

Naoko Tabata, *project researcher*

Staff at LPS/LUMSA University of Rome

Bruno Losito, *associate research co-director*

Gabriella Agrusti, *associate research co-director*

Valeria Damiani, *project researcher*

Carlo Di Chiacchio, *data analyst*

Elisa Caponera, *data analyst*

Laura Palmerio, *data analyst*

International Association for the Evaluation of Educational Achievement (IEA)

IEA provides overall support for the coordination of ICCS 2022 from both the Amsterdam and Hamburg offices. Staff at IEA Amsterdam are responsible for the coordination of translation verification, quality control monitoring, and the publication and wider dissemination of the report. Staff at IEA Hamburg are responsible for the coordination of sampling procedures, and data management and processing.

Staff at IEA Amsterdam

Julian Fraillon, *coordinator of test development*

Dirk Hastedt, *executive director*

Andrea Netten, *director at IEA Amsterdam*

Jan-Peter Broek, *financial manager*

Jan-Philipp Wagner, *research officer*

Jasmin Schiffer, *graphic designer*

Kateřina Hartmanová, *junior research officer*

Katie Hill, *head of communications*

Lauren Musu, *head of translation, verification, and international quality control (TIQ)*

Philippa Elliott, *publications manager*

Staff at IEA Hamburg

Eva Feron, *head of international studies unit*
 Falk Brese, *ICCS international data manager (since 2022)*
 Yasin Afana, *ICCS international data manager (since 2022)*
 Hannah Kowolik, *ICCS international data manager (through 2022)*
 Christine Busch, *ICCS deputy international data manager (through 2022)*
 Diego Cortes, *researcher (sampling)*
 Umut Atasever, *researcher (sampling)*
 Ralph Carstens, *senior research advisor*

ICCS 2022 project advisory committee (PAC)

The ICCS 2022 PAC has, from the beginning of the project, advised the international study center and its partner institutions during regular meetings.

Babara Malak-Minkiewicz, *IEA Amsterdam (retired), Netherlands*
 Cristián Cox, *Diego Portales University, Chile*
 Erik Amnå, *Örebro University, Sweden*
 Judith Torney-Purta, *University of Maryland, United States*
 Wiel Veugelers, *The University of Humanistic Studies Utrecht, Netherlands*

Other project advisors

ICCS 2022 sampling referee

Marc Joncas is the sampling referee for the study, providing invaluable advice on all sampling-related aspects of the study.

Expert consultant

Christian Monseur (*University of Liège*) conducted a review of link items and mode effects for cognitive test items. He provided support and invaluable advice for the implementation of equating and mode effect adjustment procedures for the cognitive data of the ICCS 2022 main survey.

ICCS 2022 National Research Coordinators (NRCs)

The national research coordinators (NRCs) played a crucial role in the study's development. They provided policy- and content-oriented advice on developing the instruments and were responsible for the implementation of ICCS 2022 in the participating countries.

Brazil

Maribel Alves Fierro Sevilla
The Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)

Bulgaria

Natalia Vassileva
Center for Control and Assessment of the Quality in School Education

Chinese Taipei

Meihui Liu
National Taiwan Normal University (NTNU)

Colombia

Natalia González Gómez
Colombian Institute for the Assessment of Education (ICFES)

Croatia

Ines Elezović

Department for Quality Assurance in Education, National Centre for External Evaluation of Education

Cyprus

Yiasemina Karagiorgi

Centre for Educational Research and Evaluation

Denmark

Jens Bruun

Danish School of Education, Aarhus University

Estonia

Meril Ümarik

Tallin University

France

Elodie Persem

Ministry of National Education

Germany (North Rhine-Westphalia and Schleswig-Holstein)

Hermann Josef Abs

University of Duisburg-Essen

Katrin Hahn-Laudenberg

Bergische Universität Wuppertal

Italy

Laura Palmerio

INVALSI

Latvia

Ireta Čekse

University of Latvia

Lithuania

Lina Pareigienė

National Agency for Education

Malta

Jude Zammit

Louis Scerri

Ministry for Education and Employment

Netherlands

Remmert Daas

University of Amsterdam

Norway

Oddveig Storstad

Department of Teacher Education, Norwegian University of Science and Technology (NTNU)

Poland

Olga Wasilewska

Educational Research Institute (IBE)

Romania

Catalina Ulrich

University of Bucharest

Serbia

Tanja Trbojević

Institute for Education quality and Evaluation

Slovak Republic

Gabriella Kopas

Kristina Čevorová

National Institute for Certified Educational Measurements

Slovenia

Eva Klemenčič-Mirazchiyski

Educational Research Institute

Spain

Ricardo Callejo Martínez

Gala Ríos Junquera

INEE (Instituto Nacional de Evaluación Educativa)

Sweden

Maria Axelsson

Ellen Almgren

Swedish National Agency for Education

Appendix B

Characteristics of National Samples

This appendix describes, for each education system participating in ICCS 2022, the population coverage, exclusion categories, stratification variables, and any deviations from the general ICCS 2022 sampling design.

The same sample of schools was selected for the student survey and the teacher survey. However, the school participation status of a school in the student and teacher survey could differ. It was particularly common for a school to count as participating in the student survey, but not in the teacher survey; however, the reverse scenario was also possible. Finally, this appendix also summarizes sampling outcomes for countries participating in the Bridge Study.

B.1. Bulgaria - Core and Bridge Samples

- Coverage is 100.
- School level exclusions consisted of special needs schools and very small schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type for a total of three explicit strata.
- Implicit stratification by location for a total of three implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.1.1: Allocation of core student sample in Bulgaria

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Profiled Schools	33	0	28	5	0	0	0
Vocational Schools	64	0	54	6	0	4	0
Other Schools	53	1	44	7	0	1	0
Total	150	1	126	18	0	5	0

Note: Four schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.1.2: Allocation of teacher sample in Bulgaria

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Profiled Schools	33	0	26	5	0	2	0
Vocational Schools	64	0	52	7	0	5	0
Other Schools	53	1	43	7	0	2	0
Total	150	1	121	19	0	9	0

Note: Six schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.1.3: Allocation of bridge student sample in Bulgaria

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Profiled Schools	14	0	9	1	0	4	0
Vocational Schools	28	0	17	7	0	4	0
Other Schools	22	0	16	5	0	1	0
Total	64	0	42	13	0	9	0

B.2. Brazil – Core Sample

- Coverage is 100.
- School level exclusions consisted of special needs schools, very small schools, indigenous schools, and inaccessible schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by region and school type for a total of 10 explicit strata.
- Implicit stratification by SES, location 1 and location 2 for a total of 16 implicit strata.
- Sampled one classroom per school and two classrooms in schools with more than 100 students.
- Small schools sampled with equal probabilities.

Table B.2.1: Allocation of core student sample in Brazil

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
North - Private	8	0	6	2	0	0	0
North - Public	20	0	16	2	0	2	0
Northeast - Private	10	0	4	2	2	2	0
Northeast - Public	50	2	41	3	1	3	0
Southeast - Private	14	0	5	3	0	6	0
Southeast - Public	66	1	54	6	1	4	0
South - Private	8	0	2	1	1	4	0
South - Public	24	0	22	1	0	1	0
Middle-West - Private	8	0	2	3	0	3	0
Middle-West - Public	14	0	8	2	0	4	0
Total	222	3	160	25	5	29	0

Note: Twenty-nine schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.2.2: Allocation of teacher sample in Brazil

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
North - Private	8	0	6	1	0	1	0
North - Public	20	0	17	2	0	1	0
Northeast - Private	10	0	4	1	2	3	0
Northeast - Public	50	2	40	2	1	5	0
Southeast - Private	14	0	4	4	0	6	0
Southeast - Public	66	1	53	5	0	7	0
South - Private	8	0	1	0	0	7	0
South - Public	24	0	19	1	0	4	0
Middle-West - Private	8	0	3	3	1	1	0
Middle-West - Public	14	0	12	2	0	0	0
Total	222	3	159	21	4	35	0

Note: Fifteen schools were regarded as non-participating because the within-school participation rate was below 50%.

B.3. Colombia - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, and foreign language schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type and urbanization, and number of classes for a total of six explicit strata.
- Implicit stratification by gender and administrative unit for a total of 417 implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.3.1: Allocation of core student sample in Colombia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Large - Public Urban	44	0	43	1	0	0	0
Large - Public Rural	4	0	4	0	0	0	0
Large - Private	6	0	6	0	0	0	0
Small - Public Urban	48	0	46	2	0	0	0
Small - Public Rural	25	0	19	3	1	2	0
Small - Private	23	0	21	1	1	0	0
Total	150	0	139	7	2	2	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.3.2: Allocation of teacher sample in Colombia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Large - Public Urban	44	0	37	0	0	7	0
Large - Public Rural	4	0	3	0	0	1	0
Large - Private	6	0	6	0	0	0	0
Small - Public Urban	48	0	35	2	0	11	0
Small - Public Rural	25	0	15	1	1	8	0
Small - Private	23	0	18	1	1	3	0
Total	150	0	114	4	2	30	0

Note: Eighteen schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.3.3: Allocation of bridge student sample in Colombia

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Large - Public Urban	16	0	16	0	0	0	0
Large - Public Rural	4	0	3	0	0	1	0
Large - Private	4	0	4	0	0	0	0
Small - Public Urban	12	0	11	1	0	0	0
Small - Public Rural	7	0	7	0	0	0	0
Small - Private	5	0	5	0	0	0	0
Total	48	0	46	1	0	1	0

Note: One schools was regarded as non-participating because the within-school participation rate was below 50%.

B.4. Chinese Taipei - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, pupil referral units, and international schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by number of classes, region, and school type for a total of 10 explicit strata.
- Implicit stratification by urbanization for a total of four implicit strata
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.4.1: Allocation of core student sample in Chinese Taipei

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public - North - Small	5	0	5	0	0	0	0
Public - North - Large	51	0	50	1	0	0	0
Public - Middle - Small	4	0	4	0	0	0	0
Public - Middle - Large	29	0	27	0	0	2	0
Public - South - Small	6	0	5	1	0	0	0
Public - South - Large	29	0	29	0	0	0	0
Public - East - Small	2	0	2	0	0	0	0
Public - East - Large	4	0	4	0	0	0	0
Private - All - Small	4	0	4	0	0	0	0
Private - All - Large	16	0	16	0	0	0	0
Total	150	0	146	2	0	2	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

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	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public - North - Small	5	0	5	0	0	0	0
Public - North - Large	51	0	49	1	0	1	0
Public - Middle - Small	4	0	4	0	0	0	0
Public - Middle - Large	29	0	29	0	0	0	0
Public - South - Small	6	0	5	1	0	0	0
Public - South - Large	29	0	29	0	0	0	0
Public - East - Small	2	0	2	0	0	0	0
Public - East - Large	4	0	4	0	0	0	0
Private - All - Small	4	0	4	0	0	0	0
Private - All - Large	16	0	16	0	0	0	0
Total	150	0	147	2	0	1	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.4.3: Allocation of bridge student sample in Chinese Taipei

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public - North - Small	2	0	2	0	0	0	0
Public - North - Large	20	0	19	1	0	0	0
Public - Middle - Small	2	0	2	0	0	0	0
Public - Middle - Large	11	0	11	0	0	0	0
Public - South - Small	2	0	2	0	0	0	0
Public - South - Large	10	0	10	0	0	0	0
Public - East - Small	1	0	1	0	0	0	0
Public - East - Large	1	0	1	0	0	0	0
Private - All - Small	2	0	2	0	0	0	0
Private - All - Large	4	0	4	0	0	0	0
Total	55	0	54	1	0	0	0

B.5. Croatia – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, minority schools, and private schools.
- Within-sample exclusions consist of students being excluded during data collection
- Explicit stratification by region for a total of six explicit strata.
- Implicit stratification by county for a total of 21 implicit strata.
- Sampled one classroom per school and two classrooms in schools with more than 90 students.
- Small schools sampled with equal probabilities.

Table B.5.1: Allocation of core student sample in Croatia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Central Croatia	32	0	30	1	0	1	0
Eastern Croatia	24	0	23	0	0	1	0
Northern Croatia	16	0	14	1	0	1	0
Western Croatia	18	0	18	0	0	0	0
Southern Croatia	32	0	30	0	0	2	0
City of Zagreb	28	0	24	0	0	4	0
Total	150	0	139	2	0	9	0

Note: Nine schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.5.2: Allocation of teacher sample in Croatia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Central Croatia	32	0	31	1	0	0	0
Eastern Croatia	24	0	24	0	0	0	0
Northern Croatia	16	0	15	1	0	0	0
Western Croatia	18	0	18	0	0	0	0
Southern Croatia	32	0	31	0	0	1	0
City of Zagreb	28	0	27	0	0	1	0
Total	150	0	146	2	0	2	0

Note: Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.6. Cyprus – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, schools in the Turkish occupied area, and language of instruction other than Greek or English.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by type of school and curriculum, and district for a total of five explicit strata.
- Implicit stratification by urbanization and school type for a total of four implicit strata.
- Sampled one classroom per school and two classrooms in schools with more than 70 students.
- All schools are sampled with certainty.

Table B.6.1: Allocation of core student sample in Cyprus

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public and Private schools with national curriculum - Larnaca/Famagusta	15	0	15	0	0	0	0
Public and Private schools with national curriculum - Limassol	23	0	23	0	0	0	0
Public and Private schools with national curriculum - Nicosia	29	0	28	0	0	1	0
Public and Private schools with national curriculum - Paphos	8	0	8	0	0	0	0
Private schools with other curriculum - All	23	0	21	0	0	2	0
Total	98	0	95	0	0	3	0

Table B.6.2: Allocation of teacher sample in Cyprus

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public and Private schools with national curriculum - Larnaca/Famagusta	15	0	13	0	0	2	0
Public and Private schools with national curriculum - Limassol	23	0	23	0	0	0	0
Public and Private schools with national curriculum - Nicosia	29	0	24	0	0	5	0
Public and Private schools with national curriculum - Paphos	8	0	7	0	0	1	0
Private schools with other curriculum - All	23	0	18	0	0	5	0
Total	98	0	85	0	0	13	0

Note: Forty-nine schools were regarded as non-participating because the within-school participation rate was below 50%.

B.7. Denmark – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, international schools, and very small schools.
- Within-sample exclusions consist of students being excluded during data collection.
- No Explicit stratification.
- Implicit stratification by area code for a total of 12 implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.7.1: Allocation of core student sample in Denmark

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Denmark	216	0	64	43	27	82	0
Total	216	0	64	43	27	82	0

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	216	0	32	25	13	146	0
Total	216	0	32	25	13	146	0

Note: Forty-nine schools were regarded as non-participating because the within-school participation rate was below 50%.

B.8. Estonia - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools and different curriculum schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by language of instruction for a total of two explicit strata.
- No Implicit stratification.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.8.1: Allocation of core student sample in Estonia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Estonian language school	157	0	130	0	0	25	2
Russian language school	45	1	36	0	0	8	0
Total	202	1	166	0	0	33	2

Note: Two schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.8.2: Allocation of teacher sample in Estonia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Estonian language school	157	0	112	0	0	43	2
Russian language school	45	1	33	0	0	11	0
Total	202	1	145	0	0	54	2

Note: Sixteen schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.8.3: Allocation of bridge student sample in Estonia

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Estonian language school	89	0	80	0	0	9	0
Russian language school	8	0	6	0	0	2	0
Total	97	0	86	0	0	11	0

Note: Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.9. France – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, overseas territories, geographically inaccessible schools, private schools without a contract, and Mayotte schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by type of school and priority level for a total of three explicit strata.
- No Implicit stratification.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.9.1: Allocation of core student sample in France

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public schools outside priority education	94	0	91	0	0	3	0
Public schools in priority education	24	0	24	0	0	0	0
Private schools	32	0	31	0	1	0	0
Total	150	0	146	0	1	3	0

Table B.9.2: Allocation of teacher sample in France

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Public schools outside priority education	94	0	60	0	0	34	0
Public schools in priority education	24	0	16	0	0	8	0
Private schools	32	0	20	0	1	11	0
Total	150	0	96	0	1	53	0

Note: Forty-one schools were regarded as non-participating because the within-school participation rate was below 50%.

B.10. Italy - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, students taught in Slovenian, students taught in Ladin, and private schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by geographic area for a total of five explicit strata.
- Implicit stratification by funding for a total of two implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.10.1: Allocation of core student sample in Italy

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
North West	42	0	39	2	0	1	0
North East	30	0	29	0	0	1	0
Center	30	0	30	0	0	0	0
South	32	0	32	0	0	0	0
South Islands	26	0	26	0	0	0	0
Total	160	0	156	2	0	2	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.10.2: Allocation of teacher sample in Italy

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
North West	42	0	40	2	0	0	0
North East	30	0	29	0	0	1	0
Center	30	0	28	0	0	2	0
South	32	0	32	0	0	0	0
South Islands	26	0	26	0	0	0	0
Total	160	0	155	2	0	3	0

Table B.10.3: Allocation of bridge student sample in Italy

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
North West	18	0	18	0	0	0	0
North East	14	0	12	2	0	0	0
Center	14	0	13	1	0	0	0
South	14	0	14	0	0	0	0
South Islands	12	0	11	1	0	0	0
Total	72	0	68	4	0	0	0

B.11. Lithuania - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, schools teaching in other language, and prison and reformatory schools.
- Within-sample exclusions consist of students being excluded during data collection
- Explicit stratification by language for a total of four explicit strata.
- Implicit stratification by urbanization and school type for a total of 13 implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.11.1: Allocation of core student sample in Lithuania

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Lithuanian only	130	4	123	0	0	3	0
Russian only	15	0	15	0	0	0	0
Polish only	24	0	24	0	0	0	0
Multilingual	15	0	15	0	0	0	0
Total	184	4	177	0	0	3	0

Table B.11.2: Allocation of teacher sample in Lithuania

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Lithuanian only	130	4	123	0	0	3	0
Russian only	15	0	15	0	0	0	0
Polish only	24	0	24	0	0	0	0
Multilingual	15	0	14	0	0	1	0
Total	184	4	176	0	0	4	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.11.3: Allocation of bridge student sample in Lithuania

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Lithuanian only	60	1	59	0	0	0	0
Russian only	3	0	3	0	0	0	0
Polish only	4	0	4	0	0	0	0
Multilingual	3	0	3	0	0	0	0
Total	70	1	69	0	0	0	0

B.12. Latvia - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, distance education schools, language of instruction differ from Latvian or Russian and special regime (criminal) schools.
- Within-sample exclusions consist of students being excluded during data collection
- Explicit stratification by language for a total of five explicit strata.
- Implicit stratification by school type and urbanization for a total of 10 implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.12.1: Allocation of core student sample in Latvia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Latvian - Gymnasium	18	0	17	0	0	1	0
Latvian - Basic	27	0	26	1	0	0	0
Latvian - Secondary	60	0	58	2	0	0	0
Russian - ALL	37	0	33	1	0	3	0
Latvian & Russian - ALL	10	0	8	1	0	1	0
Total	152	0	142	5	0	5	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.12.2: Allocation of teacher sample in Latvia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Latvian - Gymnasium	18	0	16	0	0	2	0
Latvian - Basic	27	0	23	1	0	3	0
Latvian - Secondary	60	0	47	1	0	12	0
Russian - ALL	37	0	24	0	0	13	0
Latvian & Russian - ALL	10	0	6	0	0	4	0
Total	152	0	116	2	0	34	0

Note: Twenty-five schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.12.3: Allocation of bridge student sample in Latvia

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Latvian - Gymnasium	6	0	6	0	0	0	0
Latvian - Basic	9	0	9	0	0	0	0
Latvian - Secondary	20	0	18	1	0	1	0
Russian - ALL	13	0	10	0	0	3	0
Latvian & Russian - ALL	4	0	4	0	0	0	0
Total	52	0	47	1	0	4	0

Note: Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.13. Malta - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools and language schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by geographic sector for a total of three explicit strata.
- No Implicit stratification.
- Sampled all (15) classrooms per school.
- Small schools sampled with equal probabilities.

Table B.13.1: Allocation of core student sample in Malta

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
State Schools	10	0	10	0	0	0	0
Church Schools	14	0	14	0	0	0	0
Independent Schools	5	0	5	0	0	0	0
Total	29	0	29	0	0	0	0

Table B.13.2: Allocation of teacher sample in Malta

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
State Schools	10	0	9	0	0	1	0
Church Schools	14	0	13	0	0	1	0
Independent Schools	5	0	4	0	0	1	0
Total	29	0	26	0	0	3	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.13.3: Allocation of bridge student sample in Malta

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
State Schools	4	0	4	0	0	0	0
Church Schools	7	0	7	0	0	0	0
Independent Schools	2	0	2	0	0	0	0
Total	13	0	13	0	0	0	0

B.14. the Netherlands – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, and international schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type for a total of three explicit strata.
- No Implicit stratification.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.14.1: Allocation of core student sample in the Netherlands

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Mixed	14	0	4	3	2	5	0
General	70	0	44	8	7	11	0
Vocational	66	0	38	7	11	9	1
Total	150	0	86	18	20	25	1

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 the Netherlands

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Mixed	14	0	4	2	0	8	0
General	70	0	31	4	7	28	0
Vocational	66	0	30	6	8	22	0
Total	150	0	65	12	15	58	0

Note: Thirteen schools were regarded as non-participating because the within-school participation rate was below 50%.

B.15. Norway (9) - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, and other language schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type and language for a total of three explicit strata.
- Implicit stratification by region for a total of eight implicit strata.
- Sampled two classrooms per school.
- Small schools sampled with equal probabilities.

Table B.15.1: Allocation of core student sample in Norway (9)

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Bokmål - Private	24	0	21	0	0	3	0
Bokmål - Public	100	1	98	0	0	1	0
Nynorsk - All	25	1	23	0	0	1	0
Total	149	2	142	0	0	5	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.15.2: Allocation of teacher sample in Norway (9)

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Bokmål - Private	24	0	18	0	0	6	0
Bokmål - Public	100	1	84	0	0	15	0
Nynorsk - All	25	1	21	0	0	3	0
Total	149	2	123	0	0	24	0

Note: Fourteen schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.15.3: Allocation of bridge student sample in Norway (9)

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Bokmål - Private	5	0	5	0	0	0	0
Bokmål - Public	40	0	39	0	0	1	0
Nynorsk - All	5	0	5	0	0	0	0
Total	50	0	49	0	0	1	0

B.16. Poland – Core Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, and schools with language of instruction other than Polish.
- Within-sample exclusions consist of students being excluded during data collection
- Explicit stratification by number of inhabitants for a total of four explicit strata.
- No Implicit stratification.
- Sampled one classroom per school and two classrooms in schools with more than 80 students.
- Small schools sampled with equal probabilities.

Table B.16.1: Allocation of core student sample in Poland

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Village	62	0	56	5	1	0	0
Town to 20k Inhabitants	30	0	30	0	0	0	0
Town to 100k Inhabitants	36	0	32	4	0	0	0
Town over 100k Inhabitants	42	0	37	2	2	1	0
Total	170	0	155	11	3	1	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.16.2: Allocation of teacher sample in Poland

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Village	62	0	56	5	1	0	0
Town to 20k Inhabitants	30	0	30	0	0	0	0
Town to 100k Inhabitants	36	0	31	4	0	1	0
Town over 100k Inhabitants	42	0	37	2	1	2	0
Total	170	0	154	11	2	3	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

B.17. Romania – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools and very small schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by region and urbanization for a total of 10 explicit strata.
- Implicit stratification by school average score category for a total of four implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.17.1: Allocation of core student sample in Romania

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Center - Urban	12	0	11	1	0	0	0
Center - Rural	8	0	7	1	0	0	0
Bucharest-Ilfov - Urban	14	0	11	2	0	1	0
Bucharest-Ilfov - Rural	8	0	7	0	0	1	0
North - Urban	24	0	17	5	1	1	0
North - Rural	22	0	19	3	0	0	0
South - Urban	28	0	25	2	0	1	0
South - Rural	26	0	20	5	1	0	0
West - Urban	8	0	8	0	0	0	0
West - Rural	8	0	6	2	0	0	0
Total	158	0	131	21	2	4	0

Table B.17.2: Allocation of teacher sample in Romania

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Center - Urban	12	0	11	1	0	0	0
Center - Rural	8	0	7	1	0	0	0
Bucharest-Ilfov - Urban	14	0	10	2	0	2	0
Bucharest-Ilfov - Rural	8	0	7	0	0	1	0
North - Urban	24	0	18	4	1	1	0
North - Rural	22	0	19	3	0	0	0
South - Urban	28	0	25	2	0	1	0
South - Rural	26	0	20	5	1	0	0
West - Urban	8	0	8	0	0	0	0
West - Rural	8	0	6	2	0	0	0
Total	158	0	131	20	2	5	0

Note: Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.18. Serbia – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools and very small schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by region, urbanization, and school hierarchy for a total of seven explicit strata
- No Implicit stratification.
- Sampled one classroom per school and two classrooms in schools with more than 170 students.
- Small schools sampled with equal probabilities.

Table B.18.1: Allocation of core student sample in Serbia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Belgrade - City - All Schools	28	0	22	1	0	5	0
Belgrade - Other - All Schools	8	0	5	0	0	3	0
Vojvodina - City - All Schools	24	0	23	0	0	1	0
Vojvodina - Other - All Schools	14	0	11	1	0	2	0
Central Serbia - City - All Schools	48	0	48	0	0	0	0
Central Serbia - Other - Main School	20	0	16	3	0	1	0
Central Serbia - Other - Branch Department	8	0	8	0	0	0	0
Total	150	0	133	5	0	12	0

Note: Fourteen schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.18.2: Allocation of teacher sample in Serbia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Belgrade - City - All Schools	28	0	27	1	0	0	0
Belgrade - Other - All Schools	8	0	8	0	0	0	0
Vojvodina - City - All Schools	24	0	24	0	0	0	0
Vojvodina - Other - All Schools	14	0	13	1	0	0	0
Central Serbia - City - All Schools	48	0	48	0	0	0	0
Central Serbia - Other - Main School	20	0	17	3	0	0	0
Central Serbia - Other - Branch Department	8	0	8	0	0	0	0
Total	150	0	145	5	0	0	0

B.19. Slovak Republic – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, and schools with a language of instruction other than Slovak or Hungarian.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type and language for a total of four explicit strata.
- Implicit stratification by national math score and region for a total of 27 implicit strata.
- Sampled one classroom per school and two classrooms in schools with more than 75 students.
- Small schools sampled with equal probabilities.

Table B.19.1: Allocation of core student sample in Slovak Republic

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Primary School - Slovak	130	0	122	5	0	3	0
Primary School - Hungarian	10	0	10	0	0	0	0
Gymnasium - All	10	0	9	1	0	0	0
SEN Elementary School - All	10	0	6	1	0	3	0
Total	160	0	147	7	0	6	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.19.2: Allocation of teacher sample in Slovak Republic

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Primary School - Slovak	130	0	122	5	0	3	0
Primary School - Hungarian	10	0	10	0	0	0	0
Gymnasium - All	10	0	9	1	0	0	0
SEN Elementary School - All	10	0	6	0	0	4	0
Total	160	0	147	6	0	7	0

Note: Two schools were regarded as non-participating because the within-school participation rate was below 50%.

B.20. Slovenia - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, private schools/different pedagogy, and Italian and bilingual schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by number of classes for a total of two explicit strata.
- No Implicit stratification.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.20.1: Allocation of core student sample in Slovenia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Slovenia - Small	20	0	18	1	0	1	0
Slovenia - Large	154	0	139	10	0	5	0
Total	174	0	157	11	0	6	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.20.2: Allocation of teacher sample in Slovenia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Slovenia - Small	20	0	19	1	0	0	0
Slovenia - Large	154	0	138	10	0	6	0
Total	174	0	157	11	0	6	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.20.3: Allocation of bridge student sample in Slovenia

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Slovenia - Small	8	0	8	0	0	0	0
Slovenia - Large	70	0	62	3	0	5	0
Total	78	0	70	3	0	5	0

B.21. Spain – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, and other curriculum or school structure.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by autonomous community for a total of seven explicit strata.
- Implicit stratification by school type and region for a total of 32 implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.21.1: Allocation of core student sample in Spain

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Andalusia	32	0	32	0	0	0	0
Catalonia	24	0	24	0	0	0	0
Community of Madrid	22	0	22	0	0	0	0
Valencian Community	18	0	15	1	0	2	0
Medium Size Autonomous Communities	38	0	37	1	0	0	0
Small Size Autonomous Communities	18	0	17	0	0	1	0
Very Small Size Autonomous Communities	8	0	8	0	0	0	0
Total	160	0	155	2	0	3	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.21.2: Allocation of teacher sample in Spain

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Andalusia	32	0	32	0	0	0	0
Catalonia	24	0	24	0	0	0	0
Community of Madrid	22	0	22	0	0	0	0
Valencian Community	18	0	17	1	0	0	0
Medium Size Autonomous Communities	38	0	36	1	0	1	0
Small Size Autonomous Communities	18	0	17	0	0	1	0
Very Small Size Autonomous Communities	8	0	7	0	0	1	0
Total	160	0	155	2	0	3	0

Note: Three schools were regarded as non-participating because the within-school participation rate was below 50%.

B.22. Sweden - Core and Bridge Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools, very small schools, special schools, and international schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by score for a total of four explicit strata.
- No Implicit stratification.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.22.1: Allocation of core student sample in Sweden

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Missing	9	0	9	0	0	0	0
Low	12	0	10	0	0	2	0
Medium	56	0	53	2	0	1	0
High	77	0	73	2	0	2	0
Total	154	0	145	4	0	5	0

Table B.22.2: Allocation of teacher sample in Sweden

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Missing	9	0	6	0	0	3	0
Low	12	0	10	0	0	2	0
Medium	56	0	48	1	0	7	0
High	77	0	64	2	0	11	0
Total	154	0	128	3	0	23	0

Note: Seventeen schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.22.3: Allocation of bridge student sample in Sweden

School Participation Status—Bridge Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Missing	4	0	3	1	0	0	0
Low	5	0	4	0	0	1	0
Medium	25	0	24	0	0	1	0
High	35	0	32	2	0	1	0
Total	69	0	63	3	0	3	0

B.23. North Rhine-Westphalia (Germany) – Core Sample

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools and Waldorf schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type and migration for a total of two explicit strata.
- Implicit stratification by migration for a total of four implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.23.1: Allocation of core student sample in North Rhine-Westphalia

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Gymnasium - Low proportion of immigrants	21	0	20	1	0	0	0
Gymnasium - High proportion of immigrants	21	0	18	2	0	1	0
Hauptschule - Low proportion of immigrants	8	0	7	0	0	1	0
Hauptschule - High proportion of immigrants	10	0	10	0	0	0	0
Realschule - Low proportion of immigrants	10	0	8	1	0	1	0
Realschule - High proportion of immigrants	19	0	17	0	0	2	0
Integrierte Gesamtschule - Low proportion of immigrants	10	0	9	1	0	0	0
Integrierte Gesamtschule - High proportion of immigrants	27	0	24	1	0	2	0
Förderschule - All	11	0	8	3	0	0	0
Schule mit mehreren Bildungsgängen - Low proportion of immigrants	7	0	7	0	0	0	0
Schule mit mehreren Bildungsgängen - High proportion of immigrants	8	0	8	0	0	0	0
Total	152	0	136	9	0	7	0

Note: One school was regarded as non-participating because the within-school participation rate was below 50%.

Table B.23.2: Allocation of teacher sample in North Rhine-Westphalia

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Gymnasium - Low proportion of immigrants	21	0	18	1	0	2	0
Gymnasium - High proportion of immigrants	21	0	19	2	0	0	0
Hauptschule - Low proportion of immigrants	8	0	5	0	0	3	0
Hauptschule - High proportion of immigrants	10	0	9	0	0	1	0
Realschule - Low proportion of immigrants	10	0	7	1	0	2	0
Realschule - High proportion of immigrants	19	0	14	0	0	5	0
Integrierte Gesamtschule - Low proportion of immigrants	10	0	9	1	0	0	0
Integrierte Gesamtschule - High proportion of immigrants	27	0	20	1	0	6	0
Förderschule - All	11	0	8	3	0	0	0
Schule mit mehreren Bildungsgängen - Low proportion of immigrants	7	0	5	0	0	2	0
Schule mit mehreren Bildungsgängen - High proportion of immigrants	8	0	6	0	0	2	0
Total	152	0	120	9	0	23	0

Note: Seventeen schools were regarded as non-participating because the within-school participation rate was below 50%.

B.24. Schleswig-Holstein (Germany) – Core Samples

- Coverage is 100 percent.
- School level exclusions consisted of special needs schools.
- Within-sample exclusions consist of students being excluded during data collection.
- Explicit stratification by school type and migration for a total of four explicit strata.
- Implicit stratification by migration for a total of three implicit strata.
- Sampled one classroom per school.
- Small schools sampled with equal probabilities.

Table B.24.1: Allocation of core student sample in Schleswig-Holstein

School Participation Status—Core Student Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Gymnasium - All	37	0	30	5	0	2	0
Integrierte Gesamtschule - Low proportion of immigrants	58	0	28	9	1	20	0
Integrierte Gesamtschule - High proportion of immigrants	11	1	3	3	1	3	0
Förderschule - All	14	5	4	0	0	5	0
Total	120	6	65	17	2	30	0

Note: Seven schools were regarded as non-participating because the within-school participation rate was below 50%.

Table B.24.2: Allocation of teacher sample in Schleswig-Holstein

School Participation Status—Teacher Survey

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools	Excluded Schools
			Sampled	First Replacement	Second Replacement		
Gymnasium - All	37	0	2	2	0	33	0
Integrierte Gesamtschule - Low proportion of immigrants	59	0	5	1	0	53	0
Integrierte Gesamtschule - High proportion of immigrants	11	1	1	1	0	8	0
Förderschule - All	14	5	1	0	0	8	0
Total	121	6	9	4	0	102	0

Note: Seventy-one schools were regarded as non-participating because the within-school participation rate was below 50%.

Appendix C

Civic knowledge test items with associated framework references, scaled item difficulties and proficiency levels

Table C.1: Civic knowledge items used in ICCS 2022

Item	Item label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge item difficulty	Proficiency level
CI3MAM1	Role of military	2016	MC	1	1	398	Level C
CI4JSM1	Jury selection	2022	MC	2	1	536	Level B
CI4JSM2	Jury justification	2022	MC	2	2	463	Level C
CI2WFO2	Argument in favor of requiring people	2009	CR	2	2	595	Level A
CI4SDM1	Living sustainably	2022	MC	2	1	417	Level C
CI4SDM2	Sustainable development learning	2022	MC	2	2	435	Level C
CI4SDM3	Sustainable development cooperation	2022	MC	2	1	462	Level C
CI2ECM1	Most likely reason to ask for advice	2009	MC	3	2	436	Level C
CI2ECM2	Risk by asking for advice	2009	MC	3	2	491	Level B
CI308M1	Equal opportunity employment	2016	MC	2	2	425	Level C
CI4SJM1	Social justice improvement	2022	MC	2	1	498	Level B
CI3RRM1	Representative responsibilities	2016	MC	1	1	397	Level C
CI2REM2	Help for refugees	2009	MC	2	1	485	Level B
CI2REM3	Responsibility of refugees	2009	MC	2	1	405	Level C
CI4REM1	Renewable energy	2022	MC	2	1	461	Level C
CI4FKC1	Fake news	2022	CR	1	2	589	Level A
CI4FKM2	Fact checking	2022	MC	3	1	437	Level C
CI2CCM1	Affection by damage to the environment	2009	MC	2	2	444	Level C
CI2CCM2	What advisor wants people to understand	2009	MC	4	2	519	Level B
CI4EPM1	Online petition advantage	2022	MC	3	2	407	Level C
CI4EPM2	Online petition problem	2022	MC	3	2	463	Level C
CI4CFM1	Carbon emissions	2022	MC	2	1	617	Level A
CI2FSM1	Most likely behavior against law	2009	MC	2	2	480	Level B
CI4TDM1	Treaty example	2022	MC	1	2	440	Level C
CI2PGM1	Way pressure groups contribute democracy	2009	MC	1	2	490	Level B
CI2PGM2	Advantage of independent pressure groups	2009	MC	1	2	600	Level A
CI4FWC2	Charity support	2022	CR	3	2	525	Level B
CI4NDM1	National day celebration	2022	MC	4	2	460	Level C
CI2BPM1	Important find out policies before vote	2009	MC	3	2	484	Level B
CI2BPM2	Use only friends' information	2009	MC	3	2	472	Level C
CI4OCM1	Open court cases	2022	MC	2	2	554	Level B
CI4OCM2	Closed court cases	2022	MC	1	2	516	Level B
CI4NTM1	Good neighbor	2022	MC	4	1	460	Level C

Item	Item label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge item difficulty	Proficiency level
CI3PLM1	Pet leash laws	2016	MC	1	2	350	Level D
CI4ECM1	Ethical consumer	2022	MC	3	1	561	Level B
CI4ECM2	Ethical consumer website	2022	MC	3	1	548	Level B
CI4ECM3	Local products	2022	MC	2	2	547	Level B
CI2FDM1	Adult citizens expected to decide	2009	MC	4	1	456	Level C
CI2WFO1	Reasons giving money unemployed people	2009	CR	2	2	588	Level A
CI2RCM1	Need to introduce new emblems	2009	MC	2	2	467	Level C
CI4TXM1	Taxation	2022	MC	1	1	400	Level C
CI4VRM1	Compulsory voting reason	2022	MC	1	2	517	Level B
CI4VRM2	Voluntary voting reason	2022	MC	1	2	494	Level B
CI3IVM1	Informed voting	2016	MC	3	2	442	Level C
CI4PRM1	Respectful rule	2022	MC	4	1	433	Level C
CI2PRM1	Best reason against violent protest	2009	MC	3	2	574	Level A
CI3CAM1	Art as political expression	2016	MC	2	2	571	Level A
CI4NHC2	Learning from history	2022	CR	2	2	613	Level A
CI4DDM1	Equality information	2022	MC	2	2	483	Level B
CI4DDM2	Equality school information	2022	MC	2	2	384	Level D
CI3LPM1	Online advocacy	2016	MC	3	1	411	Level C
CI4LEM1	Free lecture program	2022	MC	3	2	476	Level C
CI4LEM2	Restricted free speech	2022	MC	2	2	546	Level B
CI3CSM1	School in community	2016	MC	2	2	480	Level B
CI312M1	Democratic government action	2016	MC	1	1	491	Level B
CI4SSM1	Petition action	2022	MC	3	2	468	Level C
CI3PAM1	Political parties	2016	MC	3	2	461	Level C
CI3ULM1	Unjust law	2016	MC	2	2	520	Level B
CI2BIO1	Understanding histories and cultures	2009	CR	4	2	711	Level A
CI4MWM1	Minimum wage example	2022	MC	1	2	465	Level C
CI4MWM2	Minimum wage	2022	MC	1	1	503	Level B
CI4MWM3	Wage consultation	2022	MC	1	2	569	Level A
CI4NWM1	Citizen law enforcement legal	2022	MC	2	1	498	Level B
CI4NWM2	Citizen law enforcement illegal	2022	MC	2	2	312	Level D
CI3DDM1	Debate and discuss	2016	MC	3	2	499	Level B
CI3GMM1	Government and military	2016	MC	1	1	491	Level B
CI4IPM1	Illegal protest example	2022	MC	2	1	471	Level C
CI3NPM1	Reason for rules	2016	MC	1	1	294	Below Level D
CI3NPM2	Benefits of rules	2016	MC	4	1	349	Level D
CI4ERM1	Employee rights	2022	MC	2	1	439	Level C
CI4GLM1	Shipping and carbon emissions	2022	MC	2	1	577	Level A
CI4GCM1	Global citizen definition	2022	MC	4	2	542	Level B
CI4DVC1	Voting decrease	2022	CR	3	2	596	Level A
CI2GLM1	Factory benefit local people	2009	MC	1	2	466	Level C
CI2GLM2	Factory harm local people	2009	MC	1	2	503	Level B
CI3SCM1	Democratic selection	2016	MC	3	1	403	Level C

Item	Item label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge item difficulty	Proficiency level
CI4EAM1	Right to education	2022	MC	2	1	445	Level C
CI4INM1	Respectful discussion	2022	MC	2	1	278	Below Level D
CI2ASM1	Interview help farmers get more money	2009	MC	3	2	511	Level B
CI2ASM2	Farmers seeking	2009	MC	2	2	427	Level C
CI3UHM1	Union help	2016	MC	1	1	456	Level C
CI4RFM1	Refugee protection	2022	MC	4	1	447	Level C
CI3CPO1	Town meeting	2016	CR	3	2	516	Level B
CI3CPO2	Survey people	2016	CR	3	2	568	Level A
CI4PDM1	Political donations	2022	MC	1	2	561	Level B
CI4PDM2	Secret donations	2022	MC	1	2	489	Level B
CI4FRM1	Free press	2022	MC	2	1	498	Level B
CI4RCM1	Children's right	2022	MC	2	1	505	Level B
CI3VGM1	Volunteer group	2016	MC	3	2	339	Level D
CI4NHM1	History and connectedness	2022	MC	4	2	419	Level C
CI4DCM1	Equal under the law	2022	MC	2	1	457	Level C
CI4SLM1	Statelessness	2022	MC	4	2	441	Level C
CI4SLC3	Removal of citizenship	2022	CR	2	2	647	Level A
CI4SLC4	Continuance of citizenship	2022	CR	2	2	669	Level A
CI314M1	Law into own hands	2016	MC	2	1	426	Level C
CI2CEM1	Censorship government responsibility	2009	MC	1	2	417	Level C
CI2CEM2	Best reason for opposing censorship	2009	MC	2	2	560	Level B
CI3EPM1	Informed voters	2016	MC	3	2	441	Level C
CI4CZM1	Citizenship characteristic	2022	MC	4	1	468	Level C
CI2ORM1	Not teach official religion	2009	MC	2	2	452	Level C
CI3LSM1	Legal support	2016	MC	2	2	440	Level C
CI3PRO1	Consult electorate	2016	CR	3	2	601	Level A
CI3SDM1	Secret information	2016	MC	1	2	503	Level B
CI4JCM1	Sustainable living example	2022	MC	2	1	573	Level A
CI4JCM2	Sustainable living improvement	2022	MC	2	1	487	Level B
CI4PIM1	Court interpreter	2022	MC	2	2	427	Level C
CI4RPM1	Minister investigation	2022	MC	1	2	502	Level B
CI4EDM1	Electoral debates	2022	MC	1	2	467	Level C
CI3PEM1	Open budget	2016	MC	1	2	513	Level B
CI3LTM1	Regular elections	2016	MC	3	2	562	Level B
CI3IEM1	International environmental agreement	2016	MC	1	1	460	Level C
CI4FTM1	Fair trade	2022	MC	1	2	572	Level A
CI3MDM1	Gov controlled media	2016	MC	2	2	503	Level B
CI2CNM1	Best reason to live next door	2009	MC	2	2	473	Level C
CI2CNM2	Need to understand in order to accept	2009	MC	4	2	355	Level D
CI4FFM1	Fossil fuel effects	2022	MC	2	1	459	Level C
CI4EDC2	Public debate	2022	CR	2	2	551	Level B
CI4SWM1	Social welfare and equity	2022	MC	2	1	429	Level C

Item	Item label	Year of first use	Format	Content domain	Cognitive domain	ICCS civic knowledge item difficulty	Proficiency level
CI4SCM1	Volunteer honor	2022	MC	3	2	561	Level B
CI4UWM1	Retrain workers	2022	MC	1	2	453	Level C
CI4ZHC1_R	Budget evaluation	2022	CR	3	2	534	Level B
CI4ZHL2_R	Budget allocation	2022	LT	3	2	636	Level A
CI4ZHC3	Volunteer party	2022	CR	3	2	564	Level A
CI4ZHC4	Charity donor	2022	CR	3	2	641	Level A
CI4ZHM5	Charity information	2022	MC	3	2	410	Level C
CI4ZEM1	Fair election rules	2022	MC	3	2	361	Level D
CI4ZEC2_R	Best represent views	2022	CR	3	2	555	Level B
CI4ZED3L_R	Select election rules	2022	DD	3	2	635	Level A
CI4ZEC4	Advantaged candidate	2022	CR	3	2	710	Level A
CI4ZEC5	Rewarding voters	2022	CR	3	2	650	Level A
CI4ZSM1	Members voting	2022	MC	3	2	478	Level C
CI4ZSD2	App and democracy	2022	MC	3	2	784	Level A
CI4ZSL3_R	Voting app: identity	2022	LT	3	2	464	Level C
CI4ZSC4	App setting best	2022	CR	3	2	531	Level B
CI4ZSC5Y	Need for quorum	2022	CR	3	2	604	Level A
CI4ZSC5N	No need for quorum	2022	CR	3	2	663	Level A

Appendix D

Mapping of questionnaire items in ICCS 2022 and 2016

Table D.1: Mapping of student questionnaire items in ICCS 2022 and ICCS 2016

ICCS 2022 question	Variable name(s)	Mapping against ICCS 2016
Q1	IS4G01A, IS4G01B	Modified from ICCS 2016 Q1 (Years are changed)
Q2	IS4G02	Modified from ICCS 2016 Q2 (Optional third category)
Q2O	IS4G02O	Unchanged from ICCS 2016 Q2B
Q3	IS4G03	Unchanged from ICCS 2016 Q3
Q3O	IS4G03OA, IS4G03OB, IS4G03OC, IS4G03OD, IS4G03OE, IS4G03OF, IS4G03OG	Modified from ICCS 2016 Q3B (changed traditional parent/guardians to parent1/parent2)
Q4	IS4G04A, IS4G04B, IS4G04C	Modified from ICCS 2016 Q4 (changed traditional parent/guardians to parent1/parent2)
Q5	IS4G05	Unchanged from ICCS 2016 Q5
Q5O	IS4G05O	New for ICCS 2022
Q6a	IS4G06A	Modified from ICCS 2016 Q6A (changed traditional parent/guardians to parent1/parent2)
Q6b	IS4G06B	Modified from ICCS 2016 Q6B (changed traditional parent/guardians to parent1/parent2)
Q7	IS4G07	Modified from ICCS 2016 Q7 (changed traditional parent/guardians to parent1/parent2)
Q8a	IS4G08A	Modified from ICCS 2016 Q8A (changed traditional parent/guardians to parent1/parent2)
Q8b	IS4G08B	Modified from ICCS 2016 Q8B (changed traditional parent/guardians to parent1/parent2)
Q9	IS4G09	Modified from ICCS 2016 Q9 (changed traditional parent/guardians to parent1/parent2)
Q10	IS4G10	Modified from ICCS 2016 Q11 (change to Instructions)
Q10O	IS4G10OA, IS4G10OB, IS4G10OC	New for ICCS 2022
Q11	IS4G11A, IS4G11B, IS4G11C	Modified from ICCS 2016 Q10 (traditional parent/guardians to parent1/parent2)
Q12	IS4G12A, IS4G12B, IS4G12C, IS4G12D, IS4G12E, IS4G12F	Modified from ICCS 2016 Q14 (change to item Q12B)
Q13	IS4G13A, IS4G13B, IS4G13C, IS4G13D, IS4G13E	New for ICCS 2022 (Q13A was part of ICCS 2016 Q14)
Q14	IS4G14A, IS4G14B, IS4G14C, IS4G14D, IS4G14E, IS4G14F, IS4G14G	Modified from ICCS 2016 Q15 (change to items Q14A, Q14C) and new item (Q14D)
Q15	IS4G15A, IS4G15B, IS4G15C, IS4G15D, IS4G15E, IS4G15F, IS4G15G	Modified from ICCS 2016 Q16 (change to items Q15B, Q15E, Q15F) and new item (Q15G)
Q16	IS4G16A, IS4G16B, IS4G16C, IS4G16D, IS4G16E, IS4G16F	Unchanged from ICCS 2016 Q17
Q17	IS4G17A, IS4G17B, IS4G17C, IS4G17D, IS4G17E, IS4G17F, IS4G17G, IS4G17H, IS4G17I	Modified from ICCS 2016 Q18 (change to stem) and new items (Q17H, Q17I)
Q18	IS4G18A, IS4G18B, IS4G18C, IS4G18D, IS4G18E, IS4G18F, IS4G18G, IS4G18H, IS4G18I	Modified from ICCS 2016 Q19 (change to item Q18C)
Q19	IS4G19A, IS4G19B, IS4G19C, IS4G19D, IS4G19E, IS4G19F	New for ICCS 2022
Q20	IS4G20A, IS4G20B, IS4G20C, IS4G20D, IS4G20E, IS4G20F, IS4G20G, IS4G20H, IS4G20I	New for ICCS 2022

ICCS 2022 question	Variable name(s)	Mapping against ICCS 2016
Q21	IS4G21A, IS4G21B, IS4G21C, IS4G21D, IS4G21E, IS4G21F, IS4G21G	New for ICCS 2022
Q22	IS4G22A, IS4G22B, IS4G22C, IS4G22D, IS4G22E, IS4G22F, IS4G22G, IS4G22H, IS4G22I, IS4G22J, IS4G22K, IS4G22L, IS4G22M	Modified from ICCS 2016 Q23 (new items Q22J, Q22K, Q22L, Q22M)
Q23	IS4G23A, IS4G23B, IS4G23C, IS4G23D, IS4G23E, IS4G23F, IS4G23G, IS4G23H, IS4G23I	New for ICCS 2022
Q24	IS4G24A, IS4G24B, IS4G24C, IS4G24D, IS4G24E	Modified from ICCS 2016 European questionnaire Q4 (new item Q24E)
Q25	IS4G25A, IS4G25B, IS4G25C, IS4G25D, IS4G25E, IS4G25F, IS4G25G	Modified from ICCS 2016 Q24 (new item Q25G)
Q26	IS4G26A, IS4G26B, IS4G26C, IS4G26D, IS4G26E	New for ICCS 2022
Q26O	IS4G26OA, IS4G26OB, IS4G26OC, IS4G26OD, IS4G26OE	Unchanged from ICCS 2016 Q25
Q27	IS4G27A, IS4G27B, IS4G27C, IS4G27D, IS4G27E, IS4G27F, IS4G27G, IS4G27H, IS4G27I, IS4G27J, IS4G27K, IS4G27L, IS4G27M, IS4G27N, IS4G27O, IS4G27P, IS4G27Q	Modified from ICCS 2016 Q26 (change to Q27F, Q27G, Q27H, Q27L) and new items (Q27J, Q27K)
Q28	IS4G28A, IS4G28B, IS4G28C, IS4G28D, IS4G28E, IS4G28F, IS4G28G, IS4G28H, IS4G28I, IS4G28J, IS4G28K	Modified from ICCS 2016 Q28 (new item Q28J)
Q29	IS4G29A, IS4G29B, IS4G29C, IS4G29D, IS4G29E, IS4G29F, IS4G29G	Modified from ICCS 2016 Q29 (change to 29D) and new item (Q29G)
Q30	IS4G30A, IS4G30B, IS4G30C, IS4G30D	Unchanged from ICCS 2016 Q32
Q31	IS4G31A, IS4G31B, IS4G31C, IS4G31D, IS4G31E, IS4G31F, IS4G31G, IS4G31H, IS4G31I, IS4G31J, IS4G31K, IS4G31L, IS4G31M, IS4G31N	Modified from ICCS 2016 Q30 (new items Q31F, Q31G, Q31H, Q31I, Q31J, Q31N)
Q32	IS4G32A, IS4G32B, IS4G32C, IS4G32D, IS4G32E, IS4G32F, IS4G32G, IS4G32H, IS4G32I, IS4G32J, IS4G32K	Unchanged from ICCS 2016 Q31
Q33	IS4G33	Unchanged from ICCS 2016 Q33
Q34	IS4G34	Unchanged from ICCS 2016 Q34
Q35	IS4G35A, IS4G35B, IS4G35C, IS4G35D, IS4G35E, IS4G35F, IS4G35G	Unchanged from ICCS 2016 Q35

Table D.2: Mapping of teacher questionnaire items in ICCS 2022 and ICCS 2016

ICCS 2022 question	Variable name(s)	Mapping against ICCS 2016
TcQ-Q1	IT4G01A, IT4G01B, IT4G01C, IT4G01D, IT4G01E, IT4G01F	Unchanged from ICCS 2016 Q1
TcQ-Q2	IT4G02	Unchanged from ICCS 2016 Q2
TcQ-Q3	IT4G03	Unchanged from ICCS 2016 Q4
TcQ-Q4	IT4G04	Unchanged from ICCS 2016 Q3
TcQ-Q5	IT4G05	Modified from ICCS 2016 Q5 (optional third category added)
TcQ-Q6	IT4G06A, IT4G06B, IT4G06C, IT4G06D, IT4G06E, IT4G06F	Modified from ICCS 2016 (changes to stem and items)
TcQ-Q7	IT4G07A, IT4G07B, IT4G07C, IT4G07D, IT4G07E, IT4G07F, IT4G07G, IT4G07H, IT4G07I	Unchanged from ICCS 2016 Q7
TcQ-Q8	IT4G08A, IT4G08B, IT4G08C, IT4G08D, IT4G08E, IT4G08F, IT4G08G, IT4G08H, IT4G08I, IT4G08J	Modified from ICCS 2016 Q8 (change to items Q8B, Q8F, Q8G) and new item (Q8J)
TcQ-Q9	IT4G09A, IT4G09B, IT4G09C, IT4G09D	Unchanged from ICCS 2016
TcQ-Q10	IT4G10A, IT4G10B, IT4G10C, IT4G10D, IT4G10E, IT4G10F	New for ICCS 2022
TcQ-Q11	IT4G11A, IT4G11B, IT4G11C, IT4G11D, IT4G11E, IT4G11F	New for ICCS 2022
TcQ-Q12	IT4G12A, IT4G12B, IT4G12C, IT4G12D, IT4G12E, IT4G12F	New for ICCS 2022
TcQ-Q13	IT4G13A, IT4G13B, IT4G13C, IT4G13D, IT4G13E, IT4G13F	New for ICCS 2022
TcQ-Q14	IT4G14A, IT4G14B, IT4G14C, IT4G14D, IT4G14E, IT4G14F	Modified from ICCS 2016 Q12 (changes to items Q14A, Q14C, Q14D)
TcQ-Q15	IT4G15A, IT4G15B, IT4G15C, IT4G15D	New for ICCS 2022
TcQ-Q16	IT4G16A, IT4G16B, IT4G16C, IT4G16D, IT4G16E, IT4G16F, IT4G16G, IT4G16H, IT4G16I, IT4G16J, IT4G16K, IT4G16L, IT4G16M	Modification from ICCS 2016 Q14 (new items Q16C, Q16H, Q16M)
TcQ-Q17	IT4G17A, IT4G17B, IT4G17C, IT4G17D, IT4G17E, IT4G17F, IT4G17G, IT4G17H, IT4G17I, IT4G17J, IT4G17K, IT4G17L, IT4G17M	New for ICCS 2022
TcQ-Q18	IT4G18	Unchanged from ICCS 2016 Q15
TcQ-Q19	IT4G19A, IT4G19B, IT4G19C, IT4G19D, IT4G19E, IT4G19F, IT4G19G, IT4G19H	Modified from ICCS 2016 Q16 (changes to item Q19E)
TcQ-Q20	IT4G20A, IT4G20B, IT4G20C, IT4G20D, IT4G20E, IT4G20F, IT4G20G, IT4G20H, IT4G20I, IT4G20J	Modified from ICCS 2016 Q17 (new items Q20I, Q20J)
TcQ-Q21	IT4G21A, IT4G21B, IT4G21C, IT4G21D, IT4G21E	New for ICCS 2022
TcQ-Q22	IT4G22A, IT4G22B, IT4G22C, IT4G22D, IT4G22E, IT4G22F, IT4G22G, IT4G22H, IT4G22I, IT4G22J, IT4G22K, IT4G22L, IT4G22M, IT4G22N	Modified from ICCS 2016 Q18 (new items Q22L, Q22M)
TcQ-Q23	IT4G23A, IT4G23B, IT4G23C, IT4G23D, IT4G23E, IT4G23F, IT4G23G, IT4G23H, IT4G23I, IT4G23J, IT4G23K, IT4G23L, IT4G23M, IT4G23N	Modified from ICCS 2016 Q19 (new items Q23L, Q23M)
TcQ-Q24	IT4G24A, IT4G24B, IT4G24C, IT4G24D, IT4G24E, IT4G24F, IT4G24G, IT4G24H, IT4G24I, IT4G24J, IT4G24K, IT4G24L, IT4G24M, IT4G24N	New for ICCS 2022
TcQ-Q25	IT4G25A, IT4G25B, IT4G25C, IT4G25D, IT4G25E, IT4G25F	Modified from ICCS 2016 Q20 (new items Q25F)
TcQ-Q26	IT4G26A, IT4G26B, IT4G26C, IT4G26D, IT4G26E, IT4G26F, IT4G26G, IT4G26H	Modified from ICCS 2016 Q21 (change to items Q26B, Q26D) and new items (Q26A, Q26G, Q26H)

Table D.3: Mapping of school questionnaire items in ICCS 2022 and ICCS 2016

ICCS 2022 question	Variable name(s)	Mapping against ICCS 2016
Q1	IC4G01	Modified from ICCS 2016 Q1 (changes to response options)
Q2	IC4G02A, IC4G02B, IC4G02C, IC4G02D, IC4G02E	Unchanged from ICCS 2016 Q2
Q3	IC4G03A, IC4G03B, IC4G03C, IC4G03D, IC4G03E, IC4G03F, IC4G03G, IC4G03H, IC4G03I, IC4G03J	New for ICCS 2022
Q4	IC4G04A, IC4G04B	Unchanged from ICCS 2016 Q5
Q5	IC4G05A, IC4G05B, IC4G05C, IC4G05D, IC4G05E	New for ICCS 2022
Q6	IC4G06A, IC4G06B, IC4G06C, IC4G06D	New for ICCS 2022
Q7	IC4G07A, IC4G07B, IC4G07C, IC4G07D, IC4G07E	New for ICCS 2022
Q8	IC4G08A, IC4G08B, IC4G08C, IC4G08D, IC4G08E, IC4G08F, IC4G08G, IC4G08H, IC4G08I, IC4G08J	Modified from ICCS 2016 Q4 (change to items Q8B, Q8F) and new item (Q8I)
Q9	IC4G09A, IC4G09B, IC4G09C, IC4G09D, IC4G09E, IC4G09F	New for ICCS 2022
Q10	IC4G10A, IC4G10B, IC4G10C, IC4G10D, IC4G10E, IC4G10F, IC4G10G, IC4G10H, IC4G10I	Modified from ICCS 2016 Q9 (changes to response options and item Q10E), new items (Q10G, Q10H, Q10I)
Q11	IC4G11A, IC4G11B, IC4G11C, IC4G11D, IC4G11E, IC4G11F, IC4G11G	New for ICCS 2022
Q12	IC4G12A, IC4G12B, IC4G12C, IC4G12D, IC4G12E, IC4G12F	New for ICCS 2022
Q13	IC4G13A, IC4G13B, IC4G13C, IC4G13D, IC4G13E, IC4G13F, IC4G13G, IC4G13H, IC4G13I, IC4G13J, IC4G13K	Modified from ICCS 2016 Q11 (new item Q13K)
Q14	IC4G14A, IC4G14B, IC4G14C, IC4G14D, IC4G14E, IC4G14F, IC4G14G, IC4G14H, IC4G14I, IC4G14J, IC4G14K, IC4G14L	Modified from ICCS 2016 Q12 (change to item Q14A)
Q15	IC4G15A, IC4G15B, IC4G15C, IC4G15D	Modified from ICCS 2016 Q13 (change to items Q15A, Q15B)
Q16	IC4G16	New for ICCS 2022
Q17	IC4G17	Modified from ICCS 2016 Q15 (changes to stem/response options)
Q18	IC4G18A, IC4G18B, IC4G18C, IC4G18D, IC4G18E, IC4G18F, IC4G18G, IC4G18H	Modified from ICCS 2016 Q14 (change to item Q18E)
Q19	IC4G19A, IC4G19B, IC4G19C, IC4G19D, IC4G19E, IC4G19F, IC4G19G, IC4G19H, IC4G19I, IC4G19J, IC4G19K, IC4G19L, IC4G19M	Modified from ICCS 2016 Q16 (new items Q19C, Q19H, Q19M)
Q20	IC4G20	Unchanged from ICCS 2016 Q17
Q21	IC4G21A, IC4G21B	Unchanged from ICCS 2016 Q18
Q22	IC4G22A, IC4G22B	Unchanged from ICCS 2016 Q19
Q23	IC4G23	Unchanged from ICCS 2016 Q20
Q24	IC4G24A, IC4G24B, IC4G24C	New for ICCS 2022

Table D.4: Mapping of European regional student questionnaire items in ICCS 2022 and ICCS 2016

ICCS 2022 question	Variable name(s)	Mapping against ICCS 2016
StE-Q1	ES4G01A, ES4G01B, ES4G01C, ES4G01D, ES4G01E, ES4G01F	Unchanged from ICCS 2016 Q1
StE-Q2	ES4G02A, ES4G02B, ES4G02C, ES4G02D, ES4G02E, ES4G02F	Modified from ICCS 2016 Q2 (changes to item Q2D), new items (Q2E, Q2F)
StE-Q3	ES4G03A, ES4G03B, ES4G03C, ES4G03D, ES4G03E, ES4G03F	Modified from ICCS 2016 Q3 (changes to items Q3D, Q3E, Q3F)
StE-Q4	ES4G04A, ES4G04B, ES4G04C, ES4G04D, ES4G04E, ES4G04F, ES4G04G	Modified from ICCS 2016 Q5 (changes to all items)
StE-Q5	ES4G05A, ES4G05B, ES4G05C, ES4G05D, ES4G05E	New for ICCS 2022
StE-Q6	ES4G06A, ES4G06B, ES4G06C, ES4G06D, ES4G06E, ES4G06F, ES4G06G, ES4G06H, ES4G06I, ES4G06J, ES4G06K	New for ICCS 2022
StE-Q7	ES4G07A, ES4G07B, ES4G07C, ES4G07D, ES4G07E, ES4G07F, ES4G07G, ES4G07H, ES4G07I, ES4G07J, ES4G07K, ES4G07L, ES4G07M	Modified from ICCS 2016 Q7 (changes to items Q7C, Q7E), new items (Q7B, Q7I, Q7J, Q7K, Q7L, Q7M)
StE-Q8	ES4G08A, ES4G08B, ES4G08C, ES4G08D, ES4G08E	New for ICCS 2022
StE-Q9	ES4G09A, ES4G09B, ES4G09C, ES4G09D, ES4G09E, ES4G09F, ES4G09G, ES4G09H, ES4G09I	New for ICCS 2022
StE-Q10	ES4G10A, ES4G10B, ES4G10C, ES4G10D, ES4G10E, ES4G10F	Modified from ICCS 2016 Q9 (changes to stem and all items)
StE-Q11	ES4G11A, ES4G11B, ES4G11C, ES4G11D, ES4G11E, ES4G11F, ES4G11G, ES4G11H	New for ICCS 2022
StE-Q12	ES4G12A, ES4G12B, ES4G12C, ES4G12D, ES4G12E, ES4G12F, ES4G12G, ES4G12H, ES4G12I, ES4G12J	Modified from ICCS 2016 Q11 (changes to stem, modified, and new items)

Table D.5: Mapping of Latin American regional student questionnaire items in ICCS 2022 and ICCS 2016

ICCS 2022 question	Variable name(s)	Mapping against ICCS 2016
StL - Q1	LS4G01A, LS4G01B, LS4G01C, LS4G01D, LS4G01E, LS4G01F, LS4G01G, LS4G01H	New for ICCS 2022 (adapted from ICCS 2016 European regional student questionnaire Q8)
StL - Q2	LS4G02A, LS4G02B, LS4G02C, LS4G02D, LS4G02E, LS4G02F, LS4G02G	Modified from ICCS 2016 Q1/Q2 (combination of items)
StL - Q3	LS4G03A, LS4G03B, LS4G03C, LS4G03D, LS4G03E	Unchanged from ICCS 2016 Q3
StL - Q4	LS4G04A, LS4G04B, LS4G04C, LS4G04D, LS4G04E, LS4G04F, LS4G04G, LS4G04H, LS4G04I, LS4G04J	Unchanged from ICCS 2016 Q5
StL - Q5	LS4G05A, LS4G05B, LS4G05C, LS4G05D, LS4G05E	Unchanged from ICCS 2016 Q8
StL - Q6	LS4G06A, LS4G06B, LS4G06C, LS4G06D, LS4G06E, LS4G06F, LS4G06G, LS4G06H, LS4G06I, LS4G06J, LS4G06K	Unchanged from ICCS 2016 Q9

Appendix E

Tables with coding information for conditioning variables

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
International Student Questionnaire					
S_Age	Age of student		Imputed from year and month	Copy value (Missing = mean)	Valid = 0 Missing = 1
S_GENDER	Gender	0. boy (or other, missing) 1. girl	0 1	Copy value (Missing = 0)	Valid = 0 Missing = 1
IS4G02O	Student's ethnic background	1. <A> 2. 3. <C> 4. <D>		Length of dummy code varies across countries, Dummy coding for national options, largest group is reference category (Missing=reference category = 0)	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 = 0)	Valid = 0 Missing = 1
IS4G03OA	[parent 1]	1. Yes	1	Dummy coding for each variable (0,1)	Only missing for all = 1 Valid for at least one = 0
IS4G03OB	Other [guardian 1] (for example, step-parent or foster-parent)	2. No	0		
IS4G03OC	[parent 2]				
IS4G03OD	Other [guardian 2] (for example, step-parent or foster-parent)				
IS4G03OE	Siblings (brothers or sisters including step-brothers and step-sisters)				
IS4G03OF	Grandparents				
IS4G03OG	Others				

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS4G04A IS4G04B IS4G04C	Student's country of birth Mother's or <female guardian's> country of birth Father's or <male guardian'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 = 0)	Valid = 0 Missing = 1
IS4G05	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 = 0)	Valid = 0 Missing = 1
S_P1ISEI	Mother's occupational status	ISEI score (IS4G06A)		Copy value (Missing=mean)	Valid = 0 Missing = 1
S_P1ISCED	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 percentage is reference category (Missing = reference category)	Valid = 0 Missing = 1
S_P2ISEI	Father's occupational status	ISEI score (IS4G08A)		Copy value (Missing = mean)	Valid = 0 Missing = 1
S_P2ISCED	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 = She 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
IS4G10	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 IS4G10 variable, option with the highest percentage is reference category (Missing=reference category)	Valid = 0 Missing = 1

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
IS4G10OA	Use of a desktop or [laptop] computer	1. Never or rarely	0	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
IS4G10OB	Use of a tablet device (e.g. [iPad])	2. At least once a week	1		
IS4G10OC	Use of a [smartphone] (except for using text and calling)	3. At least once a day	2		
		4. Several times a day	3		
IS4G11A	Interest in political and social issues-Student	1. Very interested	3	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
IS4G11B	Interest in political and social issues-Mother or <female guardian>	2. Quite interested	2		
IS4G11C	Interest in political and social issues-Father or <male guardian>	3. Not very interested	1		
		4. Not interested at all	0		
IS4G12A	National and international news-watching tv	1. Never or hardly ever	0	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
IS4G12B	National and international news-reading newspaper	2. Monthly (at least once a month)	1		
		3. Weekly (at least once a week)	2		
		4. Daily or almost daily	3		
S_POLDISC	Students' discussion of political and social issues outside school	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G13A	Using the internet to find information about political or social issues	1. Never or hardly ever	0	Recode as per scoring (Missing = median)	Valid = 0 Missing = 1
		2. Monthly (at least once a month)	1		
		3. Weekly (at least once a week)	2		
		4. Daily or almost daily	3		
S_ENGDM	Students' engagement with digital media	IRT scores (WLE)		(Missing = mean)	Valid = 0 Missing = 1
S_COMPART	Students' participation in the wider community	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G14F	A sports team	1. Yes, I have done this within the last twelve months	2	Recode as per scoring (Missing = median) (* Optional variables)	Valid = 0 Missing = 1
IS4G14G	A religious group or organization	2. Yes, I have done this but more than a year ago	1		
		3. No, I have never done this	0		
S_SCHPART	Students' participation at school	IRT scores (WLE)		(Missing = mean)	Valid = 0 Missing = 1
S_OPDISC	Students' perception of openness in classroom discussions	IRT scores (WLE)		(Missing = mean)	Valid = 0 Missing = 1

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
S_CIVLRN	Student reports on civic learning at school	IRT scores (WLE)		(Missing = mean)	Valid = 0 Missing = 1
S_STUTREL	Students' perception of student-teacher relations at school	IRT scores (WLE)		(Missing = mean)	Valid = 0 Missing = 1
S_INTACT	Students' perceptions of student interaction at school	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G18I	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
S_INFDEC	Students' perception of the value of participation at their school	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_SYSPPOS	Student's support for political system	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_SYSNEG	Student's cynical attitudes toward political system	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G20C	Political decisions should more often be based on advice from scientific experts.	1. Strongly agree	3	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
IS4G20G	Democracy may have some problems but it is still the best form of government for [country of test].	2. Agree	2		
		3. Disagree	1		
		4. Strongly disagree	0		
S_DEMTHRT	Student perceptions of threats to democracy	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_CITCON	Students' perception of the importance of conventional citizenship	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_CITSOC	Students' perception of the importance of social movement related citizenship	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_GLOBCIT	Students' perceptions of the importance of global citizenship	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G22C	Learning about the country's history	1. Very important 2. Quite important 3. Not very important 4. Not important at all	3 2 1 0	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
S_ENREST	Students' endorsement of restrictions in a national emergency	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_IMMRGHT	Students' attitudes toward equal rights for immigrants	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
S_GENEQL	Students' attitudes toward gender equality	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G25G	There should be equal numbers of men and women in [parliament/ congress].	1. Strongly agree 2. Agree 3. Disagree 4. Strongly disagree	3 2 1 0	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
S_ATTENV	Students' attitudes toward environmental protection	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_ETHRGHT	Students' attitudes toward equal rights for all ethnic/racial groups	IRT scores (WLE)		(Missing=mean) Optional variable	Valid = 0 Missing = 1
S_INTRUST	Students' trust in civic institutions	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G27G	Traditional media (television, newspapers, radio)	1. Completely	3	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
IS4G27H	Social media (e.g. [Instagram, Facebook, Twitter])	2. Quite a lot	2	(* O, P, Q - Optional variables)	
IS4G27I	[The Armed Forces]	3. A little	1		
IS4G27J	Scientists	4. Not at all	0		
IS4G27K	Your teachers				
IS4G27L	Schools in general				
IS4G27M	The United Nations				
IS4G27N	People in general				
IS4G27O	[State/Province] government				
IS4G27P	European Commission				
IS4G27Q	European Parliament				
IS4G28A	Pollution	1. To a large extent	3	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
IS4G28B	Violent conflict	2. To a moderate extent	2		
IS4G28C	Poverty	3. To a small extent	1		
IS4G28D	Climate change	4. Not at all	0		
IS4G28E	Unemployment				
IS4G28F	Overpopulation				
IS4G28G	Infectious diseases (e.g., [measles], [COVID-19])				
IS4G28H	Terrorism				
IS4G28I	Global financial crises				
IS4G28J	[Loss of biodiversity, extinction of living species]				
IS4G28K	Water shortages				
S_CITEFF	Students' sense of citizenship self-efficacy	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
S_SCACT	Students' willingness to participate in school activities	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_LEGACT	Students' expected participation in legal activities	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_ENVACT	Students' expected participation in activities to protect the environment	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_ILLACT	Students' expected participation in illegal protest activities	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G31N	Participate in a school strike	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)	Valid = 0 Missing = 1
S_ELECPART	Students' expected electoral participation	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
S_POLPART	Students' expected active political participation	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
IS4G32D	Help a candidate or party during an election campaign	1. I would certainly do this	3	Recode as per scoring (Missing=median) (* J, K - Optional variables)	Valid = 0 Missing = 1
IS4G32I	Volunteer time to help other people in the [local community]	2. I would probably do this	2		
IS4G32J	Vote in [state, province elections]	3. I would probably not do this	1		
IS4G32K	Vote in European elections	4. I would certainly not do this	0		
IS4G33	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
IS4G34	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
S_RELINF	Students' attitudes toward the influence of religion in society	IRT scores (WLE)		(Missing=mean) Optional variable	Valid = 0 Missing = 1
IS4G35F	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

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
European Regional Questionnaire					
E_EUIDENT	Students' sense of European identity	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
ESRG01E	I feel part of the European Union	1. Strongly agree	3	Recode as per scoring (Missing=median) (* Optional variables)	Valid = 0 Missing = 1
ESRG01F	I am proud that my country is a member of the European Union	2. Agree	2		
		3. Disagree	1		
		4. Strongly disagree	0		
E_EULRN	Student reports on opportunities for learning about Europe at school	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
ESRG02F	Role and functions of the European Union institutions (e.g., European Parliament, European Council, European Commission)	1. To a large extent	3	Recode as per scoring (Missing=median) (Optional variable)	Valid = 0 Missing = 1
		2. To a moderate extent	2		
		3. To a small extent	1		
		4. Not at all	0		
E_FREEMOVE	Students' attitudes toward freedom of migration within Europe	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_RESMIG	Students' attitudes toward restrictions of migration in Europe	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_CCOOP	Students' attitudes toward cooperation among European countries	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_COOPENV	Students' endorsement of environmental cooperation in Europe	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
ES4G06A	Women	1. A lot	3	Recode as per scoring (Missing=median) (K = optional variable)	Valid = 0 Missing = 1
ES4G06B	Young people	2. To some extent	2		
ES4G06C	Poor people	3. A little	1		
ES4G06D	Religious minorities	4. Not at all	0		
ES4G06E	People with disabilities				
ES4G06F	Older people				
ES4G06G	[Immigrants]				
ES4G06H	People with political opinions different from those of the majority				
ES4G06I	[Members of the LGTB+ community]				
ES4G06J	[People of African descent]				
ES4G06K	People from ethnic minority groups				

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
E_EUFPOS	Students' positive expectations for European future	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_EUFNEG	Students' negative expectations for European future	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_INDFUT	Students' expectations for their own individual future	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
ES4G09A	To be economically independent	1. Very important	3	Recode as per scoring (Missing=median)	
ES4G09B	To have a higher income than my [parents or guardians]	2. Quite important	2		
ES4G09C	To find a job I like	3. Not very important	1		
ES4G09D	To have the opportunity to travel abroad for leisure	4. Not important at all	0		
ES4G09E	To have the opportunity to work abroad				
ES4G09F	To be paid in line with my qualifications				
ES4G09G	To get a [post-secondary] degree				
ES4G09H	To have children				
ES4G09I	To have friends				
E_ETHCON	Students' reports on political and ethical consumerism behaviors	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_SUSBEH	Students' reports on their sustainable behaviors	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_EUATPOS	Students' positive attitudes toward European Union	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
E_EUATNEG	Students' negative attitudes toward European Union	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1

Variable	Variable label	Response options	Scoring	Conditioning variables	
				Scaling variable	Missing variable
Latin American Regional Questionnaire					
LS4G01A	I will find a steady job.	1. Very likely	3	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
LS4G01B	My financial situation will be better than that of my parents.	2. Likely	2		
LS4G01C	I will find a job I like.	3. Unlikely	1		
LS4G01D	I will have the opportunity to travel abroad for leisure.	4. Very unlikely	0		
LS4G01E	I will earn enough money to start a family.				
LS4G01F	I will move to another country.				
LS4G01G	I will move to another city or town in [country of test].				
LS4G01H	I will be able to study what I am interested in after I have finished school				
L_AUTGOV	Students' attitudes toward authoritarianism in government	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
L_ATTCORR	Students' attitudes toward corrupt practices in government	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
L_DISLAW	Students' attitudes toward disobeying the law	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
LS4G04G	When it is the only way of fighting publicly against an unfair law.	1. Strongly agree	3	Recode as per scoring (Missing=median)	Valid = 0 Missing = 1
		2. Agree	2		
		3. Disagree	1		
		4. Strongly disagree	0		
L_ATTHS	Attitudes toward homosexuals	IRT scores (WLE)		(Missing=mean)	Valid = 0 Missing = 1
LS4G06A	Women	1. A lot	3	Recode as per scoring (Missing=median) (K = optional variable)	Valid = 0 Missing = 1
LS4G06B	Young people	2. To some extent	2		
LS4G06C	Poor people	3. A little	1		
LS4G06D	Religious minorities	4. Not at all	0		
LS4G06E	People with disabilities				
LS4G06F	Older people.				
LS4G06G	[Immigrants]				
LS4G06H	[Members of the LGTB+ community]				
LS4G06I	[People of African origin]				
LS4G06J	Unemployed people				
LS4G06K	[People of Indigenous origin]				

Note: Optional items are shaded in grey.

Appendix F

Tables comparing questionnaire scale means across delivery modes

Table F.1: National average scale scores for student questionnaire scales across different delivery modes

Country	S_POLDISC			S_ENGDM			S_COMPART			S_SCHPART			S_OPDISC		
	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference
Bulgaria	52 (0.2)	52 (0.4)	-0.1 (0.5)	51 (0.3)	51 (0.5)	0.0 (0.6)	55 (0.3)	55 (0.4)	0.0 (0.5)	48 (0.3)	48 (0.5)	-0.7 (0.6)	48 (0.4)	50 (0.5)	-1.5 (0.6)
Chinese Taipei	54 (0.2)	53 (0.3)	0.7 (0.4)	50 (0.2)	48 (0.3)	1.3	46 (0.2)	45 (0.3)	0.2 (0.3)	53 (0.2)	53 (0.4)	0.3 (0.4)	54 (0.4)	53 (0.6)	1.1 (0.7)
Colombia	53 (0.3)	53 (0.4)	0.1 (0.5)	52 (0.3)	52 (0.4)	-0.8 (0.5)	53 (0.2)	52 (0.4)	0.5 (0.5)	53 (0.2)	52 (0.3)	0.6 (0.4)	51 (0.3)	52 (0.6)	-0.4 (0.7)
Estonia	55 (0.3)	55 (0.4)	0.3 (0.5)	48 (0.2)	49 (0.4)	-0.2 (0.5)	49 (0.3)	48 (0.8)	0.3 (0.8)	46 (0.4)	44 (0.7)	1.1 (0.8)	48 (0.4)	48 (0.6)	-0.2 (0.7)
Italy	55 (0.2)	57 (0.3)	-1.7	51 (0.2)	52 (0.3)	-0.5 (0.4)	49 (0.2)	49 (0.3)	-0.4 (0.4)	50 (0.6)	50 (0.6)	0.5 (0.8)	55 (0.3)	55 (0.4)	-0.4 (0.5)
Latvia ¹	55 (0.2)	54 (0.3)	0.4 (0.3)	49 (0.2)	49 (0.3)	0.2 (0.4)	50 (0.2)	50 (0.5)	-0.2 (0.5)	46 (0.3)	47 (0.7)	-1.0 (0.8)	46 (0.4)	47 (0.5)	-1.6 (0.6)
Lithuania	57 (0.2)	57 (0.2)	0.3 (0.3)	51 (0.2)	50 (0.3)	0.9	52 (0.3)	52 (0.4)	-0.3 (0.5)	49 (0.3)	50 (0.3)	-0.2 (0.4)	49 (0.3)	50 (0.3)	-0.6 (0.5)
Malta	54 (0.4)	54 (0.5)	-0.4 (0.7)	51 (0.3)	51 (0.4)	-0.1 (0.5)	52 (0.2)	52 (0.7)	-0.1 (0.7)	49 (0.5)	49 (0.5)	-0.2 (0.7)	50 (0.3)	50 (0.6)	-0.1 (0.7)
Norway (9) ¹	54 (0.1)	54 (0.3)	0.6 (0.4)	49 (0.2)	52 (0.4)	-3.0	47 (0.2)	47 (0.3)	0.0 (0.3)	53 (0.2)	50 (0.4)	2.9 (0.4)	52 (0.3)	53 (0.5)	-0.9 (0.6)
Slovenia	51 (0.2)	51 (0.3)	0.2 (0.3)	48 (0.2)	47 (0.2)	0.8	50 (0.2)	50 (0.4)	0.1 (0.4)	49 (0.2)	49 (0.3)	0.2 (0.4)	46 (0.3)	47 (0.4)	-0.5 (0.5)
Sweden ¹	55 (0.2)	55 (0.3)	0.3 (0.4)	50 (0.2)	50 (0.3)	0.3 (0.3)	47 (0.2)	47 (0.4)	-0.4 (0.4)	51 (0.3)	51 (0.3)	-0.3 (0.4)	50 (0.3)	51 (0.5)	-0.7 (0.6)
ICCS 2022 average	54 (0.1)	54 (0.1)	0.1 (0.1)	50 (0.1)	50 (0.1)	-0.1 (0.1)	50 (0.1)	50 (0.1)	0.0 (0.2)	50 (0.1)	49 (0.1)	0.3 (0.2)	50 (0.1)	50 (0.2)	-0.5 (0.2)

Notes:

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some aggregate statistics may appear inconsistent.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

- No comparable data available.

Table F.2: National average scale scores for student questionnaire scales across different delivery

Country	S_CIVLRN			S_STUTREL			S_INTACT			S_INFDEC			S_SYSSAT		
	Student reports about civic learning at school			Students' perceptions of student-teacher relations at school			Students' perceptions of student interaction at school			Students' beliefs about their influence on decision-making at school			Students' satisfaction with the political system		
	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference
Bulgaria	47 (0.3)	47 (0.4)	0.3 (0.5)	51 (0.3)	51 (0.5)	-0.2 (0.6)	49 (0.3)	50 (0.6)	-0.3 (0.7)	52 (0.3)	52 (0.4)	-0.2 (0.5)	47 (0.4)	47 (0.6)	0.4 (0.7)
Chinese Taipei	59 (0.3)	59 (0.4)	-0.3 (0.5)	56 (0.3)	56 (0.8)	0.3 (0.9)	55 (0.3)	55 (0.5)	0.1 (0.6)	56 (0.3)	54 (0.5)	1.4 (0.6)	55 (0.2)	54 (0.2)	0.3 (0.3)
Colombia	53 (0.3)	52 (0.5)	1.0 (0.6)	54 (0.3)	54 (0.5)	0.0 (0.6)	49 (0.3)	50 (0.5)	-0.3 (0.6)	57 (0.3)	56 (0.3)	0.5 (0.4)	52 (0.4)	50 (0.6)	1.1 (0.7)
Estonia	45 (0.4)	46 (0.3)	-0.3 (0.5)	49 (0.3)	50 (0.4)	-0.8 (0.6)	48 (0.4)	47 (0.6)	0.4 (0.7)	49 (0.3)	49 (0.4)	0.4 (0.5)	52 (0.2)	52 (0.4)	-0.3 (0.5)
Italy	51 (0.3)	51 (0.4)	0.2 (0.4)	52 (0.3)	52 (0.4)	-0.5 (0.5)	48 (0.4)	47 (0.4)	0.3 (0.6)	50 (0.3)	49 (0.4)	0.4 (0.5)	49 (0.2)	49 (0.4)	-0.8 (0.4)
Latvia ¹	45 (0.3)	45 (0.6)	0.3 (0.7)	46 (0.3)	47 (0.4)	-1.0 (0.5)	46 (0.3)	48 (0.5)	-1.8 (0.6)	49 (0.3)	49 (0.5)	0.4 (0.5)	48 (0.2)	48 (0.5)	-0.1 (0.5)
Lithuania	46 (0.3)	47 (0.3)	-0.3 (0.4)	49 (0.4)	49 (0.5)	0.2 (0.6)	47 (0.3)	47 (0.5)	-0.3 (0.6)	50 (0.3)	49 (0.4)	0.6 (0.5)	49 (0.2)	50 (0.4)	-0.2 (0.4)
Malta	47 (0.3)	46 (0.5)	1.0 (0.6)	52 (0.3)	51 (0.7)	0.3 (0.8)	48 (0.5)	48 (1.0)	0.1 (1.2)	52 (0.4)	52 (0.9)	0.5 (1.0)	52 (0.4)	52 (0.7)	0.3 (0.8)
Norway (9) ¹	52 (0.2)	51 (0.4)	1.1 (0.5)	52 (0.3)	53 (0.5)	-1.1 (0.6)	50 (0.3)	51 (0.4)	-0.9 (0.5)	50 (0.2)	49 (0.4)	1.0 (0.5)	57 (0.2)	56 (0.3)	0.7 (0.3)
Slovenia	52 (0.2)	51 (0.4)	0.5 (0.4)	47 (0.3)	47 (0.4)	-0.2 (0.5)	46 (0.3)	46 (0.4)	-0.6 (0.5)	48 (0.3)	48 (0.4)	-0.1 (0.4)	50 (0.2)	50 (0.3)	-0.1 (0.4)
Sweden ¹	52 (0.4)	52 (0.4)	0.2 (0.6)	51 (0.3)	52 (0.4)	-0.7 (0.5)	47 (0.3)	48 (0.5)	-0.6 (0.6)	49 (0.3)	49 (0.4)	0.1 (0.5)	54 (0.3)	54 (0.3)	0.1 (0.4)
ICCS 2022 average	50 (0.1)	50 (0.1)	0.3 (0.2)	51 (0.1)	51 (0.2)	-0.3 (0.2)	49 (0.1)	49 (0.2)	-0.4 (0.2)	51 (0.1)	51 (0.1)	0.4 (0.2)	51 (0.1)	51 (0.1)	0.1 (0.2)

Notes:

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some aggregate statistics may appear inconsistent.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

- No comparable data available.

Table F.3: National average scale scores for student questionnaire scales across different delivery modes

Country	S_SYSCRT			S_DEMTHRT			S_CITCON			S_CITSOC			S_GLOBCIT		
	Students' critical views of the political system			Students' beliefs about threats to democracy			Students' beliefs about the importance of conventional citizenship activities			Students' beliefs about the importance of social movement related citizenship activities			Students' beliefs about the importance of globally oriented citizenship activities		
	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference
Bulgaria	52 (0.3)	53 (0.3)	-0.4 (0.5)	48 (0.2)	47 (0.3)	0.5 (0.4)	47 (0.3)	45 (0.4)	2.5 (0.5)	49 (0.3)	48 (0.3)	1.0 (0.4)	48 (0.3)	47 (0.3)	1.4 (0.4)
Chinese Taipei	47 (0.2)	47 (0.3)	0.7 (0.4)	57 (0.3)	57 (0.4)	-0.3 (0.5)	54 (0.2)	53 (0.4)	1.4 (0.5)	52 (0.2)	51 (0.3)	0.5 (0.4)	52 (0.2)	52 (0.4)	0.1 (0.4)
Colombia	52 (0.2)	52 (0.4)	-0.3 (0.5)	46 (0.3)	47 (0.5)	-0.5 (0.6)	52 (0.3)	50 (0.4)	1.3 (0.5)	52 (0.3)	52 (0.3)	-0.6 (0.4)	53 (0.2)	53 (0.5)	-0.6 (0.5)
Estonia	47 (0.2)	47 (0.4)	0.3 (0.4)	49 (0.3)	50 (0.3)	-0.4 (0.5)	44 (0.2)	44 (0.3)	0.4 (0.4)	45 (0.2)	46 (0.3)	-0.5 (0.4)	48 (0.3)	49 (0.3)	-0.8 (0.5)
Italy	51 (0.3)	50 (0.3)	0.8 (0.4)	50 (0.3)	50 (0.3)	0.5 (0.4)	53 (0.3)	51 (0.3)	1.5 (0.4)	51 (0.3)	50 (0.2)	0.6 (0.3)	53 (0.2)	53 (0.3)	0.3 (0.4)
Latvia ¹	50 (0.3)	50 (0.4)	0.6 (0.5)	47 (0.2)	48 (0.3)	-0.9 (0.4)	47 (0.3)	45 (0.4)	1.7 (0.5)	46 (0.2)	47 (0.3)	-1.0 (0.4)	47 (0.2)	48 (0.2)	-0.4 (0.3)
Lithuania	51 (0.2)	50 (0.3)	0.5 (0.4)	47 (0.2)	47 (0.4)	0.5 (0.4)	49 (0.3)	48 (0.4)	1.1 (0.5)	48 (0.3)	48 (0.3)	0.2 (0.4)	50 (0.2)	50 (0.4)	0.5 (0.4)
Malta	49 (0.3)	49 (0.3)	0.3 (0.4)	48 (0.3)	48 (0.4)	-0.2 (0.5)	48 (0.3)	46 (0.6)	1.7 (0.7)	49 (0.4)	48 (0.5)	0.7 (0.7)	51 (0.4)	50 (0.6)	0.5 (0.7)
Norway (9) ¹	47 (0.2)	46 (0.3)	0.6 (0.3)	50 (0.2)	49 (0.3)	1.4 (0.4)	50 (0.2)	47 (0.2)	2.4 (0.3)	47 (0.2)	47 (0.3)	0.8 (0.3)	49 (0.2)	48 (0.3)	0.9 (0.4)
Slovenia	51 (0.2)	51 (0.3)	0.1 (0.3)	50 (0.2)	50 (0.4)	0.3 (0.4)	50 (0.3)	48 (0.4)	1.8 (0.4)	48 (0.2)	47 (0.3)	0.3 (0.4)	49 (0.2)	48 (0.3)	0.7 (0.4)
Sweden ¹	45 (0.2)	46 (0.3)	-0.2 (0.4)	54 (0.3)	55 (0.3)	-0.4 (0.4)	48 (0.2)	48 (0.3)	-0.1 (0.3)	46 (0.2)	46 (0.4)	-0.1 (0.5)	48 (0.2)	48 (0.3)	-0.2 (0.4)
ICCS 2022 average	49 (0.1)	49 (0.1)	0.3 (0.1)	50 (0.1)	50 (0.1)	0.0 (0.1)	49 (0.1)	48 (0.1)	1.4 (0.1)	48 (0.1)	48 (0.1)	0.2 (0.1)	50 (0.1)	50 (0.1)	0.2 (0.1)

Notes:

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some aggregate statistics may appear inconsistent.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

- No comparable data available.

Table F.4: National average scale scores for student questionnaire scales across different delivery modes

Country	S_ENREST			S_IMMPOS			S_GENEQL			S_ATTENV			S_ETHRGHT		
	Students' endorsement of restrictions in a national emergency			Students' positive attitudes toward immigrants			Students' endorsement of gender equality			Students' positive attitudes toward environmental protection			Students' endorsement of equal rights for all ethnic groups		
	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference
Bulgaria	49 (0.3)	48 (0.5)	0.8 (0.6)	46 (0.3)	46 (0.4)	0.4 (0.5)	46 (0.3)	47 (0.4)	-0.5 (0.5)	49 (0.3)	49 (0.4)	0.0 (0.5)	49 (0.3)	49 (0.5)	0.0 (0.6)
Chinese Taipei	54 (0.2)	53 (0.4)	0.9 (0.4)	57 (0.2)	56 (0.3)	0.7 (0.4)	58 (0.2)	59 (0.5)	-1.2 (0.5)	53 (0.2)	53 (0.3)	0.8 (0.4)	58 (0.2)	57 (0.4)	0.8 (0.4)
Colombia	51 (0.2)	51 (0.4)	0.0 (0.5)	49 (0.2)	50 (0.3)	-0.4 (0.4)	48 (0.4)	51 (0.6)	-2.2 (0.7)	52 (0.3)	53 (0.4)	-0.7 (0.5)	51 (0.3)	52 (0.5)	-0.9 (0.5)
Estonia	51 (0.2)	50 (0.4)	0.4 (0.4)	48 (0.3)	47 (0.4)	0.4 (0.5)	51 (0.4)	51 (0.6)	0.2 (0.7)	49 (0.3)	49 (0.3)	0.2 (0.4)	51 (0.3)	52 (0.4)	-0.2 (0.5)
Italy	51 (0.2)	52 (0.3)	-0.7 (0.3)	53 (0.3)	53 (0.3)	0.1 (0.4)	56 (0.3)	56 (0.4)	-0.3 (0.5)	52 (0.2)	52 (0.3)	0.3 (0.3)	54 (0.2)	54 (0.3)	0.2 (0.4)
Latvia ¹	48 (0.2)	48 (0.2)	0.1 (0.3)	46 (0.2)	47 (0.3)	-0.6 (0.4)	48 (0.2)	50 (0.4)	-1.4 (0.5)	47 (0.2)	48 (0.3)	-0.6 (0.3)	48 (0.2)	49 (0.4)	-0.8 (0.5)
Lithuania	51 (0.2)	51 (0.3)	0.6 (0.4)	48 (0.2)	47 (0.3)	0.3 (0.4)	51 (0.3)	51 (0.4)	0.6 (0.5)	50 (0.2)	50 (0.3)	0.1 (0.4)	54 (0.3)	53 (0.5)	0.8 (0.6)
Malta	51 (0.2)	51 (0.5)	0.4 (0.5)	51 (0.4)	51 (0.6)	0.2 (0.7)	54 (0.7)	54 (0.8)	-0.1 (1.1)	51 (0.5)	52 (0.7)	-0.4 (0.8)	52 (0.5)	52 (0.6)	0.1 (0.8)
Norway (9) ¹	53 (0.2)	52 (0.2)	0.6 (0.3)	53 (0.2)	53 (0.4)	0.2 (0.5)	55 (0.2)	56 (0.3)	-0.8 (0.4)	50 (0.2)	49 (0.4)	0.2 (0.4)	56 (0.3)	56 (0.5)	-0.4 (0.6)
Slovenia	48 (0.2)	47 (0.3)	0.6 (0.4)	49 (0.2)	49 (0.4)	0.2 (0.5)	50 (0.3)	51 (0.4)	-0.6 (0.5)	49 (0.2)	49 (0.3)	0.5 (0.4)	50 (0.2)	50 (0.4)	0.2 (0.4)
Sweden ¹	50 (0.3)	50 (0.3)	0.0 (0.4)	54 (0.3)	54 (0.5)	-0.5 (0.6)	56 (0.3)	57 (0.4)	-0.6 (0.5)	50 (0.2)	50 (0.4)	-0.1 (0.4)	58 (0.3)	58 (0.5)	-0.1 (0.5)
ICCS 2022 average	51 (0.1)	50 (0.1)	0.3 (0.1)	50 (0.1)	50 (0.1)	0.1 (0.1)	52 (0.1)	53 (0.1)	-0.6 (0.2)	50 (0.1)	50 (0.1)	0.0 (0.1)	53 (0.1)	53 (0.1)	0.0 (0.2)

Notes:

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some aggregate statistics may appear inconsistent.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

- No comparable data available.

Table F.5: National average scale scores for student questionnaire scales across different delivery modes

Country	S_INTRUST			S_ENVCON			S_CITEFF			S_SCACT			S_LEGACT		
	Students' trust in civic institutions			Students' concern about threats to the global environment			Students' citizenship self-efficacy			Students' willingness to participate in school activities			Students' expected participation in legal activities		
	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference
Bulgaria	47 (0.3)	47 (0.6)	0.0 (0.7)	49 (0.3)	49 (0.4)	0.3 (0.5)	52 (0.3)	49 (0.6)	2.5 (0.6)	49 (0.3)	48 (0.5)	1.0 (0.6)	50 (0.3)	49 (0.5)	1.6 (0.6)
Chinese Taipei	52 (0.2)	52 (0.4)	-0.2 (0.5)	53 (0.2)	53 (0.3)	0.0 (0.3)	54 (0.2)	51 (0.3)	2.8 (0.4)	54 (0.2)	53 (0.4)	0.8 (0.5)	48 (0.2)	47 (0.3)	1.0 (0.4)
Colombia	46 (0.4)	45 (0.4)	0.4 (0.6)	49 (0.4)	51 (0.6)	-1.4 (0.8)	51 (0.3)	49 (0.5)	1.9 (0.5)	53 (0.2)	52 (0.3)	0.8 (0.4)	53 (0.3)	51 (0.4)	1.9 (0.5)
Estonia	51 (0.3)	52 (0.5)	-1.2 (0.5)	51 (0.3)	51 (0.3)	-0.2 (0.4)	49 (0.3)	49 (0.3)	0.7 (0.4)	46 (0.3)	45 (0.6)	0.6 (0.7)	46 (0.3)	45 (0.4)	1.2 (0.5)
Italy	49 (0.2)	50 (0.3)	-0.7 (0.3)	52 (0.2)	52 (0.3)	0.1 (0.3)	52 (0.2)	52 (0.3)	0.5 (0.4)	52 (0.3)	51 (0.3)	0.3 (0.4)	49 (0.3)	48 (0.2)	0.9 (0.4)
Latvia ¹	48 (0.3)	48 (0.4)	-0.3 (0.5)	50 (0.3)	51 (0.3)	-1.0 (0.4)	49 (0.3)	48 (0.3)	0.6 (0.4)	46 (0.3)	45 (0.6)	1.0 (0.7)	46 (0.3)	45 (0.4)	0.7 (0.5)
Lithuania	51 (0.2)	51 (0.4)	-0.6 (0.4)	52 (0.2)	52 (0.3)	0.4 (0.4)	50 (0.2)	48 (0.4)	1.7 (0.5)	49 (0.3)	49 (0.5)	0.6 (0.5)	50 (0.2)	49 (0.4)	0.6 (0.5)
Malta	49 (0.3)	48 (0.4)	0.3 (0.5)	49 (0.4)	49 (0.7)	0.2 (0.8)	50 (0.4)	48 (0.7)	1.7 (0.8)	48 (0.3)	48 (0.7)	-0.2 (0.8)	47 (0.3)	46 (0.5)	0.8 (0.6)
Norway (9) ¹	55 (0.2)	55 (0.3)	0.0 (0.3)	48 (0.2)	46 (0.3)	2.7 (0.3)	50 (0.2)	49 (0.4)	1.0 (0.5)	47 (0.2)	45 (0.3)	2.2 (0.4)	46 (0.2)	46 (0.3)	0.3 (0.4)
Slovenia	47 (0.2)	47 (0.2)	0.0 (0.4)	50 (0.2)	50 (0.3)	0.2 (0.4)	50 (0.2)	49 (0.4)	1.4 (0.5)	47 (0.2)	47 (0.4)	0.4 (0.5)	48 (0.2)	47 (0.4)	1.6 (0.5)
Sweden ¹	54 (0.3)	53 (0.3)	0.2 (0.5)	49 (0.2)	48 (0.3)	0.2 (0.4)	51 (0.3)	51 (0.4)	0.4 (0.5)	46 (0.3)	47 (0.5)	-0.5 (0.6)	48 (0.3)	47 (0.3)	0.2 (0.4)
ICCS 2022 average	50 (0.1)	50 (0.1)	-0.2 (0.1)	50 (0.1)	50 (0.1)	0.1 (0.1)	51 (0.1)	49 (0.1)	1.4 (0.2)	49 (0.1)	48 (0.1)	0.6 (0.2)	48 (0.1)	47 (0.1)	1.0 (0.1)

Notes:

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some aggregate statistics may appear inconsistent.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

- No comparable data available.

Table F.6: National average scale scores for student questionnaire scales across different delivery modes

Country	S_ENVACT			S_ILFACT			S_ELECPART			S_POLIPART			S_RELINF		
	Students' expected participation in environmental protection activities			Students' expected participation in illegal protest activities			Students' expected electoral participation			Students' expected active political participation			Students' positive attitudes toward the influence of religion in society		
	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference	Core	Bridge	Difference
Bulgaria	50 (0.2)	50 (0.4)	-0.4 (0.4)	54 (0.3)	53 (0.4)	1.9 (0.5)	47 (0.3)	48 (0.5)	-0.3 (0.6)	50 (0.3)	48 (0.5)	2.0 (0.6)	50 (0.2)	50 (0.5)	0.2 (0.5)
Chinese Taipei	53 (0.2)	53 (0.3)	-0.6 (0.4)	47 (0.2)	45 (0.4)	1.7 (0.4)	51 (0.2)	51 (0.4)	0.5 (0.4)	50 (0.2)	48 (0.3)	1.4 (0.4)	45 (0.2)	45 (0.3)	0.7 (0.4)
Colombia	54 (0.3)	55 (0.4)	-1.9 (0.5)	55 (0.3)	52 (0.5)	3.0 (0.6)	50 (0.2)	50 (0.4)	-0.7 (0.4)	53 (0.3)	51 (0.5)	2.1 (0.6)	53 (0.3)	51 (0.6)	1.2 (0.6)
Estonia	47 (0.3)	47 (0.5)	0.0 (0.6)	49 (0.3)	48 (0.2)	0.6 (0.4)	46 (0.4)	46 (0.5)	-0.4 (0.7)	48 (0.2)	47 (0.3)	0.6 (0.3)	-	-	-
Italy	53 (0.2)	53 (0.3)	-0.3 (0.4)	50 (0.3)	48 (0.2)	1.6 (0.4)	52 (0.2)	52 (0.3)	-0.5 (0.4)	51 (0.2)	50 (0.2)	1.0 (0.3)	-	-	-
Latvia ¹	46 (0.3)	48 (0.4)	-1.5 (0.5)	48 (0.2)	46 (0.2)	2.5 (0.3)	44 (0.3)	45 (0.5)	-0.1 (0.6)	49 (0.2)	47 (0.4)	2.0 (0.5)	44 (0.2)	42 (0.4)	1.4 (0.5)
Lithuania	51 (0.2)	51 (0.4)	-0.3 (0.5)	52 (0.3)	51 (0.3)	0.7 (0.4)	49 (0.3)	49 (0.4)	0.3 (0.5)	50 (0.2)	49 (0.3)	0.9 (0.3)	45 (0.2)	45 (0.4)	-0.2 (0.5)
Malta	50 (0.3)	51 (0.8)	-0.9 (0.8)	49 (0.3)	49 (0.4)	0.1 (0.5)	47 (0.4)	47 (0.6)	0.0 (0.7)	50 (0.2)	48 (0.7)	1.8 (0.7)	52 (0.3)	51 (0.6)	0.7 (0.7)
Norway (9) ¹	46 (0.2)	46 (0.3)	-0.5 (0.3)	49 (0.2)	48 (0.2)	0.8 (0.3)	52 (0.2)	53 (0.3)	-0.6 (0.4)	49 (0.2)	48 (0.3)	1.2 (0.3)	44 (0.2)	43 (0.3)	0.6 (0.4)
Slovenia	49 (0.2)	49 (0.3)	0.1 (0.4)	52 (0.2)	51 (0.3)	1.1 (0.4)	48 (0.2)	48 (0.4)	-0.2 (0.4)	49 (0.2)	48 (0.4)	1.2 (0.4)	48 (0.3)	47 (0.4)	0.4 (0.5)
Sweden ¹	47 (0.2)	48 (0.3)	-0.6 (0.4)	47 (0.3)	47 (0.4)	0.2 (0.4)	51 (0.3)	52 (0.3)	-0.3 (0.4)	50 (0.2)	50 (0.3)	0.5 (0.4)	45 (0.4)	44 (0.5)	0.5 (0.7)
ICCS 2022 average	50 (0.1)	50 (0.1)	-0.6 (0.1)	50 (0.1)	49 (0.1)	1.3 (0.1)	49 (0.1)	49 (0.1)	-0.2 (0.2)	50 (0.1)	49 (0.1)	1.3 (0.1)	47 (0.1)	47 (0.2)	0.6 (0.2)

Notes:

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some aggregate statistics may appear inconsistent.

Statistically significant changes ($p < 0.05$) are displayed in bold.

(9) Country deviated from international defined population and surveyed adjacent upper grade.

¹ National defined population covers 90% to 95% of national target population

- No comparable data available.



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 third decade of the 21st century. ICCS 2022 is the third cycle of a study initiated in 2009 and collected data in 24 countries and benchmarking participants in Europe, Latin America, and Asia. For the first time, ICCS 2022 also included computer-based delivery, which was implemented in two thirds of the participating education systems.

This technical report follows the publication of several international and European reports that presented the results of ICCS 2022. It comprises detailed information on the development of the instruments including their translation and translation verification, sampling design and implementation, the implementation of computer-based delivery system, field operations and quality control of data collection, data management, sampling weights and participation rates, and the scaling, analysis and reporting of ICCS 2022 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 2022 User Guide for the International Database, will provide guidance for their own analyses.

Over the past 50 years, 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 2022 database will allow participating education systems to evaluate the strength of educational policies, from a comparative perspective, and to measure progress in achieving critical social objectives of their educational policy agenda including critical components of the United Nations' 2030 agenda for sustainable development.

