IEA Civic Education Study Technical Report

Edited by Wolfram Schulz and Heiko Sibberns

The International Association for the Evaluation of Educational Achievement

This volume contains the technical report for the IEA Civic Education Study conducted by the International Association for the Evaluation of Educational Achievement (IEA). IEA, headquartered in Amsterdam, is a consortium of educational research organizations in more than 60 countries, with a history of conducting cross-national research in education.

During 1999, researchers surveyed nearly 90,000 14-year-old students in 28 countries. Findings for this age group were released in 2001 and reported in *Citizenship and Education in Twenty-eight Countries: Civic Knowledge and Engagement at Age Fourteen* (Torney-Purta, Lehmann, Oswald, & Schulz). In the following year, over 50,000 upper secondary students (ranging in age from 16 to 19 years) from 16 countries received a similar test and the same survey of civic concepts and attitudes. Findings for the upper secondary students were released in 2002 and reported in *Civic Knowledge and Engagement. An IEA Study of Upper Secondary Students in Sixteen Countries* (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova).

This latest technical report will enable fellow researchers in the field to evaluate published reports, monographs and articles based upon these data and to conduct their own analyses of the international data sets available from IEA. This volume includes detailed information on instrument development, translation procedures, field operations, quality assurance procedures, sampling design, data management, database construction, weighting procedures, scaling methodology and reporting of data from the survey.

IEA Civic Education Study
Technical Report

IEA Civic Education Study Technical Report

Edited by
Wolfram Schulz
and
Heiko Sibberns

with contributors
Pierre Foy
Vera Husfeldt
Rainer Lehmann
Roumiana Nikolova
Judith Torney-Purta

TABLE OF CONTENTS

Chapter 1:

OVERVIEW OF THE IEA CIVIC EDUCATION STUDY

Rainer Lehmann

Twice in its long history of investigating the productivity of educational systems, IEA (the International Association for the Evaluation of Educational Achievement) has authorized surveys of student achievement and attitudes in the domain of civic education. The first such study (Torney, Oppenheim, & Farnen, 1975) was conducted as part of the so-called Six Subject Study, with data collected in 1971 (for a summary, see Walker, 1976). The second, for which a technical description is given in the present volume, was approved by the 35th IEA General Assembly in 1994 in Yokyakarta, Indonesia. The main data collection for this second study took place in 1999, and initial results were published in two volumes (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002; Torney-Purta, Lehmann, Oswald, & Schulz, 2001).

Unlike domains of learning and teaching such as mother tongue, mathematics, and science (and in many countries also foreign languages), civic education does not hold a well-defined place in an international core curriculum. Even within a country, it can be taught as a specific school subject (often under a variety of titles), be relegated to the status of a secondary or implicit area in an existing subject (typically history), or be declared a cross-curricular principle that is to govern all school-based activities. In accordance with this situation, the straightforward aim of the study could not be to measure the "achieved curriculum". Rather, the extremely broad definitions of what civic education could or should be in and across the countries under investigation had to be built into the general design of the study.

GENERAL DESIGN

The IEA Civic Education Study was conceived as a two-phased study, with the two phases referring to:

a) a series of national case studies, designed to investigate, for each participating country, the key issues in the discourse related to the role of civic education in the respective context (Phase 1); and

b) a comparative empirical study, consisting of a test of civic knowledge and a survey of civic-related concepts and attitudes as well as reports on current or expected civic activities (Phase 2).

Even before the case studies were published (Torney-Purta, Schwille, & Amadeo, 1999), information generated in Phase 1 had been used in a consensus process conducted by the International Steering Committee (ICC) with the national research coordinators (NRCs) to sharpen the focus of the empirical study of Phase 2. Three main topic areas were identified as being of

CHARACTERISTICS OF THE IEA CIVIC EDUCATION STUDY

Who Was Assessed?

- *Standard population:* Students in the grade in which most 14-year-olds were found at the time of testing.
- *Older population:* Upper secondary students in the grade where most students of the nationally defined age group (16, 17, or 18) were found at the time of testing.[1]

How Many Were Assessed?

- *Standard population:* Approximately 90,000 14-year-old students from 28 countries were assessed in 1999. In addition, questionnaire data were collected from approximately 9,000 teachers and 4,000 school principals.
- *Older population:* More than 50,000 16- to 18-year-old students at the upper secondary school level from 16 countries were assessed. In addition, questionnaire data were collected from more than 2,000 school principals.

What Was Assessed?

- Civic knowledge, including two sub-domains (civic content knowledge and skills in interpreting political communication), was assessed for both populations; economic literacy was assessed for the upper secondary students only.
- Information was obtained on the home background of the tested students.
- Students' concepts of democracy and citizenship, students' attitudes regarding institutions, minorities, and national identity, and students' civic-related actions were ascertained, as were aspects of civic-related teaching and learning as perceived by students, teachers, and school principals.
- Characteristics of sampled schools and teachers of civic-related subjects were documented.

How Were the Data Collected?

- The main assessment, of two hours, included a cognitive test, background questions, and Likert-type items.
- Students were assessed with tests consisting of 38 multiple-choice items for the 14-year-old students and 43 multiple-choice items for the upper secondary students.
- Students answered a short background questionnaire and a range of Likert-type items covering students' civic-related concepts, attitudes, and actions.
- School principals answered a questionnaire about their school and the teaching of civic education.
- Teachers answered a questionnaire about their professional background and the teaching of civic education.

What Were the Outcomes?

- Cross-country comparisons of 14-year-olds' and upper secondary students' civic knowledge, concepts of democracy and citizenship, and civic-related attitudes and actions.
- Cross-country comparisons regarding the teaching and learning of civic education.
- Analyses of potential predictors of civic knowledge and civic engagement.
- A rich database for secondary analysis of political socialization and civic education with data from 29 countries.
- Comparative civic-related data for students from different grades and age groups in 15 countries.

---

1 Throughout most of this report, these populations are referred to as 14-year-old students and upper secondary students respectively.

fundamental concern in all countries participating in Phase 1:

1. Democracy and Citizenship
2. National Identity and International Relations
3. Social Cohesion and Diversity.

Although these three areas represented the state of affairs in the participating countries in 1997/1998, the general design of the present IEA Civic Education Study has preserved some continuity with respect to the study of 1971. First, the set of criterion variables is not restricted to achievement test results, but also comprises conceptual, attitudinal, and motivational variables. Second, the aspects of civic education under investigation, though not identical between the two studies, are sufficiently similar to allow for occasional comparison at the item level. However, statistical and technical advances in the almost three decades between the two studies (as well as advances in research in the social sciences and education) are clearly reflected in the present study's design, analysis of data, and presentation of results.

Originally, it was envisaged that no country could participate in Phase 2 of the study unless it had provided a national case study chapter by the publication deadline for the Phase 1 volume (Torney-Purta et al., 1999). A few countries received final permissions and/or funding for participation after the Phase 1 book was completed, but were accepted for participation in Phase 2, particularly since some of them were likely to provide data-facilitating continuity between the 1971 and 1999 studies.

Continuity between the two studies was also a motive in defining the primary target population as the grade in which most 14-year-olds were enrolled. The 1971 study, as well as research in this area from Germany and Switzerland (Fend, 1991), had demonstrated the existence and interpretability of civic-related variation between students of that age. However, the evaluators of the proposal for the current study presented to the German Research Association (*Deutsche Forschungsgemeinschaft: DFG*) as the principal funder insisted that at least one older student population should also be tested and surveyed because, in their opinion, relevant aspects in the emergence of civic-related characteristics could better be studied at a late point in the development of adolescents.

The assessment of a population older than 14, upon which the funding of the international coordination of the study was made dependent, created a situation that entailed substantial problems for the data analysis. The countries had based their decision to participate on the original proposal, which required national support for a study of 14-year-olds. The national research centers had to be urged to seek such support also for a target population from the upper secondary school, preferably with a modal age of 17. With a great deal of goodwill and at considerable expense, 16 countries identified a national target population. Unfortunately, Germany was not one of them. As a consequence of the attempts to save funding for the study as a whole occasioned by the addition of this group, there is no internationally standardized definition of target populations at the upper secondary level, but rather considerable variation between the national samples of different countries, both in terms of

mean age (between 16.6 and 19.4) and coverage of the age cohort. The 14-year-old samples all have mean ages between 14.1 and 15.3 and high coverage (especially as this age bracket is just before the age when students begin leaving school in large numbers in the participating countries).

Table 1.1 lists the countries that contributed national case studies to Phase 1 of the study as well as the countries participating in the comparisons of 14-year-olds and upper secondary students. For more detail as to the development of the general design into a fully orchestrated study, see Chapter 2 of this report.

Table 1.1 Countries participating in the IEA Civic Education Study

Country	National Case Study	14-year-olds	Upper Secondary Students
Australia	•	•	
Belgium (French)	•	•	
Bulgaria	•	•	
Canada	•		
Chile		•	•
Colombia	•	•	•
Cyprus	•	•	•
Czech Republic	•	•	•
Denmark		•	•
England	•	•	
Estonia		•	•
Finland	•	•	
Germany	•	•	
Greece	•	•	
Hong Kong (SAR)	•	•	•
Hungary	•	•	
Israel	•		•
Italy	•	•	
Latvia		•	•
Lithuania	•	•	
Netherlands	•		
Norway		•	•
Poland	•	•	•
Portugal	•	•	•
Romania	•	•	
Russian Federation	•	•	•
Slovak Republic		•	
Slovenia	•	•	•
Sweden		•	•
Switzerland	•	•	•
United States	•	•	
Number of countries	**24**	**28**	**16**

THEORETICAL FRAMEWORK

We have already mentioned that civic education is much less rigidly anchored in existing curricula than core school subjects such as mathematics or mother-tongue instruction. We should add that schools have much less of a monopoly to convey principles and specifics of civic education than to provide content in other subjects such as mathematics. This consideration implied the choice of a theoretical framework that recognizes the multiple influences to which children and adolescents are subject. The theory of ecological development (Bronfenbrenner, 1988) and, to some degree, the theory of situated cognition (Lave & Wenger, 1991; Wenger, 1998) appeared to provide the respective reference systems that could be translated into a model governing the study (see Figure 1.1).

Figure 1.1 Model for the IEA Civic Education Study
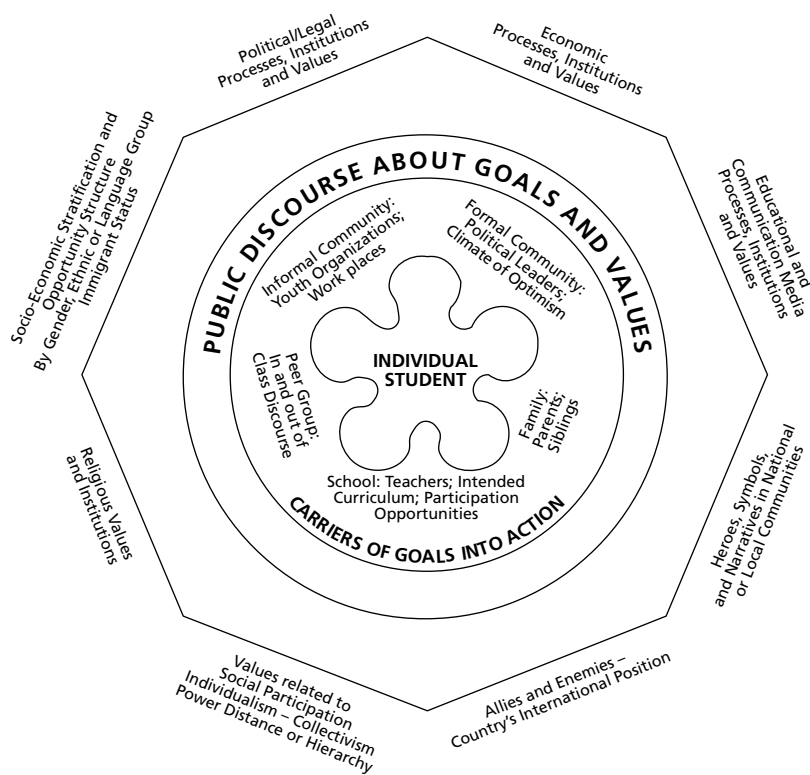
Whereas the perimeter of this model, including the reflection of public discourse about goals and values, was very much covered by the national case studies of Phase 1 of the study, an attempt was made in Phase 2 to mirror the influences of those who carry the goals into action. For this purpose, more specific theoretical contributions of sociology and political science could be drawn upon (e.g., Almond & Verba, 1965; Barnes et al., 1979; Inglehart, 1997; Kaase & Newton, 1995; Niemi & Junn, 1998). This body of literature provides much evidence (on both adolescent and adult members of the civic communities) to which the data produced by the present study can be related.

The national case studies of Phase 1 were instrumental not only in conceptualizing the study, but also in serving as backgrounds against which to interpret the findings from Phase 2, particularly in those cases where the latter

turned out to be unexpected. Theoretical references, such as the ones listed above, may serve to provide explanatory links between the case studies and specific findings from the survey, although much of this work still needs to be done.

DIMENSIONS OF STUDENT ASSESSMENT

As was stated above, care was taken not to limit the study to aspects of civic knowledge keyable as "right" or "wrong" but to give ample space to the investigation of the conceptual, attitudinal, and motivational characteristics of students. Although the details of measuring the respective constructs and of ascertaining their interrelationships are discussed in more detail in separate chapters of this report, some of the key issues encountered need to be mentioned in this present overview.

With respect to the *test of civic knowledge*, it was deemed important to tap both the students' factual or content knowledge and their ability to interpret civic-related information (for technical details, see Chapters 2 and 7 below). However, for a long time during the process of test construction, it was far from clear whether these two test components would be best understood under the assumption of a single dimension or under a two-dimensional model reflecting the course of item development and selection. Initial analyses suggest that both assumptions are defensible on the basis of existing evidence, but it is equally clear that a great deal of further research is needed before separate hierarchies of competency levels can be defined for civic content knowledge and interpretative skills respectively.

While this problem applies to both the 14-year-olds and the upper secondary students, the situation is even more complex with respect to the latter. In this case, items intended to measure economic literacy (cf., Walsted & Soper, 1989) were included, and again it was not clear from the beginning whether these would turn out to constitute a separate dimension of the test. It seemed preferable to assume that this would, indeed, be the case, but it was also clear that much additional research would be required before specific competency levels could be defined. Moreover, decisions on the dimensional structure of a test can rarely be reduced to the status of a clear-cut psychometric problem with a single—and simple—solution. Opting for a more complex structure therefore appears justifiable, if there is at least some evidence supporting it, and if the differential information thus obtained helps to solve theoretical and/or practical problems.

This ambivalence of psychometric results with respect to the definition of theoretically meaningful constructs is particularly obvious with students' concepts of democracy, citizenship, and the scope of government. To understand the tacit assumptions or beliefs underlying the students' responses to other parts of the survey, the students were asked to give their opinions on states or events "good for democracy", the importance of certain actions for being a good citizen, and the responsibilities governments should have. The items were derived from various concepts of democracy, citizenship, and government as distinguished by political scientists. Such adult differentiations found little substance in the younger students' reactions: 14-year-old students

tend to have rather vague notions of democracy, centered around such concepts as equality and liberty, with little concern for modes of participation or the scope of formal as opposed to substantive equality. It is only at a later stage that the respective distinctions appear to become important. Thus, decisions as to which scales should be defined and reported are not merely related to psychometric issues.

Many of the attitudinal items used in the study were similar to those used in earlier studies, or had even been used before. It is, however, a rather innovative feature of the present study that all of these data were scaled according to methods developed within the framework of Item Response Theory (IRT), more specifically according to partial-credit test models. We believe that this particular approach renders considerable additional insights over and above what conventional test analyses can do, let alone reports on single items. Given that many researchers in the field are relatively unfamiliar with this method, a reasonably extensive treatment of the respective analyses is in order (see Chapters 8 and 9 below).

The following scales were found to meet both the technical and the substantive criteria of being considered worth reporting in the volumes of basic analysis (Amadeo et al., 2002; Torney-Purta et al., 2001):

• Importance of Conventional Citizenship

• Importance of Social-movement-related Citizenship

• Concept of Economy-related Government Responsibilities

• Concept of Society-related Government Responsibilities

• Trust in Government-related Institutions

• Positive Attitudes toward One's Nation

• Attitudes toward Women's Political and Economic Rights

• Positive Attitudes toward Immigrants

• Confidence in School Participation

• (Expected) Conventional Political Activities

• Open Climate for Classroom Discussion.

A number of other groups of items were included in the survey but were not reported in the original reports because of limitations of time and resources.

Since the initial reporting of results, considerable additional effort has been spent on refining the scales. Chapter 8 provides the information required to evaluate the findings presented in the international descriptive reports as well as information on desirable refinements in future research, including secondary analyses of the data. See also www.wam.umd.edu/~iea, where the full survey instrument is available and where secondary analysis conducted by the IEA team is referenced on an ongoing basis.

THE CONTEXT QUESTIONNAIRES

From the very beginning of international comparisons of student achievement and related characteristics, it has been customary to collect information on students' home backgrounds as well as school principal and teacher data on school characteristics and the teaching process.

The collection of student background data is imperative if issues of the equality or inequality of educational opportunities are to be addressed, this being a criterion of international comparisons in its own right. Many of the variables used in the respective sections of the student questionnaire have been used successfully before, for example the number of books in the home as a proxy for home literacy resources. Similarly, attempts to measure the "implemented curriculum" through teacher questionnaires have been considered important for most studies in this area, even though the explanatory power of such information has often been incomplete.

Some efforts have been made to assess the explanatory relevance of the information thus collected. There have been constraints of time and staff both at the ICC for the study, located at Humboldt University, Berlin, Germany, and at the University of Maryland, College Park, United States (where an IEA-related team has been established by the Chair of the International Steering Committee). However, some in-depth analyses have been conducted (see, for example, Torney-Purta, 2002; Torney-Purta & Amadeo, 2003; Torney-Purta & Richardson, 2002). The release of the international data sets pertaining to the IEA Civic Education Study, buttressed by the present technical report, is meant to encourage the research community to undertake further analysis of the data.

SUMMARY OF THE REPORT

The present technical report for the IEA Civic Education Study, for which the main data were collected in 1999 and 2000, is meant to enable fellow researchers in the field to evaluate published reports, monographs, and articles based upon these data and to conduct their own analyses of the international data sets made available along with this report. To achieve this aim, the key elements in the conduct of this study are described with the necessary technical detail.

The present overview is followed by information on the development of the instruments for the study (Chapter 2), and then descriptions of the translation procedures (based upon English-language master versions of all instruments), the field operations, and the quality assurance procedures required by IEA (Chapter 3). An equally rigorous requirement for any study conducted under the auspices of IEA is given by the standards for probability sampling, details of which are presented in Chapter 4. For data management and the construction of databases, IEA studies can rely on the high quality of routines developed within the IEA Data Processing Center (DPC) in Hamburg, Germany, an institution that provided the services for the present study (see Chapter 5), including the verification of sampling routines and the corresponding weighting (Chapter 6). Responsibilities for the scaling of cognitive items—the civic knowledge test—was shared between the IEA DPC

in Hamburg and the ICC in Berlin; the respective technical information is presented in Chapter 7. The scaling of Likert-type items representing the students' concepts, attitudes, and (expected) actions was conducted at the ICC; the technical detail is given in Chapters 8 and 9. Chapter 10 concludes the report with information on the handling and reporting of information from the student and teacher questionnaires.

We hope that this information will not only enable, in a technical sense, the research community to conduct secondary analyses, but will also contribute, albeit indirectly, to improvements in the domain of civic education in many countries around the world.

REFERENCES

Almond, G. A., & Verba, S. (1965). *The civic culture: Political attitudes and democracy in five nations.* Boston, MA: Little, Brown and Company.

Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries.* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Barnes, S. H., Kaase, M., Allerbeck, K. R., Farah, B. G., Heunks, F., Inglehart, R., Jennings, M. K., Klingemann, H. D., Marsh, A., & Rosenmayr, L. (1979). *Political action: Mass participation in five Western democracies.* Beverly Hills, CA: Sage Publications.

Bronfenbrenner, U. (1988). Interacting systems in human development. In N. Bolger, C. Caspi, G. Downey, & M. Moorehouse (Eds.), *Persons in context: Developmental processes* (pp. 25–49). Cambridge, England: Cambridge University Press.

Fend, H. (1991). *Entwicklungspsychologie der Adoleszenz in der Moderne. Band 2: Identitätsentwicklung in der Adoleszenz-Lebensentwürfe, Selbstfindung und Weltaneignung in beruflichen, familiären und politisch-weltanschaulichen Bereichen (Developmental psychology of adolescents in modern times. Vol 2: Development of identity in the adolescent: Plans for life, self-cognition and world appropriation in professional, familiar and political-ideological domains).* Bern/Stuttgart/Toronto: Verlag Hans Huber.

Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic and political change in 43 societies.* Princeton, NJ: Princeton University Press.

Kaase, M., & Newton, K. (1995). *Beliefs in government* (Vol. 5). Oxford: Oxford University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Niemi, R., & Junn, J. (1998). *Civic education: What makes students learn?* New Haven, CT: Yale University Press.

Torney, J. V., Oppenheim, A. N., & Farnen, R. F. (1975). *Civic education in ten countries: An empirical study.* Stockholm/New York: Almqvist & Wiksell International/John Wiley and Sons.

Torney-Purta, J. (2002). Patterns in the civic knowledge, engagement, and attitudes of European adolescents: The IEA Civic Education Study. *European Journal of Education, 37*(2), 129–141.

Torney-Purta, J., & Amadeo, J. (2003). A cross-national analysis of political and civic involvement among adolescents. *Political Science and Politics, 36,* 269–274.

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

Torney-Purta, J., & Richardson, W. (2002). An assessment of what fourteen-year-olds know and believe about democracy in twenty-eight countries. In W. Parker (Ed.), *Education for democracy: Contexts, curricula, assessments* (pp. 185–209). Greenwich, CT: Information Age Publishing.

Torney-Purta, J., Schwille, J., & Amadeo, J. (1999). *Civic education across countries: Twenty-four national case studies from the IEA Civic Education Project.* Amsterdam and Washington, D.C.: IEA and National Council for Social Studies.

Walker, D. A. (1976). *The IEA Six Subject Survey: An empirical study of education in twenty-one countries.* Stockholm/New York: Almqvist & Wiksell International/John Wiley and Sons.

Walsted, W. B., & Soper, J. C. (1989). What is high school economics? Factors contributing to student achievements and attitudes. *Journal of Economic Education, 20*(Winter), 23–38.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity.* Cambridge: Cambridge University Press.

Chapter 2:

DEVELOPMENT OF THE CIVED INSTRUMENTS

Vera Husfeldt and Judith Torney-Purta

In the first section of this chapter we review the two-year process of identifying a common core of topics to form a content framework relating to citizenship and democracy valid across the 28 countries that participated in the IEA Civic Education Study (CivEd) of 14-year-old students. It will also detail the three-year process of developing a fair and valid test (items designed with keys for correct answers) and a survey (items assessing attitudes or beliefs for which there are no correct answers) that met IEA standards. In the second section we describe the development of the test and survey of the upper secondary students.

FRAMEWORK DEVELOPMENT

The Phase 1 national case studies were the basis for Phase 2 of the study, and their results provided the material from which the testing framework was developed. This framework is similar to the intended curriculum on which tests in other IEA studies have been based.

The data collected during Phase 1 included summaries of what panels of experts in participating countries believed 14-year-old students should know about 18 topics relating to democratic institutions. These topics included elections, individual rights, national identity, political participation, the role of the police and the military, organizations that characterize civil society, relation of economics to politics, and respect for ethnic and political diversity (Torney-Purta, Schwille, & Amadeo, 1999). The majority of countries submitted the following (amounting to more than 100 pages for several countries):

1. A proposal for Phase 1.
2. A review of existing research literature.
3. A set of summaries responding to 18 case-study framing questions detailing their importance and basic information about civic education in the country.
4. A set of summaries relating to the three core international domains and one relating to the three optional domains (see detail below). This material included more detailed information about textbooks, curriculum objectives, teaching, and the role of community and out-of-school groups.
5. A chapter-length case study summarizing the material and dealing with an issue of particular importance to civic education in the country.
6. A methodology questionnaire detailing the methods used and listing members of the National Expert Panel.

Following the analysis of available Phase 1 material from more than 20 countries and voting by the National Project Representatives on the most important of the 18 original topics, we chose, in 1995, three domains for extensive study as "core international domains". Early in the study it was clear that there was a common core of topics and concepts that experts in these countries believed 14-year-old students should understand. These international core domains were:

Domain I: Democracy

What does democracy mean and what are its associated institutions and practices?

The three sub-domains were:

A. Democracy and its defining characteristics

B. Institutions and practices in democracy

C. Citizenship—rights and duties.

Domain II: National Identity, Regional and International Relationships

How can the sense of national identity or national loyalty among young people be described and how does it relate to their orientation to other countries and to regional and international organizations?

The two sub-domains were:

A. National identity

B. International/regional relations.

Domain III: Social Cohesion and Diversity

What do issues of social cohesion and diversity mean to young people and how do they view discrimination?

We also identified three other issues as important: the media, economics, and local problems (including the environment). These topics were also identified as important for Phase 2 but were explored less systematically during this stage.

As a next step in developing a content framework, personnel at the Phase 1 Coordinating Center read the case study documents. They developed general statements about what young people might be expected to know and believe about the three domains, and they elaborated on and illustrated these with quotations from the national case studies. This material formed the *Content Guidelines for the International Test and Survey*, which served as a concise statement of content elements in the three domains that were important across countries (see Appendix B). These guidelines provided a focus for those writing the test items. It was clear from the case study material that the greatest emphasis in the test should be on Domain I: Democracy, Democratic Institutions, and Citizenship.

In addition to giving input on the content domains to be covered, the national research coordinators (NRCs) helped define, in 1996, the types of items to include in the instrument. These were:

- items assessing *knowledge of content*; and
- items assessing *skills in interpretation* of material with civic or political content (including short text passages and cartoons).

These items formed the *test* and had keyed correct answers.

Because civic education is an area where students' content knowledge and skills are important but not the sole focus, the NRCs suggested three other item types:

- items assessing students' understanding of *concepts* such as democracy and citizenship;
- items assessing students' *attitudes* (e.g., feelings of trust in the government); and
- items assessing students' current and expected participatory *actions* relating to politics.

These items formed the *survey* and did not have correct answers.

Intersecting these five item types with the three study domains produced the matrix in Table 2.1, which served as the basis for the test and survey design.

Table 2.1 Matrix of item types and study domains

Item Type	Knowledge of Content	Skills in Interpretation	Concepts	Attitudes	Actions
Domain I Democracy/ Citizenship					
Domain II National Identity/ International Relations					
Domain III Social Cohesion and Diversity					

A little less than half of the testing time was devoted to a test that included cognitive items that could be "keyed" with correct and incorrect answers. Slightly less than half of the remaining testing time was devoted to a survey that included non-keyed items that assessed concepts, attitudes, and actions. The rest of the instrument asked about students' perceptions of classroom climate and their confidence in participation at school, and obtained background information (including home literacy resources and the associations or organizations to which students belonged). A short period at the end of the second testing session was reserved for countries to administer nationally developed items.

THE PROCESS OF TEST AND SURVEY DEVELOPMENT FOR 14-YEAR-OLD STUDENTS

Because there were no large existing sets of items likely to yield the number of items needed to fill in the matrix, extensive item writing was required. We began by reviewing materials in the content guidelines, other summaries of Phase 1 documents, and messages exchanged during an online conference on civic issues conducted with secondary school students in seven countries. We next invited all NRCs to submit items. Our third task was to review the 1971 CivEd instrument, released items from the United States and Canadian assessments, and the published research literature. Members of the International Steering Committee (ICC) then wrote items, which were subsequently entered into an item database keyed to the content guidelines. Our fifth step involved asking groups of test specialists and content experts to review items in the database and their relation to the content framework. All NRCs whose countries were officially participating by September 1997 reviewed items before the pre-pilot testing; all NRCs whose countries were officially participating by November 1998 reviewed items before their inclusion in the final version of the test.

The result of this activity was the development of 140 knowledge and skills items (*content knowledge* and *skills in interpretation*), each with one correct answer and four distracters, and each of which was entered into the database for the 14-year-old population. All the items were judged suitable for administration in the participating countries.

The items focused on principles, pivotal ideas, and general examples, and not on the details of the political arrangements in any one country. For example, Content Knowledge/Domain I items covered the principles of democracy and its associated institutions across the countries participating in the study. The test did not include items about specific mechanisms of the electoral process or government structure in any particular country. The Content Knowledge/ Domains II and III items likewise dealt with internationally relevant or generalized matters shared across countries. This emphasis differs from that in many national tests where items about each country's political structure predominate. The CivEd Phase 2 items are based on information gathered during Phase 1 about what students are expected to know and on expert statements such as that issued under the auspices of the Council of Europe about the role of history knowledge in civic education (Slater, 1995, pp. 146– 148).

Some of the *skills in interpretation* items asked students to distinguish between statements of fact and opinion. Others were based on a leaflet of the type issued during an election campaign, on the interpretation of a short article from a mock newspaper, or on a political cartoon. The general ideas for cartoons came from those published in newspapers. They were redrawn to communicate a single message that 14-year-olds across countries could be expected to understand.

Pre-Piloting of Items of Content Knowledge and Skills in Interpretation

In September 1997, 80 items assessing *content knowledge* and *skills in interpretation* (multiple-choice items with correct answers) were sent to the NRCs for a pre-pilot. The expert panels reviewed items and the national centers then obtained convenience samples of 14-year-old students for two hours of testing. The sample sizes for the 20 participating countries ranged from 185 in Cyprus to 626 in Belgium (French).

A week-long meeting of NRCs was held in March 1998. Here, the *pre-pilot* item statistics from 20 countries were made available (Belgium (French), Bulgaria, Colombia, Cyprus, Czech Republic, Estonia, Finland, Germany, Greece, Hong Kong Special Administrative Region (SAR), Hungary, Italy, Lithuania, Norway, Poland, Portugal, Russia, Slovenia, Chinese Taipei, and the USA). NRCs from all these nations were in attendance, as were coordinators from countries planning to participate in Phase 2 who had not been able to pre-pilot (Australia, Chile, Romania, Slovak Republic, Switzerland). (In some cases these countries had been able to obtain an expert review of the items.)

A questionnaire about the 80 items was circulated to the meeting participants on the first day of the meeting to ascertain which items were likely to be problematic. The 80 items were listed, and each NRC was asked to indicate (according to his or her pre-pilot experience and the National Expert Panel review) how much discussion the item required. There were 11 items for which more than five of the 18 responding countries answered that the item required "extensive discussion". These items turned out to be the same items that had attracted, during the pre-pilot process, complaints about unsuitability for the national context or difficulties of translation. They were very briefly raised in the meeting and were then deleted from further consideration.

Many of the remaining 69 items had at least one distracter that was performing poorly in many countries, and so it was decided to change all items from five choices to four. Thus, as the remaining items were discussed, one distracter for each was cut (in accordance with the recommendation of the NRCs). Each of the 69 items was also discussed (with the text of the item shown on a transparency), and suggestions made for revision. Consensus was achieved on retaining 62 of these items as revised (and cutting seven). During the subsequent session, about a dozen "new" items were proposed (prepared by groups of NRCs, who felt there were gaps in the instrument's coverage). Six of these "new" items were further revised in a plenary session and accepted.

The surviving 68 items were sorted into two forms, A and B, in an attempt to avoid duplication, to balance item difficulties, and to include approximately equal numbers of *content knowledge* and *skills in interpretation* items in each form. Of the 62 items that had been piloted, two were so extensively revised as to be essentially new items. Thus, the pre-pilot data on 60 of the items included in forms A and B were analyzed to check the reliabilities of 30 of the items from each form. These were acceptable (in the low .80s for the large majority of countries).

Piloting of Items Assessing Content Knowledge and Skills in Interpretation and the Resulting Final Test

Between April and October 1998, 25 countries conducted pilot studies on the two forms of the test (items assessing *content knowledge* and *skills in interpretation*, described above) and the survey (items assessing *concepts, attitudes* and *actions*, described below). In each country, judgment samples of about 200 students were tested (two class periods per student). The pilot countries included Australia, Belgium (French), Bulgaria, Chile, Chinese Taipei, Colombia, Cyprus, Czech Republic, Estonia, Finland, Germany, Greece, Hong Kong (SAR), Hungary, Italy, Latvia, Lithuania, Norway, Poland, Portugal, Romania, the Russian Federation, Slovenia, Switzerland, and the United States. In addition to these countries, Denmark, England, the Slovak Republic, and Sweden participated in the final testing of 14-year-old students. (Chinese Taipei was unable to obtain funding to continue its participation after the pilot testing.)

The NRCs were provided with item statistics for their countries (difficulty, discrimination, and differential item functioning). They discussed each item within its content category at a November 1998 meeting. The small number of items that was viewed as unacceptable by more than one-fifth of the NRCs was dropped in accordance with the rule used by IEA to promote test fairness across countries. Through a process of negotiation, the research coordinators chose, by consensus, 38 items assessing *content knowledge* and *skills in interpretation* from the 68 that had been piloted. The discrimination indices were greater than .30 for most of these items. In addition to psychometric properties, coverage of the content framework and the research coordinators' preferences were decisive factors in the process of item selection (see Appendix B for a listing of items by content category).

The ratios of number of "items written" to "number piloted" to "number accepted" were similar to IEA tests in other subject areas. Confirmatory factor analysis and IRT modeling indicated a high-quality test across countries. These modern scaling methods (Frederikson, Mislevy, & Bejar, 1993) were our primary guide as we developed the test. Classical scaling methods also indicated a test of high quality. The alpha reliabilities for the final 38-item civic education test exceeded .85 in each of the countries.

Piloting of Items Assessing Concepts, Attitudes, and Actions and the Resulting Final Survey

At the March 1998 meeting, the NRCs reviewed lists of suggested topics for items assessing *concepts, attitudes* and *actions* and some prototype items. Most item sets for piloting were suggested by the research literature. Some revisions were necessary to adapt items originally designed for administration to adults in an interview, and a "don't know" option was added to each of these items.

In mid-1998, the research coordinators piloted the survey items along with two forms of the knowledge and skills test. Items for the survey were chosen through a process of negotiation similar to that described in the previous section. The final survey included 52 items assessing *concepts* (concepts of democracy, citizenship, and government), 62 items assessing *attitudes* (social and

political attitudes), and 22 items assessing *actions* (civic and political actions past, present, and expected in the future). Items assessing student background, school experience, organizational membership, and peer group involvement were also included. Policy in some of the participating countries prohibited questions about families' social or political values, and no such items were included. The final test and survey were designed so that they could be administered in two class periods. The texts of all of the items assessing *concepts, attitudes* and *actions* and about half of the items assessing *content knowledge* and *skills in interpretation* were released for use by other researchers (see Appendix F and also http://www.wam.umd.edu/~iea/).

The way in which the concepts of democracy items were developed is illustrative of the development of the survey items. We began with lists of such items used by others and circulated them in the spring of 1997 to NRCs. We also consulted the political science literature (Beetham, 1993; Held, 1996; Janoski, 1998). We developed a draft laying out several contrasting models of democracy and held several online conferences using the University of Maryland's ICONS World Wide Web model to debate these models. The models and items were further refined during the March 1998 NRC meeting.

The models included the Generic or Rule of Law Model, the Liberalism Model, the Pluralism Model, the Participation Model, the Communitarian Model, the Social Welfare Model, and the Elitism Model. Items were written to cover these models (each item usually relating to more than one model); they were phrased to ascertain whether respondents believed that a given situation or condition would be good or bad for democracy. A sample item is: *When many different organizations [associations] are available [exist] for people who wish to belong to them, that is _____ [alternatives: very good for democracy, good for democracy, bad for democracy, very bad for democracy].* This formulation allowed us to include negative as well as positive attributes (e.g., statements describing nepotism). Thirty-nine items in the pilot and 25 in the final instrument were concerned with the concept of democracy. A similar process involving a review of a range of possible items, the inclusion of 21 items in the pilot, and the final choice of 15 items for the test was used to develop the concept of citizenship item set. Content Domains I, II, and III were covered in both the concept of democracy and the concept of citizenship items.

A similar answer format (with four choices, such as "strongly agree", "agree", "disagree", and "strongly disagree", and a separate "don't know" response) was used for the large majority of the survey. Attitude items were rewritten from versions used with adults to make them shorter, more easily translated, and more suitable for 14-year-old students.

The development of short survey instruments for teachers and for school heads (principals) began at the March 1998 meeting and covered the same content domains as the student instrument, but also included questions about school context and instruction. These instruments were piloted in the same countries and at the same time as the student instruments.

The Process of Test and Survey Development for Upper Secondary Students

The test and survey for the upper secondary population was built directly on the instrument developed for the 14-year-old students. The four domains of civic education content treated in the test and survey of the upper secondary students were:

- Democracy and its associated institutions, including the rights and responsibilities of citizens.
- National identity, including relations between groups within countries and relations with other countries.
- Social cohesion and diversity.
- Economics.

For the test (keyed items), the first three domains were the same as those covered in the 14-year-old version; economics was new to the upper secondary version.

All items for the survey and many items for the test were adopted from the instruments for the 14-year-old population. The test was formed of the items assessing *knowledge of civic content* and *skills in interpretation* of material with civic or political content (including short-text passages and cartoons), as well as 14 items about economic literacy. A number of these items were modeled after the *Test of Economic Literacy* (Beck & Krumm, 1999).

The survey included items assessing students' concepts of democracy and citizenship, students' *attitudes* (e.g., feelings of trust in the government), and students' current and expected participatory *actions* relating to politics. In addition to the items that had been used for the 14-year-old students, some items on the *effectiveness of political action* and on the justifications for the *use of military force* were included. As in the 14-year-old instrument, the survey items had four response alternatives, each measuring the extent of endorsement. Similar to the data collection process for the 14-year-old students, the first half of the assessment time was scheduled for the test and the second half for the survey. The second half also included questions about students' perceptions of classroom climate, their confidence in participation at school, and background information. In contrast to the study of 14-year-old students, the study of upper secondary school students did not include a teacher questionnaire.

Piloting and the Resulting Final Test

The first pilot and final testing of the 14-year-old students showed that, although the difficulty levels of the cognitive items were generally adequate for this age group, a test composed only of these items would have been too easy for students who were two to four years older. Therefore, some new items had to be developed. A pilot test for the upper secondary students conducted in the spring of 1999 (see next paragraph) served to test these newly developed cognitive items as well as some attitude items. The final questionnaire for the upper secondary students included all attitude items from the final questionnaire for the 14-year-old students and two new eight-item scales.

In addition, the background questionnaire, which provided demographic information as well as information pertaining to the students' families and out-of-school activities, was adapted from the test for the 14-year-old students.

Between April and May 1999, six countries conducted classroom-based pilot studies of the test and survey. Samples of about 100 to 500 students in Grades 10 to 12 were tested. The pilot countries included Chile, Colombia, Estonia, Germany (which eventually decided not to participate in the main study of upper secondary students), Israel, and the Russian Federation.

In addition to undertaking the pilot study, the NRCs were asked to evaluate the newly developed cognitive items with respect to translation problems and structures, contexts, and school instruction effects, and to estimate their difficulty.

In July 1999, the analysis of the pilot data of upper secondary school students was performed, and a sub-group from the ICC and the NRCs made a final decision about which items to include. Those research coordinators who were unable to pilot the test agreed to abide by choices made in relation to the six countries. This was feasible because all non-piloting countries planning to participate at the upper secondary level had participated at the 14-year-old level using tests with a high overlap of items.

SUMMARY

IEA studies include collecting some information from participating countries before the formulation of a test instrument. The IEA Civic Education Study had a particularly intensive and extensive first phase during which detailed information was collected from participating countries about 18 topics that might be included in civic education programs. The ICC, the NRCs, and members of the IEA research centers at the University of Maryland and the Humboldt University of Berlin participated in an iterative process of consultation based on written documentation and utilizing email and face-to-face meetings to form a consensus about the content of the test and survey instruments (for 14-year-old students). This resulted in a strong basis upon which to build pilot instruments in a subject area where there had previously been little agreement about whether a common core of expectations for learning exists across democratic countries.[1] A similar (though shorter) process was undertaken for the development of the teacher and school questionnaires and the development of the test and survey for the upper secondary population.

---

1 Two books published by IEA detail this process: Torney-Purta et al. (1999) and Steiner-Khamsi, Torney-Purta, and Schwille (2002). LeTendre (2002) compares the use of case study material in the IEA Civic Education Study favorably to its use in TIMSS, especially because the case studies had a place in the process that allowed them to feed directly into the study and instrument design.

REFERENCES

Beck, K., & Krumm, V. (1999). *Wirtschaftskundlicher Bildungtest (WBT) (Manual)*. Goettingen: Hoegrefe.

Beetham, D. (1993). Key principles and indices for a democratic audit. In D. Beetham (Ed.), *Defining and measuring democracy* (pp. 25–43). Bakersfield, CA: Sage Publications.

Frederikson, N., Mislevy, R., & Bejar, I. (1993). *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Held, D. (1996). *Models of democracy*. Stanford, CA: Stanford University Press.

Janoski, T. (1998). *Citizenship and civil society: A framework of rights and obligations in liberal, traditional and social democratic regimes*. Cambridge: Cambridge University Press.

LeTendre, G. (2002). Advancements in conceptualizing and analyzing cultural effects in cross-national studies of educational achievement. In A. C. Porter and A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 198–228). Washington, D.C.: National Academy Press.

Slater, J. (1995). *Teaching history in the new Europe*. London: Cassell.

Steiner-Khamsi, G., Torney-Purta, J., & Schwille, J. (2002). *New paradigms and recurring paradoxes in education for citizenship*. Amsterdam: Elsevier Press.

Torney-Purta, J., Schwille, J., & Amadeo, J. (1999). *Civic education across countries: Twenty-four national case studies from the IEA Civic Education Project*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Chapter 3:

TRANSLATION PROCEDURES, FIELD OPERATIONS, AND QUALITY ASSURANCE [1]

Wolfram Schulz and Roumiana Nikolova

INTRODUCTION

The IEA Civic Education Study (CivEd), which involved 29 countries and two student populations with over 140,000 students, was guided by the International Steering Committee (ISC) and coordinated by the International Coordinating Center (ICC). However, the national research centers in each participating country were responsible for implementing the design of the study in accordance with the international standards. Survey and test instruments, as well as field procedures, were developed through an extensive process of cooperation and consensus among the participants. The data collection within countries was carried out by the national research centers.

Each national research team was responsible for translating the international instruments and procedures and adapting them to local conditions, implementing the field operation instructions, and organizing data collection and quality control operations. This combination of international cooperation and national implementation is an efficient and cost-effective approach to conducting international comparative studies, and it has become the standard for IEA studies. It requires close collaboration among participants, and the validity and reliability of the results depend to a large extent on each participant adhering to the prescribed procedures at all times.

The ISC and the ICC made considerable efforts to develop standardized materials and procedures to ensure that the data collected in all 29 countries would be comparable. This chapter describes these efforts, which include the processes of instrument translation and translation verification, the implementation of field operations, and the use of quality assurance measures during the testing period. Quality assurance standards for IEA studies require monitoring of instrument translation, standards for test administration, and measures of quality control during the process of data collection.

---

1 Many procedures described here and parts of this chapter were adapted from the IEA TIMSS study; see O'Connor and Malak (2000) and Gonzalez and Hastedt (2000).

Questionnaire Translation

The CivEd instruments were prepared in English and translated into 24 additional languages across the 29 participating countries (see Tables 3.1 and 3.2 for the list of languages). In addition, it sometimes was necessary to adapt the international (original) version for cultural purposes, even for the three countries that tested in English. When developing the standardized instruments across languages and countries to provide for comparability of the CivEd data, it was essential to ensure the quality of the translations and to carry out the translations as carefully and as thoroughly as possible. Explicit guidelines for translation and cultural adaptation were developed, and verification procedures were implemented. Specifically, the instrument translation procedures included development of the following:

- A translation manual with explicit guidelines for translation and cultural adaptation, translation operations, and guidelines for layout and assembly.
- Translation of the instruments by the national centers in accordance with the guidelines.
- Verification of the quality of the translations and booklet layout by independent translators, who suggested changes to the translations.
- A process whereby the verified translations were sent back to the national centers for reviewing.
- A process to monitor the implementation of changes suggested by the verifiers.

The translation guidelines developed for this study and described below are documented in the *Translation Manual* (IEA Civic Education Study, 1998a, 1999e). The guidelines recommended that for the translation of test items and survey questions, each national center engage two independent translators of each instrument. A translation team then reviewed each instrument.

In the main study, as was the case in the field trial, the national centers were instructed to apply the following procedures:

- For each target language, a minimum of two translators, preferably civic education specialists.
- Two independent forward translations or revisions of each item and instrument, which then had to be reviewed and compared by the translation review team.
- All deviations and cultural adaptations made to the instrument to be reported on the Translation Deviation Form (see Appendix C).

The national research centers were also required to translate the student questionnaire, the teacher questionnaire, the school questionnaire, the *Test Administrator Manual* (IEA Civic Education Study, 1999c, 1999i), including the test administration form, and the *School Coordinator Manual* (IEA Civic Education Study, 1999b, 1999h). International guidelines for translations assisted the national centers to develop reliable translated versions of these instruments and manuals. At a more specific level, the translators' work included:

- identifying and minimizing cultural differences;

Table 3.1 Languages and instrument verification in participating countries (14-year-old students)

Country	Language	Verification Procedure
Australia	English	Adaptations verified before testing
Belgium (French)	French	Test with unverified instrument*
Bulgaria	Bulgarian	Translation verified before testing
Chile	Spanish	Translation verified before testing
Columbia	Spanish	Translation verified before testing
Cyprus	Greek	Translation verified before testing
Czech Republic	Czech	Translation verified before testing
Denmark	Danish	Translation verified before testing
England	English	Adaptations verified before testing
Estonia	Estonian	Translation verified before testing
Estonia	Russian	Russian instrument with adaptations
Finland	Finnish	Translation verified before testing
Finland	Swedish	Translation verified before testing
Germany	German	Translation verified before testing
Greece	Greek	Test with unverified instrument
Hong Kong (SAR)	Chinese	Translation verified before testing
Hungary	Hungarian	Translation verified before testing
Italy	Italian	Translation verified before testing
Italy	German	Translation verified before testing
Latvia	Latvian	Translation verified before testing
Latvia	Russian	Russian instrument with adaptations
Lithuania	Lithuanian	Translation verified before testing
Lithuania	Russian	Russian instrument with adaptations
Norway	Norwegian	Translation verified before testing
Poland	Polish	Translation verified before testing
Portugal	Portuguese	Translation verified before testing
Romania	Romanian	Translation verified before testing
Russian Federation	Russian	Translation verified before testing
Slovak Republic	Slovak	Adapted from Czech instruments
Slovenia	Slovenian	Translation verified before testing
Sweden	Swedish	Translation verified before testing
Switzerland	German	German instruments with adaptations
Switzerland	French	Belgian instruments with adaptations
Switzerland	Italian	Italian instruments with adaptations
United States	English	Adaptations verified before testing

Note: * Translations were verified, but recommendations were not implemented in the final translation.

- finding equivalent words and phrases;
- making sure the reading level was the same in the target language as in the original English version;
- making sure the essential meaning of each item did not change;
- making sure the difficulty level of achievement items did not change; and
- being aware of changes in layout due to translation.

Table 3.2 Languages and instrument verification in participating countries (upper secondary students)

Country	Language	Verification Procedure
Chile	Spanish	Translation verified before testing
Columbia	Spanish	Translation verified before testing
Cyprus	Greek	Translation verified before testing
Czech Republic	Czech	Translation verified before testing
Denmark	Danish	Translation verified before testing
Estonia	Estonian	Translation verified before testing
Estonia	Russian	Russian instrument with adaptations
Hong Kong (SAR)	Chinese	Translation verified before testing
Israel	Hebrew	Translation verified before testing
Israel	Arabic	Translation verified before testing
Latvia	Latvian	Translation verified before testing
Latvia	Russian	Russian instrument with adaptations
Norway	Norwegian	Translation verified before testing
Poland	Polish	Translation verified before testing
Portugal	Portuguese	Translation verified before testing
Russian Federation	Russian	Translation verified before testing
Slovenia	Slovenian	Translation verified before testing
Sweden	Swedish	Translation verified before testing
Switzerland	German	Translation verified before testing

The *Translation Manual* also provided guidelines regarding decisions about vocabulary, meaning, layout, and cultural adaptations. For all countries, including the English-speaking countries, where some adaptations according to language use or the socio-cultural or political context had to be made, it was important to ensure equivalence of the items across languages. The original instruments (in English) contained alternatives or clarifications of terms in brackets for those questions where an adaptation was likely. Furthermore, the international instruments were sent with detailed translation notes to ensure accurate translation. These notes clarified acceptable adaptations, indicated changes in item wording from the field trial, and contained specific instructions (e.g., a description of how to translate and adapt graphics in the civic knowledge test).

The national teams were instructed not to simplify or clarify the test items, because doing so would have influenced the difficulty of the items. For the *multiple-choice* items used in the civic knowledge test, translators had to pay particular attention to the correspondence between words in each question and the words in the choices. Some items required an exact (verbatim) match between words in the question and words in the choices; that is, if the words in the question were repeated verbatim in the distracters of the international English version, they had to be repeated verbatim in the translation.

All changes to the format of a question required prior approval from the ICC. When in doubt, the national centers had to send a copy of the original item and a copy of the proposed item format to the international center for approval.

In an international study such as the IEA Civic Education Study, many changes in the questions are usually required from country to country because of differences in culture, quite apart from language. Terms used in the CivEd instrument needed to be equally familiar to all students, and cultural adaptations therefore were often needed to ensure this. Concepts or expressions that were not common to all cultures and not related to the substance of the questions had to be adapted to the cultural context. Changes in names, for example in fictitious newspaper articles used as stimuli in the cognitive test, had to be similar in length and complexity to the originals. National centers were asked to ensure that, in the process of translation, the changes made for cultural (or linguistic) reasons were made in a way that would not affect the difficulty levels of items, and that they were made consistently.

There were two possible levels of modifications to the instruments: the classes of allowed changes (applicable to all countries equally) and, within each of these classes, specific changes made by one or a few countries. When modifying the text of an item, translators were requested to take the meaning of the question, the reading level of the text, the difficulty level of the item, and the likelihood of another possible correct answer for the test item into account. Any changes made to items had to be noted in the Translation Deviation Form (see Appendix C).

TRANSLATION VERIFICATION

When the translation and layout had been completed, the national centers were asked to submit their national school, teacher, and student questionnaires to the international center for verification of the translations. The people who did this verification work were independent language experts selected by the international center. Each verifier was a native speaker of his or her respective language with a good command of English and who did not have any form of working relationship with the respective national center. The verifiers checked the translated questionnaires against the English original and the completed Translation Verification Report Form (see Appendix C), on which they listed any minor or major deviations from the original. These, together with any suggestions for improvement, were then sent back to the national centers. The centers were then asked to consider any suggestions for changes in the wording before deciding on the final version of the questionnaires.

The translation verification procedures were applied twice for most countries. The first verification was carried out for the instruments used in the 1998 field trial, after the data collection. The verification results were discussed with the national centers and then used to improve the translation of the final instruments for the main study in 1999. Second, the final national instruments were verified prior to the data collection. A verification of translated instruments was carried out for both the 14-year-old students and the upper secondary students.

For the 14-year-olds, the instruments were translated into 22 different languages. Table 3.1 (above), which shows the languages used in each country, also indicates whether instruments were verified before testing took place. Some countries used the already translated instruments from another country. For example, Switzerland adapted the questionnaires from Germany, Belgium (French), and Italy, and the Baltic countries used the Russian questionnaire. In these cases, questionnaires were checked for adaptations only. In Belgium (French) and Greece, due to organizational problems, the national centers did not take the verification results into account prior to the field study. However, the verification results enabled the ICC to check for translation errors when adjudicating items for scaling and preparing the analysis.

For the instruments used in the survey of the upper secondary students, a somewhat different procedure was followed. As most parts of the (student and school) questionnaires were identical to the ones used for the 14-year-old students, only the (modified) civic knowledge test and some additional questions had to be verified. An exception was Israel, which had not participated in the data collection among 14-year-old students. Here, both the Arabic and the Hebrew questionnaires were completely verified. Table 3.2 above shows the languages used for the survey and test among 16- to 18-year-old students in each country. Questionnaires for this part of the IEA Civic Education Study were translated into 16 languages.

FIELD OPERATIONS

Pilot Study

The data collection of the IEA Civic Education Study followed detailed guidelines describing the field operations that the national research centers in the participating countries were required to implement. Pilot studies in 1998 (14-year-old students) and 1999 (upper secondary students) were not only designed to pilot cognitive tests and questionnaire material (see Chapter 2) but also to replicate most of the sampling and field procedures that participants would encounter in the main study. Within-school sampling, contacting schools, test administration, and data-entry and data-cleaning procedures were carried out by a majority of countries one year before the actual data collection took place. This was done to ensure that the national centers were familiar with the data collection procedures and to test the international procedures across countries.

National Research Coordination

In international studies, similar procedures for data collection across countries are crucial for obtaining comparable data. The national research coordinators (NRCs) were the key persons in conducting the field operations in each country and were also responsible for implementing the internationally agreed procedures and preparing the data according to the international specifications. The ICC provided the national centers with the following detailed procedural manuals:

• The *Sampling Manual* (IEA Civic Education Study, 1998b, 1999f) defined the target population and described the procedure for obtaining a school sample.

- The *Field Operations Manual* (IEA Civic Education Study, 1999a, 1999g) gave a comprehensive description of all procedural aspects of the IEA Civic Education Study in a country from the delivery of test instruments and questionnaires through to the cleaning and delivery of data sets.

- The *School Coordinator Manual* (IEA Civic Education Study, 1999b, 1999h) described the activities to be undertaken by the school coordinator for organizing the test and questionnaire administration, distributing the questionnaires to teachers and school principals, and handling the test and questionnaire materials at school.

- The *Test Administrator Manual* (IEA Civic Education Study, 1999c, 1999i) described in detail the procedures from the beginning of the test administration to the time of returning test and questionnaire materials to each national center.

- The *Manual for Entering the CIVIC Data* (IEA Civic Education Study, 1999d, 1999j) provided the coding and data entry personnel in the national centers with the necessary information to enter and verify the data. It also defined the variables and file formats in the data files, provided instructions for verification of the data, and assisted national coordinators in the subsequent analysis of their data.

Selection of Civic-related Subjects

In the IEA Civic Education Study, the teacher questionnaire was administered to up to three different teachers of the sampled class who taught civic education or civic education-related subjects. However, because civic education-related subjects are labeled differently across countries and because the same subject can have different degrees of "civic-relatedness" across countries, a procedure was implemented that used the cognitive items of the civic knowledge test to identify the three most appropriate subjects and their teachers in a country. Once the subjects had been identified, schools were told which subjects and teachers they should enter in the forms.

The subjects were determined by means of a Subject Allocation Grid (see Appendix C). The rows of the grid represent the items (38 items for the 14-year-old students and 43 items for the upper secondary students) in the cognitive test; the columns represent civic education and other subjects related to civic education. Therefore, before the grid could be completed, the names of the subjects had to be inserted in the header of the grid. Then, for each of the items in the cognitive test, the national centers had to determine the subject in which the content of the respective item was taught in the target grade or had been taught prior to the target grade. Where this could be determined for an item, the respective cell in the grid then had to be checked. After clarifying to which subject the items belonged, the number of checkmarks per subject were counted and reported in the last row of the Subject Allocation Grid (sum of ticks). The subject with the most checkmarks was defined as "civic education-related subject No. 1", the subject with the second highest number of checkmarks as "civic education-related subject No. 2", and the subject with the third highest number of ticks as "civic education-related subject No. 3". The national centers were advised to undertake the subject allocation with the

assistance of the National Civic Education Expert Panel. Where two subjects received the same number of checkmarks, the national experts determined the most appropriate allocation—that is, "subject 1", "subject 2", or "subject 3".

Contacting Schools and Test Preparation

After the sample of schools had been drawn, the national centers were required to engage in a series of tasks to obtain the cooperation of the sampled schools and ensure high response rates. In countries where the Ministry of Education could require every selected school to participate in the study, obtaining permission for the testing was relatively straightforward. However, in those countries where permission had to be obtained from various school authorities, getting permission from all these bodies tended to be a long process. In some countries, the national centers had to obtain the permission of regional, state, and/or district authorities.

Often, the authorities at higher levels gave permission to undertake the assessment but allowed the school principal in each selected school to decide whether the school should participate. In such instances, permission had to be sought directly from the school principal, with the national centers having to develop particular strategies to ensure permission (e.g., personal letters signed by higher officials, telephone calls). To obtain the highest possible acceptance rate at this stage, special guidelines regarding the letter to school principals were given to the national centers. This had to include a description of the purpose and the significance of the study, and it had to be written in easily understood language so that copies could be distributed to parents, local community leaders, local newspapers, etc.

In the letter, the cooperation of the schools was invited by giving precise details about what participation would involve (e.g., the grade and number of classes involved, the length of the testing sessions, the approximate time required to complete the questionnaires). Proposed dates were given for administration of the instruments. Any benefits that individual schools might receive from their participation in the study (e.g., feedback on school data, exposure of teachers from participating classes to a range of student testing techniques) and a guarantee that students and schools would be granted anonymity in all published results were explicitly stated. The letter also asked the school principals to appoint an individual in their school who could serve as the school coordinator (see below) and take responsibility for administrative arrangements (e.g., arrange a room for testing) for the study in the school. Finally, the school principals were invited to contact the members of the national centers if they needed further information before making their decision.

After a reasonable period of time, those school principals who had not replied to the letter were contacted by phone calls or through their regional offices, and encouraged to respond with a positive decision as soon as possible. Once a school principal had decided to participate, he or she was sent a follow-up letter, which offered thanks for the school's cooperation and confirmed that future correspondence regarding the study would be addressed to the school coordinator.

The school coordinator was the key person in arranging the actual assessment within the school. The school principal usually identified a teacher as school coordinator, but the national centers could also appoint a member of their central team or a member of their team of test administrators to be the school coordinator. Sometimes the school coordinator was also the test administrator. The school coordinators were also required to provide the list of civic education classes in the target grade in the school, help the national center determine the dates of test administration, and provide the necessary information about the class in the target grade. They had to make the arrangements for testing, distribute parental permission forms (if necessary), and ensure that these were signed and returned on time. They were also asked to convey the teacher and school questionnaires to the appropriate members of staff and to return the completed testing materials, questionnaires, and forms to the national center.

Before classes could be selected, eligible classes within the selected schools had to be identified and listed. The classes to be sampled were those relating to some form of civic education instruction and were not to be "streamed" or "tracked"; that is, all classes had to have a similar level of student abilities and instruction. In educational systems where all students in a class or homeroom receive the same instruction in all subjects, a "civic class" was deemed simply to correspond to this section within the school (such as 8a, 8b, etc.). However, in other educational systems, students form different groups or courses for instruction according to subject areas. In these countries, the national centers had to sample these groups or courses instead of classes or homerooms. The instructional units were sampled for instruction in the "civic education-related subject No. 1" as determined by the Subject Allocation Grid.

School coordinators were asked for a list of all civic-related classes in the target grade, and the national centers prepared a Class Sampling Form for sampling classes and selecting the civic education class (see Appendix C). After classes had been selected by the national center, the school coordinators were given a Teacher-Student Linkage Form on which to record student and teacher names for each sampled class (see Appendix C).

In the procedures for the assessment of the upper secondary students in 1999/2000, a teacher questionnaire was not administered. Instead, the school coordinators were asked to provide information about the students in the sampled class on a Student Tracking Form, which the national centers had prepared for the selected classes and then sent to each school coordinator. This form was of crucial importance for identifying students to be tested and for assigning them unique student IDs. The form also included a record of the participation status for each sampled student. The school coordinator was responsible for listing the students of the selected classes on the form and then submitting it to the NRC.

Exclusions of Students

All students who were enrolled in the target grade belonged to the target population. However, some students within the selected schools were unable to take part in the testing because of various disabilities. The number of students in this category was held to a minimum to avoid biasing international comparisons. It was therefore important to carefully define the conditions and categories under which students could be excluded.

The following general categories for the exclusion of students within schools were used and were adapted to fit the context of each educational system.

- *Functionally disabled students permanently physically disabled in such a way that they could not perform in the IEA Civic Education Study testing situation:* It was expected that functionally disabled students who could respond to the test would be included in the testing.

- *Educable mentally retarded students who were considered in the professional opinion of the school principal or by other qualified staff members to be educable mentally retarded or who had been psychologically tested as such:* This included students who were emotionally or mentally unable to follow even the general instructions of the test.

- *Students with limited proficiency in the test language and so unable to read or speak the language of the test:* These students were expected to be unable to overcome the language barrier in the test situation. Typically, a student who had received less than one year of instruction in the language of the test was excluded from testing.

In some educational systems, these students are in special schools or special classes, and it was possible to exclude these schools or classes from the sample. However, in other educational systems this is not the case. For example, in countries where such students are integrated into the mainstream schooling system, these students are not in separate classes. They accordingly had to be excluded from the test administration by the school principal or other qualified staff member. Detailed instructions were given to the school coordinator.

Test Administration and Data Entry

Using the information about students and teachers provided by the school, the national centers prepared the Student Tracking Form[2] and the Teacher Tracking Form (for the 14-year-old students only) (Appendix C), which contained the names of students and teachers to be assessed or surveyed. These were sent to the schools together with the test material and questionnaires. After the test administration, the school coordinators returned the two forms (including participation status) and testing materials to the national centers.

Test administrators were responsible for ensuring that each student received the correct testing materials, for administering the test in accordance with the instructions in the *Test Administrator Manual,* for ensuring the correct timing of the test and recording the time spent on it, and for recording student

---

2 As outlined above, the Student Tracking Form for the survey of the upper secondary students had been previously completed by the school coordinators.

participation on the Student Tracking Form. In some countries, as noted previously, the school coordinators were also responsible for the test administration. Ideally, the test administrators were not to be the class teachers of the tested class but could be other teachers of the tested class, other teachers from the school, other school personnel, civic education teachers, or external staff.

In the study of the 14-year-olds, 10 of the participating countries engaged external test administrators. In seven countries, the school coordinators administered the tests. In the majority of the countries, class teachers were used as test administrators, followed by other teachers of the tested class, civic education teachers, and other school personnel. In the study of the upper secondary students, the majority of the 16 participating countries involved civic education teachers and other teachers of the tested class; some countries also used class teachers. Only two countries reported the appointment of external test administrators. In both populations, the test administrators received training mainly through formal sessions and/or written instructions. In a few countries, instructions were also given by telephone.

The national centers had to ensure the security of testing material from the time of receiving the international instruments until submitting data to the IEA Data Processing Center (DPC). Test materials were sent to schools and returned to the national centers immediately after test administration. Data entry was organized at the national centers or through sub-contractors. About half of the participating countries used their own staff for data entry. In this instance, special training sessions were conducted for the staff. In the majority of the countries a percentage of test booklets were entered twice as a verification procedure to assure the quality of the data entry process. The percentage of double-entered data ranged from 1 to 100 percent. A data entry program developed by the IEA DPC together with instructions about how to enter, clean, and submit the data were distributed to the participating countries.

QUALITY CONTROL PROCEDURES DURING AND AFTER THE DATA COLLECTION

Standards for IEA studies require that certain quality control measures be undertaken during the data collection process. Ideally, quality control measures for IEA studies should include personal interviews with the NRCs and visits to a sub-sample of tested schools to observe test administration (see, for example, O'Connor & Stemler, 2000). However, limitations in funding for the IEA Civic Education Study imposed constraints on the extent of quality control and did not permit these kinds of monitoring activities. However, three measures were implemented after completion of the testing:

• phone calls to schools;

• quality control questionnaires completed by national centers;

• NRC survey reports.

The national centers were requested to make *phone calls to 25 percent of the tested schools* the day after the testing had taken place. The school coordinator in each school was asked how the testing had been done, if there had been any

problems, and if there had been deviations from the testing procedure. Most of the participating countries (both populations tested) reported that national center staff made check-up calls after the testing had taken place. In three countries, external agencies were hired to do the telephone check-up procedures. The national centers were provided with guidelines for the telephone interviews. The sub-sample of tested schools was a simple random sample of participating schools. The size of the quality assurance sample of schools in the different countries ranged between 10 and 176 schools. Between 10 and 60 percent of the tested schools received check-up calls after testing. In a number of countries, organizational problems made it impossible for the national centers to implement the phone calls. Some countries used other measures of quality control, such as quality control questionnaires completed by the school coordinator, or quality control visits made by national center staff to some of the tested schools.

After completion of the main study, the national centers were asked to complete a *quality control questionnaire* to determine if procedures had been followed during the process of data collection and whether any deviations had occurred, and to give the national research centers an additional guideline for their work. According to the completed questionnaires, national centers in general had followed the procedures outlined in the manuals. However, in some cases, funding and staffing restrictions within the national center had made it extremely difficult to implement additional measures of quality control.

The quality control questionnaire also asked the national centers to report any special circumstances that could have affected the testing or students' responses in their countries. Although the majority of the centers reported no such mitigating events, in some countries the study had been affected by teacher or student strikes. One European country referred to possible effects of the presidential elections running at that time, while the USA reported that hurricanes had caused problems with testing. In some Central and Eastern European countries, testing and the survey were overshadowed by the war in Kosovo.

Eight national centers commented on the difficulty of obtaining the participation of schools in their countries. Special faxes, letters, recommendations from their ministries or teachers' organizations, and persistent phone calls were the most important measures applied to improve participation rates. Most of the national centers, however, said they encountered no difficulties in obtaining cooperation from schools. Interestingly, in the Eastern European countries that took part in the study and in countries where test administration was carried out on behalf of a central authority, such as a ministry of education, few difficulties were reported.

The procedures associated with the printing and security of test material was another important aspect covered in the quality control questionnaire. The national centers were asked whether errors had been detected during the printing process and what procedures had been in place to protect the security of the test material. Several of the national centers discovered printing errors. The most frequently reported concerns were "pages missing", "binding problems", and "page ordering". All national centers reported that they had

followed procedures to protect the security of the test material. In most cases the materials had been kept locked either in the schools or at the national centers and had been sent directly after testing to the centers. No breaches of security were reported.

After completing the data collection, the national centers were requested to submit a NRC Survey Report to the ICC. The report required each center to describe and document any unusual occurrences or deviations from prescribed procedures with regard to:

- sampling;
- test instruments;
- field operations;
- test administration;
- data entry;
- anything that might help to interpret the main survey data or explain possible anomalies.

Most national centers complied with this request. Although many minor delays were recorded, most national centers generally managed to follow the prescribed procedures and to collect their data in a satisfactory manner.

SUMMARY

Considerable efforts were made to ensure a high quality of testing procedures. The translations of instruments were monitored through a two-phased verification for both the pilot and main studies. Detailed documents helped the national centers follow the internationally agreed field procedures, and some additional quality assurance measures were implemented to ensure international comparability. Reports from the national centers and the quality control questionnaires indicate that the procedures implemented for the IEA Civic Education Study were generally followed and that the obtained data were of high quality.

REFERENCES

Gonzalez, E., & Hastedt, D. (2000). TIMSS field operations and data preparation. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 119–132). Chestnut Hill, MA: Boston College.

IEA Civic Education Study. (1998a). *Translation manual* (CIVIC Reference No. 98–0102). Berlin: HU Berlin.

IEA Civic Education Study. (1998b). *School sampling manual* (CIVIC Reference No. 98–0101). Hamburg: IEA Data Processing Center.

IEA Civic Education Study. (1999a). *Field operations manual* (CIVIC Reference No. 99–0001). Hamburg/Berlin: IEA Data Processing Center/HU Berlin.

IEA Civic Education Study. (1999b). *School coordinator manual* (CIVIC Reference No. 99–0002). Hamburg: IEA Data Processing Center.

IEA Civic Education Study. (1999c). *Test administrator manual* (CIVIC Reference No. 99–0003). Hamburg: IEA Data Processing Center.

IEA Civic Education Study. (1999d). *Manual for entering the CIVIC data* (CIVIC Reference No. 99–0004). Hamburg: IEA Data Processing Center.

IEA Civic Education Study. (1999e). *Translation manual: Main study of the older population* (CIVIC Reference No. 99–0204). Berlin: HU Berlin.

IEA Civic Education Study. (1999f). *School sampling manual* (CIVIC Reference No. 99–0209). Hamburg: IEA Data Processing Center.

IEA Civic Education Study. (1999g). *Field operations manual* (older population) (CIVIC Reference No. 99–0206). Berlin: HU Berlin.

IEA Civic Education Study. (1999h). *School coordinator manual* (older population) (CIVIC Reference No. 99–0207). Berlin: HU Berlin.

IEA Civic Education Study. (1999i). *Test administrator manual* (older population) (CIVIC Reference No. 99–0208). Berlin: HU Berlin.

IEA Civic Education Study. (1999j). *Manual for entering the CIVIC data* (CIVIC Reference No. 99–0210). Hamburg: IEA Data Processing Center.

O'Connor, K. M., & Malak, B. (2000). Translation and cultural adaptation of the TIMSS instruments. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 89–100). Chestnut Hill, MA: Boston College.

O'Connor, K. M., & Stemler, S. E. (2000). Quality control in the TIMSS data collection. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 135–154). Chestnut Hill, MA: Boston College.

Chapter 4:

THE CIVED SAMPLING DESIGN[1]

Heiko Sibberns and Pierre Foy

OVERVIEW

This chapter describes the procedures implemented in the IEA Civic Education (CivEd) Study to ensure proper sampling of the student population in each participating country. To be acceptable, the national sample designs had to result in probability samples that gave accurate weighted estimates of population parameters, and for which estimates of sampling variance could be computed. The CivEd sample design was identical to the one used for TIMSS 1999, which in turn was based on the TIMSS 1995 sample design. The CivEd sample design was chosen to balance analytical requirements and operational constraints. However, it was kept simple enough for all participants to implement. The quality of the samples depended very much on the accuracy of the information available at the design stage and on the sampling procedures.

The national research coordinators (NRCs) were aware that the CivEd sampling procedures were complex. Centrally collecting the necessary information about the national educational systems would have placed considerable demands on resources and expertise. At the same it was realized that only a limited number of qualified sampling personnel were available at the national centers. Therefore, it was mandatory to keep all procedures as simple as possible. All sampling procedures within schools, like the selection of classes or students, especially needed to be straightforward.

The International Coordination Center (ICC) and the IEA Data Processing Center (DPC) provided manuals and expert advice to help the NRCs adapt the CivEd sample design to their educational system. The CivEd school sampling manuals for the 14-year-old students and for the upper secondary students described how to implement the international sample design and adapt it to the participating countries' specific requirements. The manuals included advice on planning, implementing, and documenting all phases of the school sampling process. The survey operations manuals and the school coordinator manuals described the selection of classes within schools and the administration and monitoring of the procedures used to identify and track respondents and non-respondents.

In addition, the NRCs had access to expert support. The IEA DPC, in close cooperation with Statistics Canada, reviewed and approved the national sampling plans, sampling data, sampling frames, and sample selections. Statistics Canada assisted nearly half of the countries that participated in the survey of the 14-year-olds with their sampling procedures. The samples for these countries were drawn simultaneously with the TIMSS 1999 samples.

---

1 This chapter is based on Chapter 2 of the *TIMSS 1999 Technical Report* (Foy & Joncas, 2000).

The NRCs were allowed to modify the CivEd sample design so as to meet the requirements of their national educational systems or to address national research questions requiring changes in the design. However, the ICC had to approve these adjustments, and the IEA DPC to monitor them.

TARGET POPULATION AND EXCLUSION

In all IEA studies, the target population is known as the *international desired population*. The international desired target population for the survey of 14-year-olds was as follows:

- All students enrolled on a full-time basis in that grade in which most students aged 14:00 to 14:11 [years: month] are found at the time of testing.

The international desired target population for the survey of upper secondary students was as follows:

- All students enrolled on a full-time basis in that grade in which most students aged 16:00 to 16:11 [years: month] or 17:00 to 17:11 [years: month] or 18:00 to 18:11 [years: month] are found at the time of testing.

Given that no international definition of the upper secondary target population could be agreed upon within the constraints of the different participating countries, each country was free to select that grade closest to the end of the upper secondary school that interested the national research groups and/or funders.

School and Within-School Exclusions

The IEA Civic Education Study expected all participating countries to define *national desired populations* to correspond as closely as possible to the *international desired population*. However, sometimes it was necessary to make changes in order to adapt to national restrictions. For example, countries had to restrict coverage by excluding remote regions or segments of their educational system. Appendix D documents any deviations from the international desired population.

Using the *national desired population* as a basis, participating countries had to operationally define their population for sampling purposes. These definitions determined the *national defined population* and the sampling frame from which schools were randomly selected during the first stage of sampling. The *national defined population* could be a subset of the national desired population. All schools and students excluded from the *national desired population* were referred to as the *excluded population*.

Participants in the IEA Civic Education Study were required to keep the excluded population under five percent of the national desired population. Exclusions could occur at the school level, at the student level, or at both levels. Because the national desired population was restricted to the target grade, schools not containing the target grade were considered to be outside the

scope of the sampling frame. Participating countries could exclude schools for the following reasons:

- They were in geographically remote regions.
- They were of extremely small size.
- They offered a curriculum, or school structure, that was different from the mainstream educational system.
- They provided instructions only to students in the exclusion categories defined as "within-sample" exclusions.

Within-sample exclusions were limited to students who, because of a disability, were unable to take the civic knowledge test. The NRCs were asked to define the anticipated within-school exclusion. Because definitions can vary internationally, the NRCs were also asked to follow certain rules for defining these exclusion groups. The size of exclusions had to be estimated. It had to be in line with the five percent rule specified above.

The international within-school exclusion rules were specified as follows:

- *Educable mentally retarded students:* These were students who were considered, in the professional opinion of the school principal or other qualified staff members, to be educable mentally disabled, or students who had been diagnosed by psychological tests as such. This group included students who were emotionally or mentally unable to follow even the general instructions of the civic knowledge test.

- *Functionally disabled students:* These were students who were permanently physically disabled in such a way that they could not perform the civic knowledge test. Functionally disabled students who could perform the test were included.

- *Non-native speakers:* These were students who could not speak or read the language of the test. Typically, a student who had received less than one year of instruction in the language of the test was excluded. However, this definition was adapted by the national centers in order to match the national context.

The aim of the sampling procedures was to ensure that the effective target population would be as close as possible to the international desired population. Any exclusion of students from the international desired target population had to be accounted for, both at the school and the student level.

The size of the excluded population was documented and served as an index of coverage and representativeness of the selected samples. The different stages of possible exclusions between the international desired target population and the effective target population are illustrated in Figure 4.1.

Figure 4.1 Population definitions and exclusions
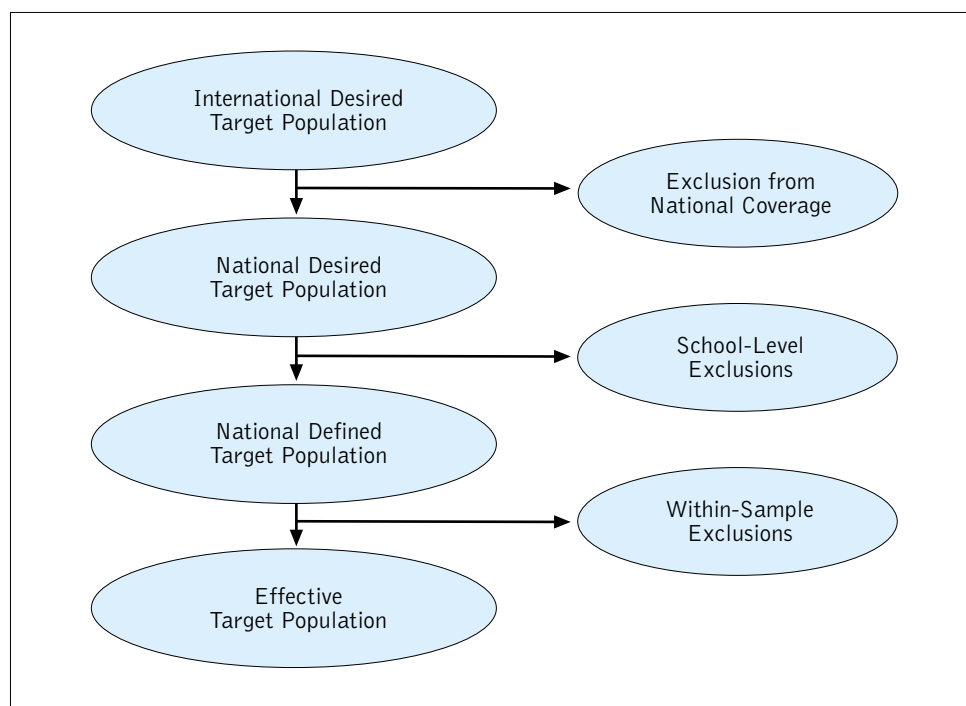
SAMPLE DESIGN

The basic sample design of the IEA Civic Education Study is generally called a two-stage stratified cluster sample design. The first stage consisted of a sample of schools that had been stratified in a number of countries. The second stage consisted of a single civic education class selected from the target grade in sampled schools. It was also permissible to add a third stage where students were sampled within classrooms.

Units of Analyses and Sampling Units

The analytical focus of the IEA Civic Education Study was the learning and attitudes of students and the instructional and school characteristics influencing the learning and attitudes. The sample design had to address the measurement of variables at student level and classroom/school level that were deemed to influence learning and attitudes and the collection of data on specific characteristics of instruction and the school. Because schools, classrooms, and students were all considered potential units of analyses, they had to be regarded as sampling units.

Although intact classrooms were sampling units in the second sampling stage, the ultimate sampling elements were students. Consequently, it was important that every student from the target grade be a member of one, and only one, civic education class (or untracked homeroom) in a school from which the sampled class was to be selected. In most educational systems, the civic education class coincided with a homeroom class. However, it was also possible that a different subject could be selected for sampling purposes, for example history, if civic education as a subject was not taught in the target grade of the selected school (see Chapter 3 for details).

Sampling Precision and Sample Size

Sample sizes in the IEA Civic Education Study had to be determined in order to meet the analytical requirements. Since students were the principal units of analysis, it was important to produce reliable estimates of students' characteristics. The CivEd standard for sampling precision required that all population samples had an effective sample size of at least 400 for the cognitive scales. Phrased differently, the standard errors of the samples were not to be greater than those obtained by a simple random sample of 400 students.

An effective sample size of 400 results approximately in the following 95 percent confidence limits for sample estimates of population means, percentages, and correlation coefficients:

- *Means: M $\pm$ 0.1 s* (where *M* is the mean estimate and s is the estimated standard deviation.

- *Percentage:* p $\pm$ 5% (where p is the estimated percentage).

- *Correlations: r $\pm$ 0.1* (where *r* is the correlation estimate).

To obtain sufficient data for reliable analyses at the school and class levels, at least 150 schools had to be selected from the target population. A sample of 150 schools results in 95 percent confidence limits for school-level and class-level mean-estimates, which are precise to within $\pm$ 16 percent of their standard deviation. To achieve this level of precision, countries had to sample at least 150 schools, even if smaller school samples would have met the student-level requirements.

The precision of multi-stage cluster sample designs is generally affected by the so-called clustering effect. A classroom as a sampling unit is a cluster of students who tend to be more like each other than like other members of the population. The intra-class correlation is a measure of this similarity. Sampling 30 students from a single classroom, when the intra-class correlation is different from zero, results in less information than a simple random sample of 30 students across all classrooms in a school. Such a sample design is less efficient, in terms of information per sampled student, than a simple random sample of the same size. This clustering effect had to be taken into account when determining the overall sample sizes in the IEA Civic Education Study.

The magnitude of the clustering effect is determined by the size of the cluster and the size of the intra-class correlation. When planning the sample, each country had to choose a value for the intra-class correlation and a value for the expected cluster size (known as minimum cluster size). The intra-class correlation was estimated from previous studies like TIMSS 1995, or from national assessments. In the absence of any measures, an intra-class correlation of 0.3 was assumed. Since all countries except one chose to test intact classrooms, the minimum cluster size was set to the average classroom size. The specification of the minimum cluster size affected the treatment of small schools and small classrooms as well as the number of schools to be sampled.

Sample design tables were produced and included in the CivEd sampling manuals. These tables presented the number of schools to be sampled to meet

the precision requirements for a range of intra-class correlations and minimum cluster sizes (for an example, see Table 4.1).

Table 4.1 Example of a sample design table (95% confidence limits for means ±0.1s / percentages ±5.0%)

MCS		Intra-class Correlation								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	A	112	144	176	208	240	272	304	336	368
	N	560	720	880	1,040	1,200	1,360	1,520	1,680	1,840
10	A	76	112	148	184	220	256	292	328	364
	N	760	1,120	1,480	1,840	2,200	2,560	2,920	3,280	3,640
15	A	64	102	139	176	214	251	288	326	363
	N	960	1,530	2,085	2,640	3,210	3,765	4,320	4,890	5,445
20	A	58	96	134	172	210	248	286	324	362
	N	1,160	1,920	2,680	3,440	4,200	4,960	5,720	6,480	7,240
25	A	150	150	150	172	211	249	287	326	364
	N	3,750	3,750	3,750	4,300	5,275	6,225	7,175	8,150	9,100
30	A	52	91	130	168	207	246	284	323	362
	N	1,560	2,730	3,900	5,040	6,210	7,380	8,520	9,690	10,860
35	A	150	150	150	169	208	246	285	324	363
	N	5,250	5,250	5,250	5,915	7,280	8,610	9,975	11,340	12,705
40	A	49	88	127	166	205	244	283	322	361
	N	1,960	3,520	5,080	6,640	8,200	9,760	11,320	12,880	14,440
45	A	150	150	150	167	206	245	284	323	362
	N	6,750	6,750	6,750	7,515	9,270	11,025	12,780	14,535	16,290
50	A	48	87	126	165	204	244	283	322	361
	N	2,400	4,350	6,300	8,250	10,200	12,200	14,150	16,100	18,050

Note: A = number of sampled schools; N = number of sampled students in the target grade.

Stratification

Stratification is the grouping of sampling units in the sampling frame according to some attribute or variable prior to the sample selection. It is generally used for the following reasons:

• To improve the efficiency of the sample design, thereby making survey estimates more reliable.

• To apply different sample designs, or disproportionate sample-size allocations, to specific groups of schools (such as those in specific regions or school types).

• To ensure adequate representation in the sample of specific groups from the target population.

Examples of stratification variables for school samples are geography (such as regions or provinces), school type (such as public or private), and level of urbanization (rural/urban). Stratification variables in the CivEd sample design could be explicit, implicit, or both.

Explicit stratification consists of building separate school lists, or sampling frames, according to the stratification variables under consideration. If, for example, "regions" was an explicit stratification variable, separate sampling frames were constructed for each region. Different sample designs, or different sampling fractions, could be applied to each sampling frame. In practice, the main reason for constructing explicit strata was disproportionate allocation of the school sample across strata. For example, a country might require an equal number of schools for each stratum regardless of the relative size of the stratum.

Implicit stratification makes use of a single school sampling frame, but sorts the schools in the frame according to a set of stratification variables. This is a simple way of guaranteeing proportional allocation without the complexity of explicit stratification. Implicit stratification can also improve the reliability of survey estimates, provided the stratification variables are related to school mean student achievement.

Replacement Schools

Although participants in the IEA Civic Education Study placed great emphasis on securing school participation, it was anticipated that it would be difficult for many countries to achieve a 100 percent participation rate. To address this situation, two replacement schools were selected together with the originally sampled schools. The use of implicit stratification variables and the subsequent ordering of the school sampling frame by school size ensured that characteristics of the sampled school were similar to those of the replacement schools. Although this procedure does not guarantee the absence of potential bias in population estimates, it helps to minimize it.

First Sampling Stage

The sample-selection method used in this study involved a systematic probability proportional to size (PPS) technique. This technique requires some measure of size (MOS) of the sampling units. Ideally, this is the number of sampling elements within the unit (e.g., the number of students in the target grade in the school). If this information is not available, some highly correlated measure like total school enrolment is used.

Schools in each explicit stratum were listed in order of the implicit stratification variables, together with the MOS of each school. They were also sorted by MOS. The measure of size was accumulated from school to school, and the running total was listed next to each school. The total cumulative MOS was a measure of the size of the population of sampling elements. Dividing this figure by the number of schools to be sampled gave the sampling interval.

The first school was sampled by choosing a random number between one and the sampling interval. The school whose cumulative MOS contained the

random number was the first sampled school. By adding the sampling interval to the random number, a second school was identified. This process of consistently adding the sampling interval to the previous selection number resulted in a PPS sample of the required size.

As each school was selected, the next school in the sampling frame was designated as a replacement school for use should the sampled school not participate. The school after the first replacement was the second replacement and was to be used if neither the sampled school nor the first replacement participated.

Major benefits of the PPS sample selection method are that it is easy to implement and that it is easy to verify that it has been implemented properly.

Small Schools

Small schools tend to be problematic in PPS samples because students sampled from these schools get disproportionately large sampling weights; when the school size falls below the minimum cluster size, it reduces the overall sample. A school was defined as a small school in the IEA Civic Education Study if it was smaller than the minimum cluster size. If, for example, the minimum cluster size for a country was set at 20, then a school with fewer than 20 students in the target grade was considered a small school.

The following two steps were implemented to deal with small schools:

- *Extremely small schools:* Extremely small schools were defined as schools with students numbering under half the minimum cluster size. If student enrolment in these schools was less than two percent of the eligible population, they were excluded, provided the overall exclusion rate did not exceed the five percent criterion.

- *Explicit stratum of small schools:* If less than 10 percent of the eligible students were enrolled in small schools, no additional action was required. However, if more than 10 percent of the students were enrolled in small schools, an explicit stratum of small schools was required. The number of schools sampled from this stratum remained proportional to the stratum size, but all schools had an equal probability of selection. This action ensured greater stability in the resulting sampling weights.

Optional Preliminary Sampling Stage

Some very large countries chose to introduce a preliminary sampling stage before sampling schools. This consisted of a PPS sample of geographic regions. A sample of schools was then selected from each sampled region. This design was mostly a cost reduction measure where the construction of a comprehensive list of schools would have been either impossible or too expensive. Also, the additional sampling stage reduced the dispersion of the school sample, thereby reducing travel costs. Sampling guidelines were put in place to ensure that an adequate number of units were sampled at this preliminary stage. The sampling frame had to consist of at least 80 primary sampling units, of which at least 40 had to be sampled.

Second Sampling Stage

The second sampling stage consisted of selecting classrooms within sampled schools. As a rule, one classroom was selected per school, although one participant opted to sample two classrooms. Classrooms were selected with equal probability or with probability proportional to size. When all students in a selected classroom were tested, the classrooms had to be selected with equal probability.

Small Classrooms

Classrooms in an educational system generally tend to be of similar size. But small classrooms sometimes are devoted to special activities, such as remedial or accelerated programs. These may become problematic because classrooms selected with PPS can lead to a shortfall in sample size and thereby introduce some instability in the sample weights. To avoid this problem, the classroom sampling procedure specified that any classroom of less than minimum cluster size must be combined with another classroom from the same grade and school. The resulting pseudo-classroom then constituted a sampling unit.

PARTICIPATION RATES

Weighted and unweighted response rates were computed for each participating country at school level and at student level.

School-level Participation Rate

The minimum acceptable school-level participation rate, before the use of replacement schools, was set at 85 percent. This criterion was applied to the unweighted school response rate. School response rates were computed both weighted and unweighted, with and without replacement schools. Only the weighted response rates were reported for the 14-year-old students. For the upper secondary students, the unweighted participation rate after replacement was also reported (see Chapter 6 for details on the calculation of unweighted and weighted school-level participation rates).

Student-level Participation Rate

Like the school-level participation rate, the minimum acceptable student-within-school participation rate was set at 85 percent. This criterion was applied to the unweighted participation rates. Both weighted and unweighted participation rates were computed, but only the weighted participation rates were reported (see Chapter 6 for details on the calculation of unweighted and weighted student-level participation rates).

Overall Participation Rates

The minimum overall participation rate was set at 75 percent. This rate was calculated as the product of the weighted school-level participation rate without replacement schools and the weighted student-level participation rate. Weighted overall participation rates were calculated and reported both with and without replacement schools. (See Chapter 6 for details on the calculation of unweighted and weighted overall participation rates.)

COMPLIANCE WITH SAMPLING GUIDELINES

Generally, the NRCs made every attempt to meet the sampling requirements. However, not all countries met these requirements. Ten countries out of the 28 that participated in the survey of the 14-year-olds and seven countries out of the 16 that participated in the survey of the upper secondary students failed to achieve a weighted overall participation rate of 75 percent *before* replacement (see Table 4.2).

Three countries in the survey of the 14-year-old students had a weighted overall participation rate of less than 75 percent even *after* replacement schools had been included; these countries were reported with footnotes in the international reports (see Table 4.3). In the survey of the upper secondary students, four countries failed to meet a participation rate of 75 percent *after* replacement (see Table 4.4).

Colombia and Hong Kong (SAR) did not satisfy the guidelines for sampling procedures in the survey of the upper secondary students. Their data could not be weighted and were not reported with the data from the other participating countries. (They were, however, reported in an appendix.)

In eight out of the 28 countries participating in the survey of the 14-year-olds, the effective sample sizes for the combined civic knowledge scale were below 400, and in four of them they were below 300 (see Table E.1 in Appendix E). Among the countries participating in the survey of the upper secondary students, seven out of 14 countries fell short of achieving an effective sample size of 400 students for the combined civic knowledge scale; four had an effective sample size below 300 students (see Table E.2 in Appendix E).

However, failure to achieve the standard of collecting data from an effective sample size of 400 students in some of the countries did not jeopardize the cross-country comparability of these data. As the size of standard errors increased with decreasing effective sample sizes, the lack of precision in estimating population means from smaller samples was taken into account in the calculation of significance tests for cross-country comparisons (see Chapter 10 for details of the reporting of cross-country comparisons).

Table 4.2 Countries grouped according to their compliance with sampling implementation and participation rates

14-year-old Students	Upper Secondary Students
Countries satisfying guidelines for sample participation rates	
Bulgaria	Chile
Chile	Cyprus
Cyprus	Czech Republic
Czech Republic	Latvia
Estonia	Poland
Finland	Russian Federation
Greece	Slovenia
Hong Kong (SAR)	
Hungary	
Italy	
Latvia	
Lithuania	
Poland	
Portugal	
Romania	
Russian Federation	
Slovenia	
Sweden	
Countries satisfying guidelines for sample participation rates only after inclusion of replacement schools	
Australia	Israel
Colombia	Sweden
England	Switzerland (German)
Germany*	
Slovak Republic	
Switzerland	
United States	
*Countries **not** satisfying guidelines for sample participation rates*	
Belgium (French)	Denmark
Denmark	Estonia
Norway	Norway
	Portugal
*Countries **not** satisfying guidelines for sampling procedures*	
	Colombia
	Hong Kong (SAR)

Note: *National defined population covers less than all of the international desired population.

Table 4.3 Weighted participation rates and sample sizes for the survey of 14-year-old students

Country	School Participation Rate		Total Number of Schools that Participated	Student Participation Rate	Total Number of Students Assessed	Overall Participation Rate	
	Before replacement	*After replacement*				*Before replacement*	*After replacement*
Australia	75	94	142	92	3,331	69	86
Belgium (French)	57	75	112	93	2,076	53	70
Bulgaria	86	86	148	93	2,884	80	80
Chile	98	100	180	97	5,688	94	97
Colombia	66	94	144	96	4,926	64	90
Cyprus*	100	100	61	96	3,106	96	96
Czech Republic	91	99	148	95	3,607	86	94
Denmark	71	71	178	93	3,208	66	66
England	54	85	128	93	3,043	50	79
Estonia	84	85	145	90	3,434	76	77
Finland	93	98	146	93	2,782	86	91
Germany	63	94	169	89	3,700	56	84
Greece	88	93	142	97	3,460	85	90
Hong Kong (SAR)	90	100	150	99	4,997	89	99
Hungary	99	99	146	95	3,167	94	94
Italy	93	100	172	96	3,808	89	96
Latvia	89	91	130	91	2,572	81	82
Lithuania	93	97	169	90	3,494	84	87
Norway	75	77	154	93	3,321	70	71
Poland	83	90	179	94	3,376	78	84
Portugal	98	99	149	95	3,261	93	95
Romania	97	97	146	99	2,993	96	96
Russian Federation	96	98	185	97	2,129	94	95
Slovak Republic	79	97	145	94	3,463	74	91
Slovenia	93	99	149	96	3,068	89	95
Sweden	93	94	138	94	3,073	88	88
Switzerland	71	87	157	97	3,104	69	84
United States	65	83	124	93	2,811	61	77

Note: *In Cyprus, two classes per school were sampled.

Table 4.4 Participation rates and sample sizes for the survey of upper secondary students

| Country Before | School Participation Rate | | | Total Number of Schools that Participated | Student Participation Rate | Total Number of Students Assessed | Overall Participation Rate | |
	Before replacement (weighted)	After replacement (weighted)	After replacement (unweighted)				Before replacement (weighted)	After replacement (weighted)
Chile	99	100	100	180	96	5,777	95	96
Colombia	*	*	92	149	*	5,347	*	*
Cyprus	100	100	98	41	89	1,700	89	89
Czech Republic	96	98	98	147	92	3,362	89	91
Denmark	71	71	71	141	88	2,761	62	62
Estonia	75	77	86	142	91	3,215	68	70
Hong Kong (SAR)	*	*	54	81	*	5,810	*	*
Israel	88	98	98	233	84	4,430	74	83
Latvia	84	87	86	125	97	2,756	81	85
Norway	83	83	83	124	88	2,099	72	72
Poland	90	100	100	150	93	4,050	83	93
Portugal	94	94	95	149	79	2,795	74	74
Russian Federation	98	98	95	165	97	1,787	94	94
Slovenia	96	96	94	145	88	3,746	84	84
Sweden	90	91	89	88	76	2,678	69	76
Switzerland (German)	69	95	95	69	94	1,270	65	90

Note: *Participation rates are not reported for Colombia and Hong Kong (SAR) because only unweighted data were available.

REFERENCE

Foy, P., & Joncas, M. (2000). Sample design. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 29–45). Chestnut Hill, MA: Boston College.

DATA MANAGEMENT AND CONSTRUCTION OF THE CIVED DATABASE

Heiko Sibberns

OVERVIEW

The IEA Civic Education Study (CivEd) data were processed in a close cooperation between the national research centers, the IEA Data Processing Center (DPC), and the International Coordinating Center (ICC) at the Humboldt University Berlin. Each of these institutions was responsible for specific aspects of the data processing. The ICC was responsible for the overall coordination of data management procedures.

Data processing consisted of the following general tasks: data entry, creation of the international database, calculation of sampling weights, scaling of cognitive and attitude items, and the analyses of the background data. This chapter describes the data entry tasks undertaken by each national research coordinator (NRC), the data checking and database creation implemented by the IEA DPC, and the steps taken to ensure data quality and the accuracy of the international database. It also illustrates the responsibilities of each participant in creating the international database, the structure of the data files submitted by each country, the resulting files in the international database, and the rules, methods, and procedures used for data verification and manipulation and the software used in this process.

Data Flow

The data collected for the IEA Civic Education Study were entered at the national research centers of the participating countries into data files with a common, internationally predefined format. These files were transmitted to the IEA DPC for editing and verification. The major responsibilities of the DPC were as follows: to define the structure of the database to be transmitted by the NRCs; to check that incoming data matched the international standards; to adapt the structure where necessary; to implement and apply editing rules to verify the consistency and accuracy of the data; to correspond with the NRCs to ensure accuracy and comparability; to produce summary statistics for all relevant variables, including weighting statistics for the ICC; to add all relevant weighting variables to the database; and, finally, on the basis of input from the NRCs, to construct the core of the international database, that is, all data without scale scores.

Upon data verification and weighting, all data were sent to the ICC for scaling and scoring (see Chapter 7 for further information). Once scale scores had been calculated at the ICC, they were sent to the DPC for inclusion into the international database. The ICC then prepared the tables and figures for the reports and coordinated the publication of the results of this study.

DATA ENTRY AT THE NATIONAL RESEARCH CENTERS

Each national research center was responsible for transcribing the information from the student, teacher, and school questionnaires into computer data files. The data-entry software *DataEntryManager* (DEM), which had been adapted for the needs of the IEA Civic Education Study, was provided to each participating country, together with a *Manual for Entering the Civic Education Data* (IEA Civic Education Study, 1999a, 1999b) and codebooks defining the structure of the data. A codebook defined the variable names for each variable in the questionnaire, its position in the data file, and its field length, labels, valid ranges, missing values, and default values.

Because of the importance of the structure as defined by the codebooks provided with the DEM software, the NRCs were strongly encouraged to use DEM for entering the CivEd data. If, for example, for contractual reasons, countries had to use other data entry systems, output from other software had to meet the structure as defined in the codebook.

To facilitate data entry, the codebooks and their corresponding data files were designed to match the questionnaires. As a consequence, there was one data file per instrument. Each country was supposed to submit three data files, the *student data file*, the *teacher data file*, and the *school data file*.

- The *student data file* contained one record for every student listed on the student tracking form for each sampled class, including students who did not participate and students who were excluded or had left the school permanently. The data on the student data file came from the student tracking form, the teacher-student linkage form, and the student questionnaires.
- The *teacher data file* contained one record for each teacher. The data for this file came from the teacher tracking form and the teacher questionnaires.
- The *school data file* contained one record for each school sampled to take part in the study regardless of whether the school participated or not. The data for this file came from the school tracking form and the school questionnaires.

INFORMATION RECEIVED BY THE IEA DATA PROCESSING CENTER

In addition to the data file and the corresponding codebooks, countries were requested to submit additional information necessary for data verification. This included a report of all field-work activities, completed data management forms (including all national adaptations or changes to the questionnaires and the students' participation status), and copies of the national questionnaires.

DATA EDITING AT THE IEA DATA PROCESSING CENTER

Once each national research center had entered all data, it submitted the files to the IEA DPC for review and further processing into the international database, a process that is generally referred to as data cleaning or data editing. On the one hand, data editing for the IEA Civic Education Study aimed at identifying, documenting, and, if necessary, correcting deviations from the international file structure. On the other hand, the goal was to correct key-puncher errors, systematic deviations from the international data formats, linkage problems across data files, inconsistent tracking information both within and across files, and inconsistencies within and across observations. The main objective of this process was the creation of a database containing accurate, consistent, and comparable data.

Data editing consisted of several checks, some of which were repeated until satisfactory results could be achieved. During the first step, all incoming data were compared to the international file structure. Deviations like changes of field length and deletion of variables were corrected in order to perform all further checks with equally structured data files. In a second step, all problems with identification variables, participation status variables, valid ranges, and inconsistent answers were documented and, if necessary, solved. During this step, distributions of all variables were generated in order to detect unusual data. In a final step, data files were reviewed for possible linkage errors across files and, if necessary, corrected.

Two main processes could be distinguished in the data editing. First, if problems could be solved unambiguously, automatic changes were performed by a program. Second, if automatic changes could not be made, IEA DPC staff had to use auxiliary information to make the necessary changes to the data.

STANDARDIZATION OF THE NATIONAL FILE STRUCTURE

In the first step of data editing, the incoming data were compared to the internationally defined file structure. This step was necessary in order to perform all subsequent editing steps requiring a standardized file structure. Even though the codebooks provided with the DEM program gave clear and unambiguous instructions about the structure of the data files that each country was expected to submit to the DPC, some countries decided to use different data entry systems or to submit the data in different formats. In most cases, these changes were due to adaptations to the national design.

The instructions for entering the CivEd data asked the countries to use DEM and to submit all data emerging from DEM in extended dBase format files. However, data were received in fixed format ASCII files, tab-delimited, SPSS, or SAS formats.

After all data had been transformed into one standard format, the structure of the files was reviewed and compared to the international structure. All deviations were identified. A program scanned the files of each participating country to identify the following problems:
- international variables omitted;
- national variables added; and
- different variable length or decimal positions.

Together with the data files, the documentation provided by the national centers was reviewed. On the basis of the initial checks, the IEA DPC outlined and implemented the changes necessary to standardize the file structure.

CLEANING RULES AND PROCEDURES

After the transformation of national data into the international file structure, standard cleaning routines were applied to the data. In a first step, a diagnostic program detected possible problems in the data. During this step, the following groups of problems were under consideration:

- valid ranges;
- duplicate records;
- flaws in the ID system;
- participation status (from the tracking form) against data (from the questionnaire);
- participation status against exclusion status; and
- inconsistencies in participants' answers.

All problems detected in this step were reported in an error database. In a second step, routine programs were used to solve many of these errors. Most of them were due to the wrong usage of missing values. In a third step, remaining problems were addressed by manual operation, which required thorough reviews of tracking forms or further communication with the NRCs.

When all automatic and manual changes had been applied to the data, the diagnostic program used in the first step was applied again, and the resulting error database was compared to the previous error database. Ideally, at the end of this procedure, all problems should have been removed from the data. However, it sometimes happened that problems persisted or new problems were created in the data. The data cleaning steps then had to be repeated until no problems remained.

The review of file linkages between files was based on the information from all files involved in the study. Only records from participating respondents were used in this final data-cleaning step. During NRC meetings, the NRCs received detailed reports on database problems and the solutions applied to the data.

DATA PRODUCTS

Data products sent by the IEA DPC to the ICC included both data files and data summaries, depending on the needs of the ICC.

Data summaries contained unweighted summary statistics for each participating country on each variable in the survey instrument. The display for categorical variables, which represented the majority of variables in the data files, included the percentage of respondents choosing each of the options on the question, the percentage of respondents choosing none of the valid responses, and the percentage of students to whom the question had not been administered. These summaries were used for an in-depth review of the data. If, for example, no answer was found in a particular response category, the corresponding question as used in the country was scrutinized to detect possible printing errors. Table 5.1 shows an example of a summary statistic.

Table 5.1 Example of summary statistics for student questionnaire item

78 SQ2-2/7A BSGADU1 MOTHER/STEPM/GUAR LIVES AT HOME ALL THE TIME
Valid Ranges: BSGADU1$'1298'

Country	Omit	NR	0	1	2	3	4	5	6	7	8	9	10
AUS	5.6	•	•	6.3	88.1	•	•	•	•	•	•	•	•
BFR	5.2	•	•	14.6	80.1	•	•	•	•	•	•	•	•
BGR	3.2	•	•	14.1	82.7	•	•	•	•	•	•	•	•
CHE	3.6	•	•	6.7	89.7	•	•	•	•	•	•	•	•
CHL	5.4	•	•	9.2	85.4	•	•	•	•	•	•	•	•
COL	7.8	•	•	11.7	80.5	•	•	•	•	•	•	•	•
CYP	5.0	•	•	3.6	91.4	•	•	•	•	•	•	•	•
CZE	1.6	•	•	6.0	92.4	•	•	•	•	•	•	•	•
DEU	4.8	•	•	4.5	90.7	•	•	•	•	•	•	•	•
DNK	4.1	•	•	6.3	89.7	•	•	•	•	•	•	•	•
ENG	4.3	•	•	4.7	91.0	•	•	•	•	•	•	•	•
EST	5.5	•	•	6.0	88.4	•	•	•	•	•	•	•	•
FIN	4.4	•	•	5.8	89.8	•	•	•	•	•	•	•	•
GRC	6.0	•	•	10.3	83.7	•	•	•	•	•	•	•	•
HKG	2.8	•	•	4.8	92.4	•	•	•	•	•	•	•	•
HUN	4.5	•	•	10.6	84.9	•	•	•	•	•	•	•	•
ITA	•	•	•	•	•	•	•	•	•	•	•	•	•
LTU	5.9	•	•	3.9	90.2	•	•	•	•	•	•	•	•
LVA	7.3	•	•	8.0	84.7	•	•	•	•	•	•	•	•
NOR	5.3	•	•	6.9	87.8	•	•	•	•	•	•	•	•
POL	3.1	•	•	9.1	87.7	•	•	•	•	•	•	•	•
PRT	3.9	•	•	8.7	87.4	•	•	•	•	•	•	•	•
ROM	6.0	•	•	14.5	79.4	•	•	•	•	•	•	•	•
RUS	6.9	•	•	6.8	86.4	•	•	•	•	•	•	•	•
SVK	1.3	•	•	4.1	94.6	•	•	•	•	•	•	•	•
SVN	4.5	•	•	8.5	86.9	•	•	•	•	•	•	•	•
SWE	7.1	•	•	6.1	86.9	•	•	•	•	•	•	•	•
USA	2.2	•	•	5.3	92.5	•	•	•	•	•	•	•	•

Data files at various stages of the study, reflecting different states of data editing, were submitted to the ICC. First, the student files were submitted that had been processed completely, followed by teacher and school data. Once weights were available, the student file was augmented by the corresponding variables and submitted to the ICC.

In addition to the original data, a student-teacher linkage file had to be created, reflecting the existence of multiple teachers per student. This file, which could be matched to both the student data and the teacher data, contained adjusted student weights. Here, the weights of all teacher-linkages for one student, who could have up to three different teachers, added up to one student weight.

COMPUTER SOFTWARE

dBase was used as the standard database program for the incoming data. All subsequent programs, like diagnostic tools, programs to change the data, tools to create tables for item statistics or graphics related to this, and programs to create summary statistics, were produced in the SAS environment. For item analyses, the DPC made use of ACER Quest software (Adams & Khoo, 1996) and ACER *ConQuest* software (Wu, Adams, & Wilson, 1997).

Since SAS is rarely used by the Civic Education NRCs, all data were converted into SPSS format in a final step. These data files were submitted to the ICC for further analyses and later dispatched to the NRCs.

SUMMARY

The data processing for the IEA Civic Education Study was a cooperative process involving the NRCs, the IEA DPC, and the ICC. The structures and procedures implemented for processing data were found to be very successful and ensured a high quality of the database. The national research centers were provided with data entry facilities and detailed guidelines about their use and possible adaptation. The DPC then implemented complex procedures for data cleaning and editing, reviewed possible data errors, checked file linkages, provided data summary statistics, and successfully created the international fully weighted CivEd database consisting of student, teacher, and school data.

REFERENCES

Adams, R. J., & Khoo, S. (1996). *Quest: The Interactive Test Analysis System* (computer program). Camberwell, Vic.: Australian Council for Educational Research.

IEA Civic Education Study. (1999a). *Manual for entering the CivEd data* (CIVIC Reference No. 99–0004). Hamburg: IEA Data Processing Center.

IEA Civic Education Study. (1999b). *Manual for entering the CivEd data* (CIVIC Reference No. 99–0210). Hamburg: IEA Data Processing Center.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-Aspect Test Software* (computer program). Camberwell, Vic.: Australian Council for Educational Research.

Chapter 6:

WEIGHTING PROCEDURES[1]

Heiko Sibberns and Pierre Foy

OVERVIEW

The basic sample design used in the IEA Civic Education Study (CivEd) was the same as that used in TIMSS 1999—a two-stage stratified cluster design, with schools as the first stage and classrooms as the second. The design required schools to be sampled using a probability proportional to size (PPS) systematic method and classrooms to be sampled with equal probabilities. IEA Civic Education Study participants adapted the basic design to the requirements of their education systems, with guidance from the CivEd sampling consultants at the IEA Data Processing Center (DPC) and Statistics Canada. Very large countries could add an extra, preliminary stage, where districts or regions were sampled first, and then schools within districts. Participants used stratifications to improve the precision of their samples where appropriate. These adaptations could be quite complex, as can be seen from the information in Appendix D, which shows how the CivEd design was implemented in each country.

While CivEd's multi-stage stratified cluster design provided the basis for an efficient and manageable data collection in a school environment, it resulted in differential probabilities of selection for the ultimate sampling elements, the students. Consequently, one student in the assessment did not necessarily represent the same proportion of students in the population as another, as would have been the case with a simple random sampling approach. To account for differential probabilities of selection due to the design and to ensure proper survey estimates, CivEd computed a sampling weight for each participating student. The ability to provide proper sampling weights was an essential characteristic of an acceptable sample design in CivEd, since appropriate sampling weights were essential for the computation of accurate survey estimates. This chapter describes the procedure for calculating sampling weights for the CivEd data.

WEIGHTING PROCEDURE

The weighting procedure required three steps, reflecting the CivEd sample design. The first step consisted of calculating a school weight; this also incorporated weighting factors from any additional front-end sampling stages such as districts or regions. A school-level participation adjustment was then made to the school weight to compensate for any sampled schools that did not participate. That adjustment was calculated independently for each explicit stratum.

---

1 This chapter is based on Chapter 11 in the *TIMSS 1999 Technical Report* (Foy, 2000).

In the second step, a classroom weight was calculated. No classroom-level participation adjustment was necessary, since, in most cases, a single classroom was sampled in each school. If a school agreed to take part in the study but the sampled classroom refused to participate, the non-participation adjustment was made at the school level. If one of two selected classrooms in a school did not participate, the classroom weight was calculated as though a single classroom had been selected in the first place. The classroom weight was calculated independently for each school.

The third and final step consisted of calculating a student weight. A non-participation adjustment was made to compensate for students who did not take part in the testing. The student weight was calculated independently for each sampled classroom. The basic sampling weight attached to each student record was the product of the three intermediate weights: the first stage (school) weight, the second stage (classroom) weight, and the third stage (student) weight. The overall student sampling weight was the product of these three weights and the two non-participation adjustments—school level and student level.

SAMPLING WEIGHTS

The First Stage (School) Weight

The first stage weight represented the inverse of the first stage selection probability assigned to a sampled school. The CivEd sample design required that school selection probabilities be proportional to the school size (PPS), with school size being the enrolment in the target grade. The basic first-stage weight $BWsc$ for the ith sampled school was thus defined as

$$BW_{sc}^{i} = \frac{M}{n \cdot m_i}$$

where n was the number of sampled schools and m_i was the measure of size for the ith school. M—the total number of students in the stratum under consideration—was defined as

$$M = \sum_{i=1}^{N} m_i$$

where N was the total number of schools in the explicit stratum.

The basic first stage weight also incorporated weighting factors resulting from any additional front-end sampling stages that were applied. The calculation of such weighting factors was similar to that of the first stage weight, since geographical regions were also sampled using the PPS method. The resulting first stage weight in such cases was simply the product of the "region" weight and the first stage weight, as described earlier.

In some countries, schools were selected with equal probabilities. This generally occurred when no reliable measure of school size was available. In some countries, explicit strata had been defined to deal with very large schools or with small schools. In these strata, schools were sampled with equal probability.

Under equal probability sampling, the basic first-stage weight for the ith sampled school was defined as

$$BW^i_{sc} = \frac{N}{n}$$

where n was the number of sampled schools and N was the total number of schools in the explicit stratum. The basic weight for all sampled schools in an explicit stratum was identical in this context.

School Non-Participation Adjustment

First stage weights were calculated for all sampled schools and replacement schools that participated. A school-level participation adjustment was required to compensate for schools that were sampled but did not participate and were not replaced. Sampled schools that were found to be ineligible were removed from the calculation of this adjustment. The school-level participation adjustment was calculated separately for each explicit stratum. This adjustment, A_{sc}, was calculated as follows:

$$A_{sc} = \frac{n_s + n_{r1} + n_{r2} + n_{nr}}{n_s + n_{r1} + n_{r2}}$$

where n_s was the number of originally sampled and participating schools, n_{r1} and n_{r2} the number of participating first and second replacement schools, and n_{nr} the number of originally sampled schools that did not participate.

The final first stage weight, *FWsc*, for the ith school, corrected for non-participating schools, thus became

$$FW^i_{sc} = A_{sc} \cdot BW^i_{sc}$$

The Second Stage (Classroom) Weight

The second stage weight represented the inverse of the second stage selection probability assigned to a sampled classroom. Although almost all CivEd participants sampled intact classrooms using equal probability sampling, it also was permissible to sub-sample students within classes. Procedures for calculating sampling weights are presented below for both approaches.

Equal probability weighting: For the ith school, C^i was the total number of classrooms and c^i the number of sampled classrooms. Using equal probability sampling, the final second stage weight assigned to all sampled classrooms in the ith school was

$$FW^i_{cl1} = \frac{C^i}{c^i}$$

As a rule, c^i took the value 1 or 2 and remained fixed for all sampled schools. In those cases where c^i took the value 2 and only one of the sampled classrooms participated, the second stage weight was adjusted by multiplying it by 2.

The Third Stage (Student) Weight

The third stage weight represented the inverse of the third stage selection probability attached to a sampled student. Although almost all participants sampled intact classrooms where all eligible students were to be tested, some countries with large classes took a sub-sample of students from within the sampled classes. Procedures for calculating weights are presented below for both sampling approaches. The third stage weight was calculated independently for each sampled classroom.

Sampling intact classrooms: If intact classrooms were sampled, the basic third stage weight for the jth classroom in the ith school was simply

$$BW_{st1}^{i,j} = 1.0$$

Sub-sampling students within sampled classrooms: If students were sub-sampled within the sampled classrooms, the basic third stage weight for the jth classroom in the ith school was

$$BW_{st1}^{i,j} = \frac{N_{gi}}{r_{gi}}$$

where N_{gi} was the number of enrolled and eligible students in classroom g and r_{gi} the number of sampled students in this classroom.

Adjustment for Student Non-Participation

The student non-participation adjustment was calculated separately for each participating classroom as follows:

$$A_{st}^{i,j} = \frac{S_{rs} + S_{nr}}{S_{rs}}$$

Here, S_{rs} was the number of participating students and S_{nr} the number of non-participating students in the sampled classroom. This adjustment is the inverse of the unweighted student participation rate, R_{st}, for the corresponding classroom:

$$A_{st}^{i,j} = \frac{1}{R_{st}^{i,j}}$$

The third and final stage weight for the jth classroom in the ith school thus became

$$FW_{st1}^{i,j} = A_{st}^{i,j} \cdot BW_{st1}^{i,j}$$

Overall Sampling Weight

The overall sampling weight was simply the product of the final first stage weight, the final second stage weight, and the final third stage weight. When intact classrooms were tested, the overall sampling weight was

$$W^{i,j} = A_{sc} \cdot BW_{sc}^{i} \cdot FW_{cl1}^{i,j} \cdot A_{st}^{i,j} \cdot BW_{c/1}^{i,j}$$

or

$$W^{i,j} = FW_{sc}^{i} \cdot FW_{cl1}^{i,j} \cdot FW_{st1}^{i,j}$$

It is important to note that sampling weights varied by school and classroom, but that students within the same classroom had the same sampling weights.

PARTICIPATION RATES

Since lack of participation by sampled schools or students could lead to bias in the results, a variety of participation rates were computed to reveal how successful countries had been in securing participation from their sampled schools. To monitor school participation, three school participation rates were computed: originally sampled schools only; sampled and first replacement schools; and sampled and both first and second replacement schools. In addition, student and overall participation rates were computed.

Unweighted School Participation Rates

Two unweighted school participation rates were computed:

R_{unw}^{sc-s} = unweighted school participation rate for originally sampled schools only

R_{unw}^{sc-r2} = unweighted school participation rate, including sampled, first, and second replacement schools.

Each unweighted school participation rate was defined as the ratio of the number of participating schools to the number of originally sampled schools, including any ineligible schools. The rates were calculated as follows:

$$R_{unw}^{sc-s} = \frac{n_s}{n_s + n_{r1} + n_{r2} + n_{nr}}$$

$$R_{unw}^{sc-r2} = \frac{n_s + n_{r1} + n_{r2}}{n_s + n_{r1} + n_{r2} + n_{nr}}$$

Unweighted Student Participation Rate

The unweighted student participation rate for j classrooms within i schools was computed as follows:

$$R_{unw}^{st} = \frac{\sum_{i,j} s_{rs}^{i,j}}{\sum_{i,j} s_{rs}^{i,j} + \sum_{i,j} s_{nr}^{i,j}}$$

Unweighted Overall Participation Rates

Two unweighted overall participation rates were computed for each country. They were as follows:

R_{unw}^{ov-s} = unweighted overall participation rate for originally sampled schools only

R_{unw}^{ov-r2} = unweighted overall participation rate, including sampled, first, and second replacement schools.

For each country, the overall participation rate was defined as the product of the unweighted school participation rate and the unweighted student participation rate. They were calculated as follows:

$$R_{unw}^{ov-s} = R_{unw}^{sc-s} \cdot R_{unw}^{st}$$

$$R_{unw}^{ov-r2} = R_{unw}^{sc-r2} \cdot R_{unw}^{st}$$

Weighted School Participation Rates

In TIMSS 1995, the weighted school-level participation rates were computed using school sampling frame information. However, as in TIMSS 1999, the IEA Civic Education Study used student-level information. This alternate method has two advantages:

1. All data users can easily replicate school participation rates since all required data are available from the international database.

2. These rates more accurately reflect the current size of the target population because they rely on up-to-date within-school sampling information.

The TIMSS 1995 method relied on school data as reported on the sampling frame (see Dumais, 1998; Foy, Rust, & Schleicher, 1996), which often were not up to date with regard to current school enrolment. However, both methods are equivalent when assuming an up-to-date sampling frame, and so should yield comparable results in practice.

Two weighted school-level participation rates were computed using the alternate method. They were as follows:

$R_{wtd}^{sc-s} =$ weighted school participation rate for originally sampled schools only

$R_{wtd}^{sc-r2} =$ weighted school participation rate, including sampled, first, and second replacement schools.

The weighted school participation rates were calculated as follows:

$$R_{wtd}^{sc-s} = \frac{\sum_{i,j}^{s} BW_{sc}^{i} \cdot FW_{clx}^{i,j} \cdot FW_{stx}^{i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{clx}^{i,j} \cdot FW_{stx}^{i,j}}$$

$$R_{wtd}^{sc-r2} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot FW_{clx}^{i,j} \cdot FW_{stx}^{i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{clx}^{i,j} \cdot FW_{stx}^{i,j}}$$

Here, both the numerator and denominator were summations over all responding students, and the appropriate classroom-level and student-level sampling weights were used. Note that the basic school-level weight appears in the numerator, whereas the final (adjusted) school-level weight appears in the denominator.

The denominator remains unchanged in both equations and is the weighted estimate of the total enrolment in the target population. The numerator, however, changes from one equation to the next. Only students from originally sampled schools were included in the first equation; students from first and second replacement schools were added in the second equation.

Weighted Student Participation Rate

The weighted student response rate was computed as follows:

$$R_{wtd}^{st} = \frac{\sum\limits_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot FW_{clx}^{i,j} \cdot BW_{stx}^{i,j}}{\sum\limits_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{clx}^{i,j} \cdot FW_{stx}^{i,j}}$$

Here, both the numerator and denominator were summations over all responding students, and the appropriate classroom-level and student-level sampling weights were used. Note that the basic student weight appears in the numerator, whereas the final student weight appears in the denominator. Furthermore, the denominator in this formula was the same quantity that appeared in the numerator of the weighted school-level participation rate for all participating schools, including originally sampled and replacement schools.

Weighted Overall Participation Rates

Two weighted overall participation rates were computed. They were as follows:

$R_{wtd}^{ov-s} =$ weighted overall participation rate for originally sampled schools only

$R_{wtd}^{ov-r2} =$ weighted overall participation rate, including sampled, first, and second replacement schools.

The weighted overall participation rate was defined as the product of the appropriate weighted school participation rate and the weighted student participation rate. They were computed as:

$$R_{wtd}^{ov-s} = R_{wtd}^{sc-s} \cdot R_{wtd}^{st}$$

$$R_{wtd}^{ov-r2} = R_{wtd}^{sc-r2} \cdot R_{wtd}^{st}$$

SUMMARY

The multi-stage nature of the CivEd sampling design meant that students were sampled with varying probabilities. Consequently, if statistics computed from the sample data were to accurately reflect population values, the CivEd sampling weights had to be used when analyzing the data.

REFERENCES

Dumais, J. (1998). Implementation of the TIMSS sampling design. In M. O. Martin & D. L. Kelly (Eds.), *TIMSS technical report, volume III: Implementation and analysis—final year of secondary school* (pp. 15–35). Chestnut Hill, MA: Boston College.

Foy, P. (2000). Sampling weights. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 189–200). Chestnut Hill, MA: Boston College.

Foy, P., Rust, K., & Schleicher, A. (1996). Sample design. In M. O. Martin & D. L. Kelly (Eds.), *TIMSS technical report, volume I: Design and development* (pp. 4.1–4.17). Chestnut Hill, MA: Boston College.

Chapter 7:

SCALING PROCEDURES FOR COGNITIVE ITEMS

Wolfram Schulz and Heiko Sibberns

OVERVIEW

This chapter describes the procedures used to scale the test items in the IEA Civic Education Study (CivEd) that were administered to assess the domains of *civic content knowledge, skills in interpreting political communication*, and *economic literacy* (upper secondary students only). These procedures comprised the following steps:

- Analysis of dimensionality to determine the existence of sub-dimensions.
- Calibration of test items within participating countries and review of item statistics.
- Analysis of country-by-item interaction to safeguard the appropriateness of test items across participating countries.
- Adjudication of items based on national and international item statistics.
- Calibration of international items parameters.
- Subsequent scaling of test items using the international item parameters and transformation of maximum likelihood (logit) scores into an international metric.

The two basic reports on the IEA Civic Education Study (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002; Torney-Purta, Lehmann, Oswald, & Schulz, 2001) contain appendices that lay out the content framework around which the test was developed and serve as a reference for the topics of individual items.

THE SCALING MODEL

The tests used in the IEA Civic Education Study consisted of 38 multiple-choice items for the standard population of 14-year-old students and 43 multiple-choice items for the older population of upper secondary students. Sixteen items used for the assessment of the 14-year-old students were also included in the test for the upper secondary students to provide a link between both tests that would allow reporting of students' test scores for the two assessments on the same international metric.

Each multiple-choice item consisted of four answer categories, one with the correct response and three with incorrect responses (distracters). For scaling purposes, correct answers were coded as 1, and incorrect and not-attempted items as 0. The IRT One-Parameter Model (Rasch, 1960) for dichotomous data describes the probability of scoring 1 rather than 0 depending on the ability

θn. This is modeled as

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

where $P_i(\theta)$ is the probability of person *n* to score 1 on item *i*. θ_n is the ability of person *n* and δ_i the difficulty of item *i*.

Maximum likelihood (ML) estimates of ability are produced by maximizing the above equation with respect to θ_n, that is, solving the likelihood expressions

$$\sum_{i\delta\Omega} \left[r - \sum_{j=1}^{k} \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \right] = 0$$

for each case *n*, where r_x is the sum score obtained from a set of *k* items. This can be achieved by applying the Newton-Raphson method.[1] However, solving the above equation does not provide estimates for zero or perfect scores, that is, no estimates will be obtained for students with no correct answers or those who have answered all test items correctly.[2] Therefore, a small constant of .3 was added to the lowest category, and the same quantity was subtracted from the maximum scores in order to obtain estimates for these cases.[3]

The international item parameters were estimated with ACER ConQuest (Wu, Adams, & Wilson, 1997), using marginal maximum likelihood estimation. Here, item difficulty parameters are estimated conditional on the assumed distribution of abilities in a population; the likelihood depends on the item parameter and the parameters of the ability distribution with mean μ and standard deviation σ. ACER ConQuest was used for the estimation of the international item parameters and the ML ability estimates.[4]

In the one-parameter *Rasch* model, each raw score corresponds to one ability estimate, so that the percentage of correct answers in a single test design—as was used in the IEA Civic Education Study—typically shows a very high correlation with ML estimates. However, using the Rasch model instead of the percentage of correct answers or simple raw scores has the following advantages:

- The use of an IRT scaling model enables the researcher to apply a wider range of scaling analysis, including the assessment of item fit and the analysis of differential item functioning.

- Items with problematic item characteristics in one or more countries can be omitted in the scaling of the respective student data sets without jeopardizing the comparability of the results.

- In regard to the IEA Civic Education Study, the use of link items allowed comparable proficiency estimates for both populations to be estimated from

---

1 This method is described in detail in Hambleton and Swaminathan (1985).

2 In the survey of 14-year-olds, the percentage of students with zero scores was negligible and only an average of one percent had perfect scores. Similar figures were found in the survey of upper secondary students.

3 An alternative would have been the use of weighted likelihood estimation (see Warm, 1989), which provides ability estimates for zero and perfect scores.

4 For the national sub-samples, items were calibrated with the scaling software ACER Quest, which uses unconditional maximum likelihood (see Adams & Khoo, 1996).

the whole set of available test items. Comparing percentages of correct answers or raw scores would have limited reporting to the common subset of items used in both assessments.

The ML estimates reported in this study are attenuated, that is, they do include measurement error. Furthermore, a well-known property of ML estimates is that they are biased to a certain extent. In particular, variances tend to be overestimated when this kind of ability estimate is used. More sophisticated techniques, like the computation of so-called plausible values that are drawn from an expected distribution based on the scaling model and other available background information (conditioning variables), usually provide better population estimates and are used in international assessments like TIMSS and PISA (see Adams, 2002; Adams, Wu, & Macaskill, 1997; Mislevy, 1991). However, within the scope and timelines of the IEA Civic Education Study, it was not possible to apply more complex scaling methods.

ANALYSIS OF TEST DIMENSIONALITY

The civic knowledge test for the 14-year-old students consisted of two different types of items: those aiming primarily at *content knowledge* and those designed to measure *skills in interpreting political communication*. In addition, the civic knowledge test for the upper secondary students used a third type of item to assess *economic literacy*. Consequently, it was necessary to analyze the dimensionality of test items to find out whether sub-scales could be reported in addition to a combined total civic knowledge scale.

Confirmatory factor analyses (CFA) with structural equation modeling (SEM) of the dimensional item structure were presented in the international reports. The LISREL program (Jöreskog & Sörbom, 1993) was used for the parameter estimation, and graphical displays were included to illustrate the relationship between latent dimensions and items (Amadeo et al., 2002, p. 49; Torney-Purta et al., 2001, p. 61).

In addition to the CFA, multi-dimensional IRT models were estimated using the ACER ConQuest software (see Wu et al., 1997, pp. 85ff), which are more appropriate for the analysis of dichotomous test items. Table 7.1 shows the latent correlations between factors as derived from the multi-dimensional IRT models and the respective changes in deviances and degrees of freedom that illustrate the model fit. The estimated latent correlations are similar to those obtained from the CFA with structural equation modeling. Furthermore, the difference in deviance and its corresponding degrees of freedom[5] indicates that for both populations the multi-dimensional models provide a significantly better model fit than the alternative one-dimensional models.

The results of the analysis of item dimensionality indicated that both dimensions referred to highly similar but not identical components of student performance. Therefore, it was decided to report content knowledge and interpretative skills as sub-scales in addition to the combined civic knowledge scale.

---

5 The deviance is a statistic that indicates how well the item response model fits the data. When comparing the fit for two different models, this value can be compared to a chi-square distribution, where the degrees of freedom are equal to the difference in the number of parameters to be estimated for each model.

Table 7.1 Latent correlations between sub-dimensions of civic knowledge

	14-year-old Students	Upper Secondary Students
Latent correlation between...		
Content knowledge and interpretative skills	.925	.943
Economic literacy and content knowledge	–	.890
Economic literacy and interpretative skills	–	.929
Deviance for...		
One-dimensional model	570,045	316,998
Multi-dimensional model	569,557	317,493
Model fit		
Difference in deviance	488	494
Degrees of freedom	2	5
N of cases	14,000	8,000

Note: ACER ConQuest estimates for the international calibration sample of 14-year-olds.

Although the estimated latent correlations between content knowledge and interpretative skills were very high, the results showed that both dimensions referred to highly similar but not identical components of student performance. Therefore, it was decided to report two sub-scales in addition to the combined civic knowledge scale. Given that economic literacy was not generally regarded as part of civic education and that it also had a lower correlation with content knowledge, this item type was not included in the combined civic knowledge scale, and the results on economic literacy were reported separately for the upper secondary students.

ITEM ANALYSIS

The scaling properties of the test items were reviewed at two levels:

- At the international level, analyses were based on the international calibration sample, with 500 students randomly selected from each country.
- At the national level, within-country item analysis based on the country data sets indicated whether a test item had been translated correctly and whether it was appropriate for the national context of a country.

Items should have a good discrimination, that is, the correct answer to a test question needs to be highly correlated with the ability measured by a test. Conversely, the distracters used in a multiple-choice question should be negatively correlated. Test items should also cover a suitable range of abilities in the target population; test questions that are too hard or too easy do not contribute sufficiently to the estimation of ability.

The goodness of fit for individual test items can be determined by calculating a mean square statistic (MNSQ) (Wright & Masters, 1982), which can be obtained from the standardized residuals

$$z_i = \frac{x_i - \mathrm{E}(x_i)}{w_i}$$

where x_i is the observed scores on item i, $\mathrm{E}(x_i)$ the expected scores according to the Rasch model, and w_i the variance of item i. The variance for a dichotomous test item is calculated as

$$w_i = \sum_{n=1}^{N} P_{in} \left(1 - P_{in}\right)$$

where P_{in} is the probability for person n to answer item i correctly, and N denotes the number of test respondents. An unweighted (outfit) mean square statistic, u_i, can be obtained from

$$u_i = \sum_{n=1}^{N} z_i^2$$

As this statistic is likely to be affected by outliers, a weighted (infit) MNSQ statistic v_i is used, which is calculated as

$$v_i = \frac{\sum_{n=1}^{N} w_i z_i^2}{\sum_{n=1}^{N} w_i}$$

where the sum of standardized residuals z_i is weighted by the item variance w_i.

Values greater than 1.2 are typically viewed as an indication of poor item fit, whereas values below .8 indicate a discrimination that is considerably higher than expected. So-called item characteristic curves (ICCs) are plots of the average percentage of correct answers for ability groups against their expected performance on an item, and give additional graphical information on item fit.

The item statistics that were used for the item review at the national level were computed with the ACER QUEST software. The item analysis at the international level, however, was done with the ACER ConQuest program. The fit statistics provided by this software have a similar interpretation but their computation differs from the one outlined above (see Wu et al., 1997, p. 140ff).

International Item Analysis and Calibration

Figure 7.1 shows how the difficulty of test items developed to measure civic knowledge corresponded to the estimated latent abilities in the target population of 14-year-old students (ACER ConQuest estimates). The overall test was relatively easy but still covered an appropriate range of abilities.

The match was less satisfactory for the upper secondary students (see Figure 7.2). This was partly due to the need to use items from the test designed for 14-year-olds in order to obtain comparable ability scores. Furthermore, it was obvious that there were not sufficient items to describe higher levels of civic knowledge in this population. A closer match between abilities and test items was achieved for economic literacy (see Figure 7.3). However, the limited number of items places some constraints on the description of abilities in this domain.

Figure 7.1 Latent ability estimates and item thresholds for the total civic knowledge test items (14-year-old students)

```
                                                      |
  5                                                   |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                 X    |
  4                                                   |
                                                 X    |
                                                 X    |
                                                XX    |
                                                XX    |
                                               XXX    |
  3                                         XXXXXX    |
                                           XXXXXXX    |
                                             XXXXX    |
                                           XXXXXXX    |
                                       XXXXXXXXXXX    |
                                      XXXXXXXXXXXXX    |
                                      XXXXXXXXXXXXX    |
  2                                  XXXXXXXXXXXXXX    |
                                 XXXXXXXXXXXXXXXXXX    |
                                XXXXXXXXXXXXXXXXXXX    |
                               XXXXXXXXXXXXXXXXXXXX    |
                            XXXXXXXXXXXXXXXXXXXXXXXX   | 34
                       XXXXXXXXXXXXXXXXXXXXXXXXXXXXX   |
                          XXXXXXXXXXXXXXXXXXXXXXXXX    |
  1          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX    | 22
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX      | 21
                    XXXXXXXXXXXXXXXXXXXXXXXXXXXX       | 27  38
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX       | 15  17  26
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX       | 29  31
                XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX       | 18  36
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX    | 10  14  25  28
  0              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX        | 5  8  19  30
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXX          | 23  32
                      XXXXXXXXXXXXXXXXXX               | 4  7  13  35
                    XXXXXXXXXXXXXXXXXX                 | 9  24
                 XXXXXXXXXXXXXXXXXXXX                  | 11  12  37
                    XXXXXXXXXXXXX                     | 3  6
                     XXXXXXXXXXX                      | 2  20  33
 -1                 XXXXXXXXX                         | 1
                     XXXXXX                           |
                      XXXX                            |
                      XXX                             | 16
                       XX                             |
                       XX                             |
                        X                             |
 -2                                                   |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
                                                      |
 -3                                                   |
                                                      |
```
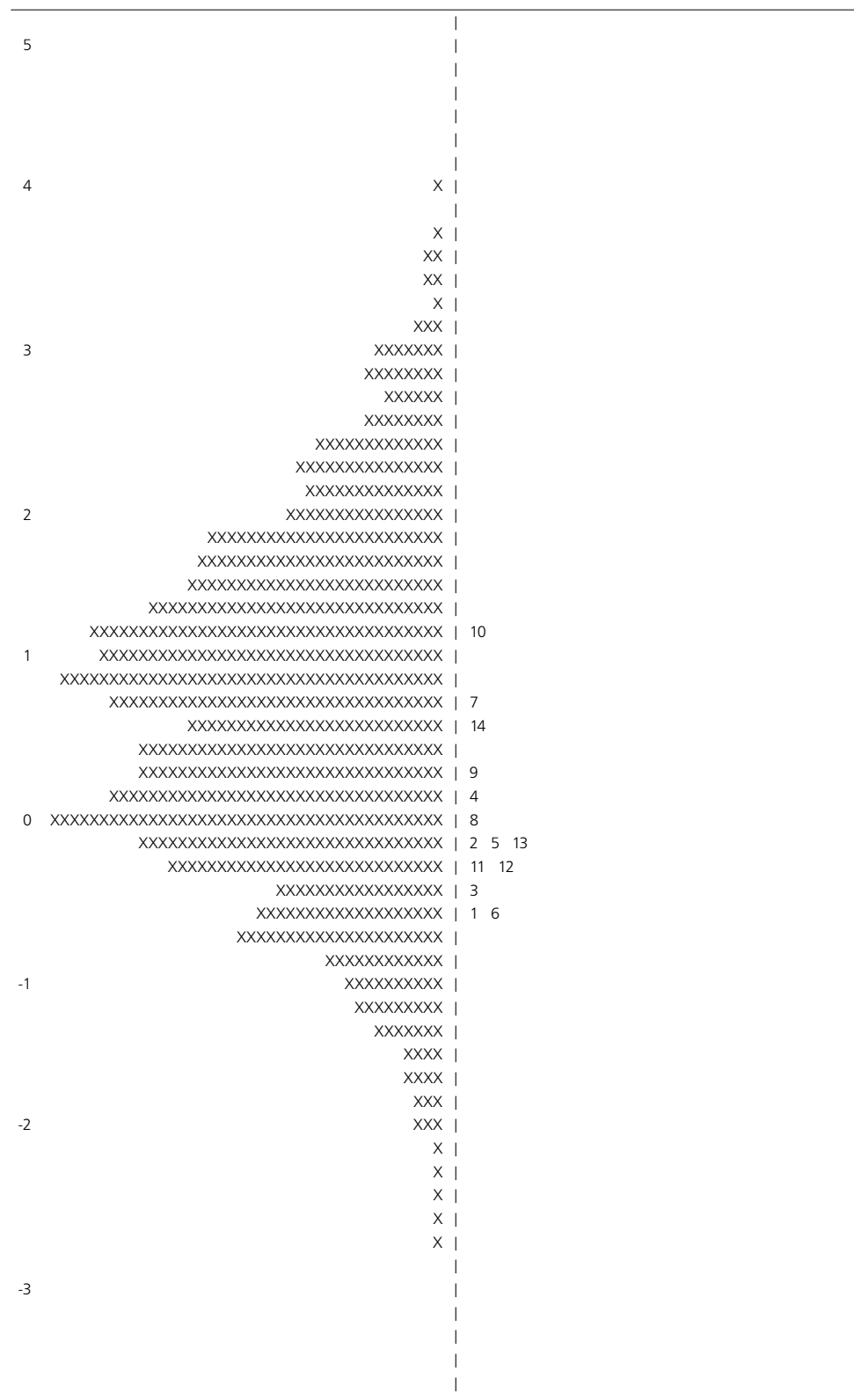
Note: ACER ConQuest estimates of latent ability distribution and test item thresholds (*rp* =.5) for the international calibration sample of 14-year-olds.

Figure 7.2 Latent ability estimates and item thresholds for the total civic knowledge test items (upper secondary students)

```
                                      |
 6                                    |
                                      |
                                      |
                                      |
                                      |
                                      |
 5                                  X |
                                      |
                                    X |
                                    X |
                                   XX |
                                   XX |
                                  XXX |
 4                             XXXXXXX |
                              XXXXXXX |
                               XXXXXX |
                             XXXXXXXX |
                          XXXXXXXXXXX |
                        XXXXXXXXXXXXXX |
 3                      XXXXXXXXXXXXXX |
                        XXXXXXXXXXXXXX |
                    XXXXXXXXXXXXXXXXXXX |
                     XXXXXXXXXXXXXXXXXX |
                    XXXXXXXXXXXXXXXXXXX |
                  XXXXXXXXXXXXXXXXXXXXXX |
 2           XXXXXXXXXXXXXXXXXXXXXXXXXXXX |
              XXXXXXXXXXXXXXXXXXXXXXXXXXX |
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
              XXXXXXXXXXXXXXXXXXXXXXXXXX |
              XXXXXXXXXXXXXXXXXXXXXXXXXXX | 9
             XXXXXXXXXXXXXXXXXXXXXXXXXXXX | 19
 1           XXXXXXXXXXXXXXXXXXXXXXXXXXX |
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXX | 10  15  17  21
             XXXXXXXXXXXXXXXXXXXXXXXXXX | 20  22  25
                XXXXXXXXXXXXXXXXX | 5  14  18  23
                XXXXXXXXXXXXXXXXX | 7  13
 0             XXXXXXXXXXXXXXXXXX | 26
                  XXXXXXXXXXXX | 6
                  XXXXXXXXXXXX | 11  24
                   XXXXXXXXXX | 12
                   XXXXXXXX | 3
                     XXXX | 4  16
-1                   XXXX |
                      XXX |
                      XXX |
                        X | 1  8
                        X |
                        X | 2
                          |
-2                      X |
                          |
                          |
                          |
                          |
                          |
-3                        |
```

Note: ACER ConQuest estimates of latent ability distribution and test item thresholds ($rp=.5$) for the international calibration sample of upper secondary students.

Figure 7.3 Latent ability estimates and item thresholds for economic literacy test items (upper secondary students)

```
                                                        |
    5                                                   |
                                                        |
                                                        |
                                                        |
                                                        |
                                                        |
    4                                               X   |
                                                        |
                                                X   |
                                               XX   |
                                               XX   |
                                                X   |
                                              XXX   |
    3                                      XXXXXX   |
                                          XXXXXXX   |
                                           XXXXX   |
                                          XXXXXXX   |
                                      XXXXXXXXXXXX   |
                                    XXXXXXXXXXXXXX   |
                                    XXXXXXXXXXXXX   |
    2                               XXXXXXXXXXXXXXX   |
                            XXXXXXXXXXXXXXXXXXXXXXX   |
                          XXXXXXXXXXXXXXXXXXXXXXXX   |
                          XXXXXXXXXXXXXXXXXXXXXXX   |
                          XXXXXXXXXXXXXXXXXXXXXXX   |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 10
    1            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   |
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   |
                XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 7
                   XXXXXXXXXXXXXXXXXXXXXXXXXX   | 14
                  XXXXXXXXXXXXXXXXXXXXXXXXXXX   |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 9
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 4
    0  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 8
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 2  5  13
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX   | 11  12
                 XXXXXXXXXXXXXXXXXX   | 3
                XXXXXXXXXXXXXXXXXXX   | 1  6
               XXXXXXXXXXXXXXXXXXX   |
                 XXXXXXXXXXX   |
   -1             XXXXXXXXX   |
                  XXXXXXXX   |
                  XXXXXXX   |
                   XXXX   |
                   XXXX   |
                    XXX   |
   -2               XXX   |
                      X   |
                      X   |
                      X   |
                      X   |
                      X   |
                        |
   -3                   |
                        |
                        |
                        |
                        |
```

Note: ACER ConQuest estimates of latent ability distribution and test item thresholds (*rp* =.5) for the international calibration sample of upper secondary students.

Figures 7.4 and 7.5 show the ICCs for test items 21 and 27 that had the highest weighted MNSQ values in the civic knowledge test for the upper secondary students. These items had a relatively low discrimination. As illustrated by the ICC plots, for both test items the observed ICCs were slightly shallower than the expected ones, that is, the percentages of correct answers were higher than expected for low achievers but lower than expected for more knowledgeable students. However, the item fit for both items was still within the acceptable range at the international level.

Figure 7.4 Item characteristic curve of test item number 21 (weighted MNSQ = 1.07)
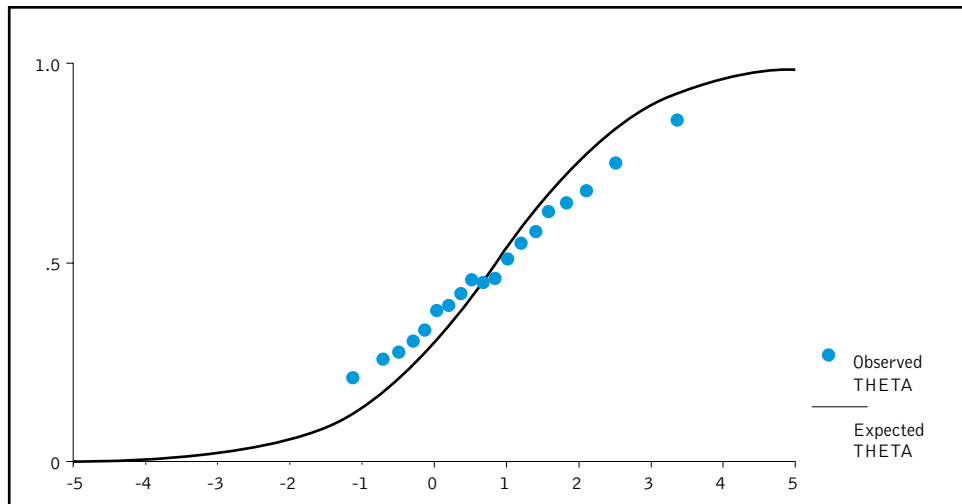
Figure 7.5 Item characteristic curve of test item number 27 (weighted MNSQ = 1.03)
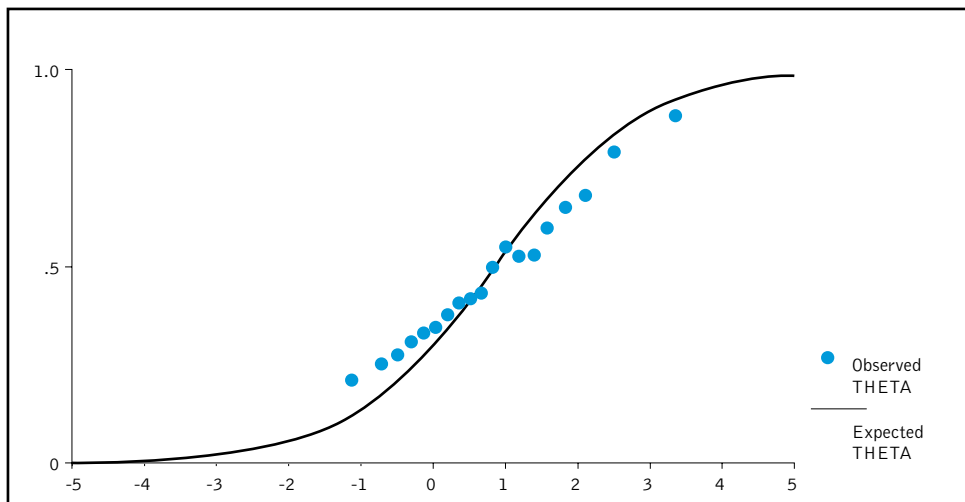
An additional criterion for the use of a test item on the international level was an acceptable item fit in more than 20 of the 28 participating countries. All 38 items used in the civic knowledge test for the 14-year-olds met this criterion and were included in the international item calibration. Three items used in the assessment of the upper secondary students did not meet this criterion and were excluded from scaling.

National Item Analyses and Item Adjudication

In addition to the international item analyses, a number of diagnostic statistics were computed for each item in each country. These statistics were carefully reviewed for any occurrence of unusual item behavior. The following indicated problematic item characteristics.

• An item was found to be relatively easier or harder than expected.

• The discriminating power of an item, as measured by its correlation with the overall score, was unusually low.

• An item did not meet the minimum requirements for fit statistics.

Whenever any of the above happened, the test booklets were reviewed for possible translation or printing problems, and, when necessary, the NRC was consulted. In addition, NRCs were asked to advise on whether a problematic item might be inappropriate for their national context. Items that were detected to have a flaw in a country were removed from the database for that country.

The basic statistics for the item review were calculated at the IEA Data Processing Center (DPC) and summarized in tabular and graphical form.[6] Item statistics were calculated for each of the 28 countries participating in the study of the 14-year-old students and for each of the 16 countries[7] participating in the study of the upper secondary students. If a country tested more than one grade, separate analyses for each grade were performed. For each item, the basic item analyses display included the following information (for an example, see Table 7.2):

• The number of students who responded in each country.

• The difficulty level (percentage of students that answered that item correctly).

• The discrimination index (point biserial correlation between success and the total score).

• The percentage of students that chose each option, including the percentages for omitting, not reached, and invalid answers.

• The point biserial correlation between each option and the total score.

• Preliminary Rasch results.

• The average item difficulty and the average item discrimination.

The item review tables contained the following statistics:

• *N (correct answers):* This is the number of students to whom the item was administered. If the item had not been attempted due to lack of time, the student data were coded as incorrect for the item analyses.

• *DIFF (correct answers):* The item difficulty was the percentage of students that provided a correct response to that item. When computing this statistic, not reached items were treated as wrong.

---

6 For the upper secondary students, the item statistics were prepared at the IEA International Coordinating Center.

7 Note that among these countries, data for Colombia and Hong Kong (SAR) were not included in the tables of the international report due to missing student weights.

Table 7.2 Example of statistics for item review (Item 1, 14-year-old students)

Population: '2'   Scale: ALL COL1 and COL 2   Item: BS101   Label: ROLE OF CITIZEN IN DEMOCR. COUNTRY   Page 1

COUNTRY	N	Correct Answer DIFF	DISCR	FLAGS	Percentages for Each Alternative A	B	C	D	W	OMIT	NR	Point Biserials for Each Alternative A	B	C	D	W	OMIT	NR	Rasch RDIFF	SE	FIT	Group	Difficulties	Intern. Mean IDIFF	IDISCR
AUS1	3301	84.4	.37	..q........	8.2	84.4*	3.9	2.6	0.0	0.8	0.0	-.21	.37*	-.23	-.15	.00	-.10	.00	-1.21	0.05	1.03	.	.	82.3	.37
BFR1	2074	79.8	.40	..q........	11.2	79.8*	4.8	2.7	0.1	1.4	0.0	-.22	.40*	-.23	-.17	.00	-.12	.00	-1.11	0.06	0.97	.	.	82.3	.37
BGR1	2854	86.7	.34	..q........	5.8	86.7*	4.4	2.6	0.0	0.5	0.0	-.17	.34*	-.25	-.12	.00	-.07	.00	-1.65	0.06	0.97	.	.	82.3	.37
CHL1	5677	53.3	.41	....	17.2	53.3*	11.3	14.7	0.0	3.5	0.0	-.16	.41*	-.23	-.14	.00	-.13	.00	0.17	0.03	1.05	.	.	82.3	.37
COL1	5038	36.4	.28	...Q..F...	43.9	36.4*	9.0	8.3	0.0	2.4	0.0	-.14	.28*	-.13	-.06	.00	-.07	.00	0.54	0.03	1.14	.	.	80.4	.37
CYP1	3093	82.0	.39	..q........	10.7	82.0*	1.9	4.8	0.0	0.5	0.0	-.24	.39*	-.22	-.19	.00	-.07	.00	-0.57	0.05	0.99	.	.	82.3	.37
CZE1	3599	93.7	.38	..q........	2.8	93.7*	1.6	1.8	0.0	0.1	0.0	-.24	.38*	-.19	-.21	.00	-.05	.00	-1.75	0.07	0.91	.	.	82.3	.37
DEU1	3750	82.4	.36	..q........	7.9	82.4*	2.5	3.8	0.7	2.7	0.0	-.16	.36*	-.17	-.20	-.09	-.15	.00	-1.13	0.05	1.00	.	.	82.1	.37
DNK1	3191	77.2	.44	..q........	12.6	77.2*	7.5	0.8	0.0	2.0	0.0	-.24	.44*	-.25	-.11	.00	-.20	.00	-0.80	0.05	1.00	.	.	80.5	.37
DNK2	2653	84.0	.41	..q........	9.8	84.0*	4.9	0.8	0.0	0.6	0.0	-.25	.41*	-.26	-.13	.00	-.12	.00	-0.90	0.06	1.01	.	.	80.5	.37
ENG1	3011	75.1	.36	..q........	16.1	75.1*	4.6	2.9	0.0	1.3	0.0	-.21	.36*	-.23	-.12	.00	-.10	.00	-0.64	0.05	1.06	.	.	82.1	.37
EST1	3418	82.4	.33	..q........	13.8	82.4*	1.5	1.3	0.1	0.9	0.0	-.26	.33*	-.12	-.11	-.03	-.09	.00	-1.33	0.05	0.98	.	.	82.3	.37
FIN1	3004	87.5	.42	..q........	7.5	87.5*	2.0	2.4	0.2	0.4	0.0	-.29	.42*	-.20	-.14	-.11	-.12	.00	-1.08	0.06	0.94	.	.	82.3	.37
FIN2	2942	93.3	.38	..q........	3.8	93.3*	1.1	1.2	0.1	0.5	0.0	-.24	.38*	-.21	-.14	-.03	-.14	.00	-1.41	0.08	0.95	.	.	82.3	.37
GRC1	3711	84.8	.41	..q........	11.0	84.8*	1.2	2.7	0.0	0.4	0.0	-.29	.40*	-.19	-.19	.00	-.07	.00	-0.89	0.05	0.97	.	.	82.3	.37
HKG1	4993	79.6	.34	..q...F...	11.5	79.6*	7.0	1.6	0.0	0.3	0.0	-.16	.34*	-.27	-.11	.00	-.07	.00	-0.73	0.04	1.15	.	.	80.5	.37
HUN1	3160	93.8	.29	..q........	4.1	93.8*	1.2	0.7	0.0	0.1	0.0	-.20	.29*	-.17	-.11	.00	-.04	.00	-2.18	0.08	0.96	.	.	82.3	.37
ITA1	3808	80.8	.36	..q........	14.7	80.8*	1.6	2.7	0.0	0.2	0.0	-.25	.36*	-.17	-.19	.00	-.04	.00	-0.60	0.05	1.04	.	.	82.3	.37
ITA2	3300	78.8	.36	..q........	15.4	78.8*	2.2	3.4	0.0	0.2	0.0	-.24	.36*	-.19	-.19	.00	-.03	.00	-0.75	0.05	1.03	.	.	82.3	.37
LTU1	3494	80.7	.36	..q........	13.7	80.7*	2.3	1.8	0.0	1.5	0.0	-.26	.36*	-.16	-.12	.00	-.11	.00	-1.26	0.05	0.98	.	.	82.3	.37
LVA1	2840	83.6	.36	..q........	9.3	83.6*	2.6	3.4	0.0	1.2	0.0	-.23	.36*	-.17	-.18	.00	-.08	.00	-1.51	0.05	0.96	.	.	82.3	.37
NOR1	3325	79.5	.43	..q........	11.4	79.5*	4.1	3.4	0.5	1.1	0.0	-.25	.43*	-.22	-.19	-.07	-.10	.00	-0.72	0.05	0.99	.	.	82.3	.37
NOR2	1016	83.7	.46	..q........	9.5	83.7*	3.4	1.9	0.1	1.4	0.0	-.31	.46*	-.25	-.11	.01	-.17	.00	-0.65	0.09	0.96	.	.	82.3	.37
POL1	3759	93.2	.32	..q........	3.9	93.2*	0.7	1.8	0.0	0.3	0.0	-.23	.32*	-.13	-.16	.00	-.06	.00	-1.88	0.07	0.96	.	.	80.8	.37
PRT1	3234	62.9	.39	....	14.4	62.9*	10.4	11.3	0.1	0.9	0.0	-.11	.39*	-.25	-.22	-.06	-.04	.00	-0.07	0.04	1.02	.	.	82.3	.37
PRT2	3162	74.5	.45	....	11.1	74.5*	6.5	7.2	0.0	0.7	0.0	-.20	.45*	-.26	-.23	-.02	-.10	.00	-0.09	0.05	0.96	.	.	82.3	.37
ROM1	3031	82.6	.39	..q........	6.3	82.6*	3.2	7.6	0.1	0.2	0.0	-.20	.39*	-.18	-.24	-.06	-.06	.00	-1.54	0.05	0.94	.	.	82.3	.37
RUS1	2129	83.3	.35	P.q........	9.1	83.3*	5.3	1.8	0.0	0.5	0.0	-.20	.35*	-.23	-.15	.00	-.02	.00	-1.25	0.06	1.02	.	.	82.3	.37
SVK1	3554	96.1	.26	P.q........	2.3	96.1*	0.6	0.9	0.0	0.1	0.0	-.19	.26*	-.08	-.15	.02	-.01	.00	-2.27	0.09	0.97	.	.	82.3	.37
SVN1	3081	92.7	.29	..q........	2.8	92.7*	1.5	2.4	0.0	0.6	0.0	-.17	.29*	-.16	-.15	.02	-.06	.00	-2.06	0.07	0.96	.	.	82.3	.37
SWE1	3157	72.6	.41	..q........	17.9	72.6*	5.6	2.7	0.3	0.9	0.0	-.22	.41*	-.25	-.18	-.06	-.08	.00	-0.49	0.04	1.03	.	.	80.5	.37
SWE2	3110	79.3	.44	..q........	13.5	79.3*	3.8	2.5	0.4	0.5	0.0	-.27	.44*	-.27	-.17	-.06	-.09	.00	-0.56	0.05	1.02	.	.	80.5	.37
USA1	2802	75.1	.37	....	12.5	75.1*	6.3	5.9	0.0	0.3	0.0	-.23	.37*	-.16	-.18	.00	-.05	.00	-0.37	0.05	1.10	.	.	82.3	.37

- *DISCR (correct answers):* The item discrimination was computed as the correlation between correctly answering the item and the total score. This correlation should always be moderately positive for items with good measurement properties.
- *ALTERNATIVES A, B, C, D (percentages):* These represent the percentage of students choosing each response option for the item. Not reached items were included in the denominator for these calculations.
- *W (percentages):* This was the percentage of students that provided an invalid response to a multiple-choice item. Invalid responses were typically the result of choosing more than one response option.
- *OMIT (percentages):* This was the percentage of students who did not provide a response to the item. Not reached items were included in the denominator for these calculations.
- *NR (percentages):* This is the percentage of students who did not reach the item. An item was coded as not reached if there was no response to any of the items following it and if the response to the preceding item had been omitted.
- *ALTERNATIVES A, B, C, D (point biserials):* These present the correlation between choosing each of the response options A, B, C, or D and the total score. Items with good psychometric properties have zero or negative correlations for the distracter options (the incorrect answers) and positive correlations for the correct answers. The point biserials were not corrected for auto-correlation, that is, the total score was not reduced by the value for the item under consideration.
- *W (point biserials):* This presents the correlation between an invalid response (usually caused by selecting more than one response option) and the total score. This correlation should always be negative or zero.
- *OMIT (point biserials):* This is the correlation between a binary variable indicating an omitted response to the item and the total score. This correlation should always be negative or zero.
- *NR (point biserials):* This is the correlation between a binary variable indicating a not reached response to the item and the total score. This correlation should always be negative or zero.
- *RDIFF (Rasch):* This is an estimate of the item difficulty based on a one-parameter IRT model. The difficulty of the item is expressed on the logit metric (with a positive logit indicating a difficult item) and was scaled so that the sum of all Rasch item difficulties was zero within each country.
- *SE (Rasch):* This is the standard error of the item difficulty parameter.
- *FIT (Rasch):* This is the weighted MNSQ value. Items that fit the Rasch model have a fit value of 1.

As an aid to reviewers, the item analyses tables included a variety of flags, signaling the presence of one or more possible problems. The following conditions were flagged:

- Item difficulty exceeds 95 percent in the sample as a whole.
- Item difficulty is less than 25 percent in the sample as a whole.

- One or more distracter percentages are less than five percent.
- One or more distracter percentages are greater than the percentage of the correct answer.
- The point biserial correlation for one or more of the distracters exceeds zero.
- Item discrimination is less than 0.2.
- Rasch goodness-of-fit index is less than 0.88 or greater than 1.12.

Although not all of these conditions necessarily indicate a problem, flags are a useful way to draw attention to possible sources of concern.

To provide an immediate overview of the key findings, the results of the item analyses were also displayed in a graphical format. Figure 7.6 shows examples as they were used in the CivEd item reviews. Four graphs were produced:

- Percentage correct across countries.
- Item discrimination (point-biserial for correct answer).
- Fit statistics (weighted MNSQ statistic minus 1).
- Item-by-country interaction.

Here, the expectation was that countries showing an average high performance would perform relatively well on each of the items used in the test, while low-performing countries would do less well on each of the items. When this did not occur, that is, a high-performing country showed low performance on an item on which other countries did well, this was called an *item-by-country interaction*. Since large item-by-country interactions can indicate some kind of item bias, this aspect of item performance was also included in the figures.

The graphical display for each item shows the average probability of a correct response across all countries, compared with the probability of a correct response in a particular country. The probability for each country is presented as a 95 percent confidence interval.

Prior to the final scaling of the CivEd data, item statistics, as described above, were reviewed carefully to guarantee comparable results. Although only a small number of inappropriate items were detected, several causes for problematic item characteristics were discovered:

- Errors during translation that had not been corrected.
- Item analyses showing negative biserial correlations.
- Item-fit statistics that indicated that an item did not fit the model in a country.

If, in a country, a serious problem[8] with an item was detected during the review, it was discarded from this particular national dataset for the international calibration and scaling. Table 7.3 shows the items that were excluded from scaling for each participating country and illustrates that in no country were more than 10 percent of the items excluded from the scaling.

---

8 Problems with items were typically due to translation errors not detected in the translation verification or a different meaning for the item in the particular national context.

Figure 7.6 Example of displays for item percentage correct, discrimination, fit statistics, and item-by-country interaction

Percent Correct

Discrimination

Fit (-1)

Item-by-country Interaction

IEA CIVIC EDUCATION STUDY **TECHNICAL REPORT**
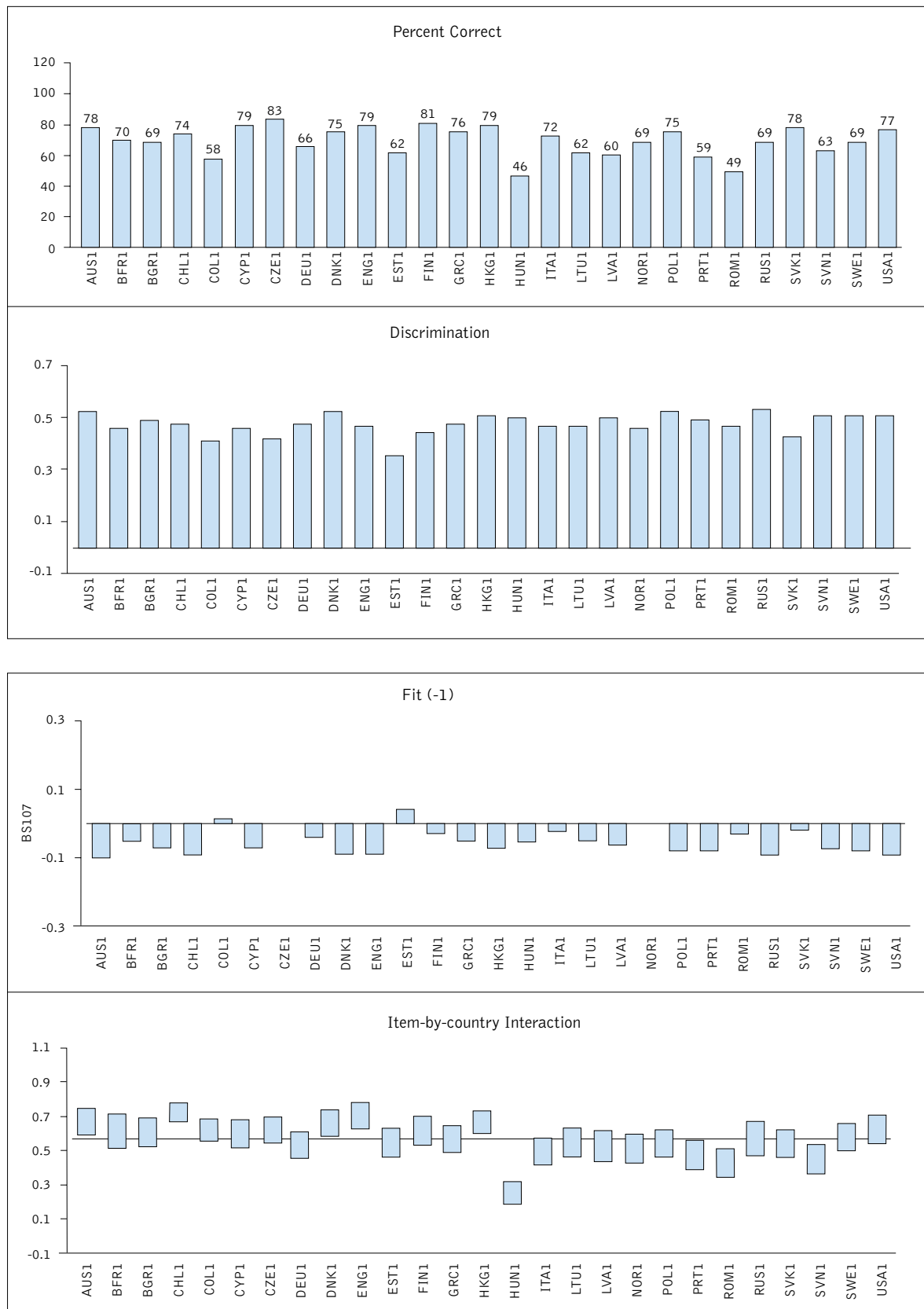
Table 7.3 Items excluded from scaling after item adjudication (14-year-old students)

Country	Items Excluded from Scaling
Australia	12
Belgium (French)	27
Bulgaria	21, 25
Chile	22
Colombia	22
Cyprus	2, 22, 27
Czech Republic	
Denmark	10, 20
England	12
Estonia	21
Finland	25, 27
Germany	
Greece	2, 22, 35
Hong Kong (SAR)	2, 12, 22
Hungary	20
Italy	35
Latvia	
Lithuania	
Norway	
Poland	31
Portugal	25
Romania	25
Russia	21
Slovak Republic	
Slovenia	27
Sweden	21
Switzerland	
United States	22, 34

Similar procedures were followed for the upper secondary students. Table 7.4 shows that only a small number of civic knowledge and economic literacy items were excluded from scaling.

Table 7.4 Items excluded from scaling after item adjudication (upper secondary students)

Country	Items Excluded from Scaling
Chile	
Colombia	
Cyprus	9, 25, 38
Czech Republic	
Denmark	
Estonia	
Hong Kong (SAR)	
Israel	34
Latvia	
Norway	
Poland	29
Portugal	
Russian Federation	29
Slovenia	
Sweden	
Switzerland	36

Linking Test Results

The tests for the 14-year-old students and the upper secondary students included a common set of 16 items. Table 7.5 shows the number of common and unique items for each of the proficiency scales.

Table 7.5 Numbers of common and unique test items

	14-year-old Students	Upper Secondary Students
Unique for 14-year-old students	22	
Unique for upper secondary students		10
Common items	16	
Total civic knowledge	38	26
Economic literacy items		14

To place ability estimates for civic knowledge on the same scale, item parameters for common items from the calibration for the 14-year-olds were used for the scaling of the upper secondary students. Three of the 16 common items had to be discarded because their relative difficulties for the upper secondary students were too different compared to the ones found for the 14-year-olds. Therefore, only 13 out of the common set of 16 items were used for the equating. However, the three common items not used for anchoring were included in the scaling with re-estimated item parameters, that is, they were treated as if they were unique items. One of these three items (number 10)

was not used for the anchoring of the total civic knowledge scale but for the anchoring of the interpretative skills scale.

ITEM CALIBRATION AND COMPUTATION OF STUDENT SCORES

The international item parameters were estimated for each scale based on a calibration sample with 500 randomly selected students per country (14,000 students from the 14-year-old population). Based on the results of the item adjudication, some items were excluded from both item calibration and scaling.

Notably in two countries (Chile and Colombia), a considerable number of students were not able to attempt all test items.[9] Missing students' responses that were likely to be due to problems with test length ("not reached items")[10] were omitted from the calibration of item parameters but were treated as incorrect for the scaling of student responses. Table 7.6 shows the international percentages of correct responses and the item parameters for the three cognitive scales for each item used for the assessment of 14-year-olds in 1999.

The international item parameters and percentage correct for the three civic knowledge scales for the upper secondary students are shown in Table 7.7; anchored item parameters are printed in bold. The item parameters for the economic literacy scale are given in Table 7.8.

The international item parameters obtained from the calibration samples were used to compute ML estimates for each sub-sample. The person parameters (logits) were transformed to the international metric with an international mean of 100 and a standard deviation of 20. The transformation was achieved by applying the formula

$$\theta'_n = 100 + 20 \left(\frac{\theta_n - \bar{\theta}}{\sigma_\theta} \right)$$

where θ'_n are the student scores in the international metric, θ_n the original logit scores, $\bar{\theta}$ the international mean of student logit scores with equally weighted country sub-samples, and σ_θ its corresponding international standard deviation.

Table 7.9 shows the means and standard deviations of student logit scores (14-year-olds tested in 1999) used for the transformation into the international metric of student scores. The same values as for the upper secondary students were used to compute the international student scores for the three civic knowledge scales for the upper secondary students. For the economic literacy scale, which was used only for the assessment of the upper secondary students, the means and standard deviations of the equally weighted country sub-samples of these students were used, so that the mean on the economic literacy scale was 100 and the standard deviation 20 for all 16 participating countries (including Colombia and Hong Kong (SAR), whose data were not included in the reporting).

---

9 For example, more than 10 percent of the Colombian students did not attempt any of the last eight test items.

10 "Not reached items" were defined as all consecutive missing values starting from the end of the test except for the first missing value of the missing series, which was coded as "missing".

Table 7.6 Percentage correct and item parameters for civic knowledge test items (14-year-old students)

Item Number	International Percentage of Correct Answers	Item Parameters for...		
		Total civic knowledge	Content knowledge	Interpretative skills
1	79	-0.874	-0.786	
2	78	-0.858	-0.772	
3	78	-0.756	-0.665	
4	70	-0.311	-0.216	
5	65	-0.010	0.087	
6	77	-0.715	-0.623	
7	69	-0.249	-0.153	
8	64	0.055	0.154	
9	72	-0.411	-0.317	
10	59	0.241	0.341	
11	75	-0.557	-0.465	
12	71	-0.579	-0.477	
13	67	-0.153	-0.059	
14	61	0.201		0.028
15	53	0.625	0.725	
16	85	-1.289	-1.206	
17	53	0.577	0.676	
18	57	0.401	0.499	
19	65	-0.048	0.046	
20	77	-0.839	-0.753	
21	47	0.856	0.949	
22	42	1.058	1.163	
23	65	-0.020		-0.202
24	71	-0.355		-0.542
25	58	0.223		0.054
26	50	0.734		0.558
27	47	0.856	0.959	
28	62	0.122	0.216	
29	54	0.560	0.657	
30	66	-0.071	0.022	
31	53	0.580		0.409
32	66	-0.122		-0.311
33	77	-0.813		-1.014
34	35	1.551		1.365
35	67	-0.249		-0.429
36	57	0.355		0.173
37	72	-0.511		-0.703
38	49	0.794		0.616

Table 7.7 Percentage correct and item parameters for civic knowledge test items (upper secondary students)

Item Number	International Percentage of Correct Answers	Item Parameters for...		
		Total civic knowledge	Content knowledge	Interpretative skills
1	91	-1.078	-0.801	
2	92	**-0.715**	**-0.623**	
3	85	**-0.249**	**-0.153**	
4	86	**0.055**	**0.154**	
5	72	**0.625**	**0.725**	
6	80	**0.577**	**0.676**	
7	75	**0.401**	**0.499**	
8	90	**-0.756**	**-0.665**	
9	54	1.684	1.717	
10	67	0.845		**0.173**
11	80	**-0.020**		**-0.202**
12	83	**-0.355**		**-0.542**
13	75	**0.223**		**0.054**
14	71	**0.560**	**0.657**	
15	68	**0.734**		**0.558**
16	86	**0.122**	**0.216**	
17	68	0.993	1.079	
18	71	0.610	0.728	
20	60	1.499		0.978
23	67	1.057		0.588
25	68	0.826	0.921	
28	70	0.895		0.429
29	71	0.817		0.394
30	80	0.211		-0.237
31	70	0.770		0.308
43	75	0.291		-0.159

Note: Anchored item parameters are given in bold type.

Table 7.8 Percentage correct and item parameters for economic literacy test items (upper secondary students)

Item Number	Percentage of Correct Answers	Item Parameters for Economic Literacy
19	75	-0.628
21	66	-0.077
22	72	-0.393
26	61	0.220
32	66	-0.135
33	73	-0.493
34	50	0.653
35	63	-0.008
36	58	0.212
37	39	1.218
38	69	-0.435
39	68	-0.353
40	66	-0.279
42	52	0.500

Table 7.9 International means and standard deviations for IRT scales (logits)

Scale	Mean	Standard Deviation
Content knowledge	0.95	1.36
Interpretative skills	0.63	1.49
Total civic knowledge	0.82	1.29
Economic literacy	0.67	1.55

Note: Based on the equally weighted country data for the 14-year-olds tested in 1999 for content knowledge, interpretative skills, and total civic knowledge. Means and standard deviation for economic literacy are for the population of upper secondary students tested in 1999/2000.

Tables 7.10 and 7.11 respectively show the country means and standard deviations on the total civic knowledge scale for the 14-year-old students and the upper secondary students; the standard errors were estimated using the *jackknife repeated replication* (JRR) technique (see Chapter 10).

Table 7.10 Means and standard deviations for the civic knowledge scale (14-year-old students)

Country	Means		Standard Deviations	
Australia	101.7	(0.79)	20.3	(.31)
Belgium (French)	94.7	(0.94)	18.3	(.49)
Bulgaria	97.6	(1.25)	19.5	(.95)
Chile	88.4	(0.69)	16.9	(.26)
Colombia	86.4	(0.91)	15.1	(.40)
Cyprus	108.1	(0.49)	19.4	(.38)
Czech Republic	102.6	(0.83)	18.8	(.36)
Denmark	100.4	(0.54)	20.9	(.30)
England	99.4	(0.62)	18.9	(.39)
Estonia	94.0	(0.54)	16.2	(.34)
Finland	109.3	(0.68)	20.5	(.29)
Germany	99.8	(0.50)	18.8	(.41)
Greece	107.9	(0.76)	21.4	(.38)
Hong Kong (SAR)	106.6	(1.10)	22.7	(.44)
Hungary	101.6	(0.63)	17.9	(.28)
Italy	105.4	(0.77)	19.3	(.41)
Latvia	91.5	(0.85)	16.5	(.47)
Lithuania	93.6	(0.71)	16.6	(.39)
Norway	102.9	(0.48)	20.5	(.36)
Poland	110.6	(1.69)	22.4	(.80)
Portugal	96.2	(0.73)	16.5	(.47)
Romania	91.6	(0.88)	16.5	(.54)
Russian Federation	99.6	(1.33)	21.2	(.89)
Slovak Republic	105.4	(0.72)	17.0	(.48)
Slovenia	100.6	(0.46)	17.7	(.30)
Sweden	99.1	(0.78)	19.6	(.50)
Switzerland	98.3	(0.80)	17.3	(.47)
United States	106.5	(1.18)	22.4	(.52)

Note: Standard errors are given in brackets.

STANDARDIZING THE INTERNATIONAL ITEM DIFFICULTIES

To assist readers with understanding the cognitive scales, item difficulty maps were produced showing the location of items on the international civic knowledge scale. Item difficulty parameters consequently had to be transformed from their original metric to the international metric (with a student mean of 100 and a standard deviation of 20).

The estimated item difficulties indicated the location on the ability scale where a student had a 50 percent chance of giving the correct answer. For the *item difficulty maps,* the preference was to map items to a level of greater student proficiency. It was therefore decided to describe the scale using those points of proficiency where a student had a 65 percent chance of giving the correct answer.

Table 7.11 Means and standard deviations for the civic knowledge scale (upper secondary students)

Country	Means		Standard Deviations	
Chile	107.7	(0.59)	17.7	(.27)
Cyprus	118.2	(1.53)	20.0	(.49)
Czech Republic	121.1	(0.71)	19.9	(.37)
Denmark	132.3	(0.46)	19.0	(.24)
Estonia	124.8	(0.91)	20.1	(.36)
Israel	117.6	(0.97)	22.7	(.44)
Latvia	104.8	(1.45)	20.1	(.54)
Norway	121.8	(1.23)	23.3	(.43)
Poland	117.8	(1.08)	21.1	(.49)
Portugal	121.1	(0.65)	17.8	(.30)
Russia	110.9	(1.25)	21.6	(.59)
Slovenia	115.3	(1.46)	22.0	(.95)
Sweden	130.1	(0.89)	24.2	(.69)
Switzerland	123.7	(1.92)	19.8	(.82)

Note: Standard errors are given in brackets.

The transformation was achieved by adding the natural log of the odds of 65 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 (in logits). The standardized item difficulty d^l_i for each item was obtained as:

$$d^l_i = 100 + 20 \text{ x} \left(\frac{d_i + 1n(.65/.35) - \bar{\theta}}{\sigma_\theta} \right)$$

where d_i is the item difficulty in its original metric, $\bar{\theta}$ the international student mean, and σ_θ its corresponding standard deviation in the original metric (see Table 7.7). For the described scale, only the item difficulties on the total civic knowledge scale were used; the reporting of the results for the upper secondary students in addition presented a described scale for economic literacy.

SUMMARY

The IRT Rasch model was used for scaling of cognitive data. IRT methodology provides a useful tool for analysing test items, produces sample-independent measures, and allows the exclusion of problematic items from scaling for individual countries without jeopardizing the comparability of scores.

Item fit was assessed both for the international sample and the national sub-samples, and problematic items were excluded from scaling. In general, the cognitive test items had good scaling properties and most of the cognitive material was used for the measurement of civic knowledge and economic literacy.

Both CFA and multi-dimensional IRT scaling were used to assess the dimensionality of the cognitive items used in the IEA Civic Education Study.

Two sub-scales were derived from the cognitive test of 14-year-olds and three sub-scales from the assessment of the upper secondary students.

The use of IRT scaling allowed the computation of comparable proficiency scores for both populations through the use of a common set of anchor items. The international metric of the content knowledge, interpreting skills, and combined civic knowledge scales referred for both populations to a mean of 100 with a standard deviation of 20 in the assessment of 14-year-olds in 1999. The metric of the economic literacy scale had a mean of 100 and a standard deviation of 20 for the upper secondary students.

REFERENCES

Adams, R. (2002). Scaling PISA cognitive data. In R. Adams & M. Wu (Eds.), *Technical report for the OECD Programme for International Student Assessment* (pp. 99–108). Paris: OECD Publications.

Adams, R., & Khoo, S.-T. (1996). *Quest: The interactive test analysis system* (computer program). Camberwell, Vic.: Australian Council for Educational Research.

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, volume II: Implementation and analysis—primary and middle school years* (pp. 111–145). Chestnut Hill, MA: Boston College.

Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries.* Amsterdam: International Association for the Evaluation of Educational Achievement.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston, MA: Kluwer Academic Publications.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: User's reference guide.* Chicago: Scientific Software International.

Mislevy, R. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*(2), 177–196.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielsen & Lydiche.

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

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450.

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

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ACER ConQuest: Multi-aspect test software* (computer program). Camberwell, Vic.: Australian Council for Educational Research.

Chapter 8:

SCALING PROCEDURES FOR LIKERT-TYPE ITEMS ON STUDENTS' CONCEPTS, ATTITUDES, AND ACTIONS

Wolfram Schulz

OVERVIEW

This chapter describes the scaling procedures for Likert-type items in the student data of the IEA Civic Education Study (CivEd). It describes the analyses of the dimensional structure of items using structural equation modeling (SEM) and the item response theory (IRT) scaling methodology used to obtain individual scores on the latent dimensions.

The items on concepts, attitudes, and actions formed a vital part of the design in this subject area and took up nearly half of the instrument (see Appendix F, section F.3). They included:

- *Concepts:* Items that assess the concepts of students regarding *democracy* (Section A), *citizenship* (Section B), and *government responsibilities* (Section C).

- *Attitudes:* Items reflecting attitudes towards *trust in institutions* (Section D), *national identity* (Section E), *rights for women, minorities, and anti-democratic groups* (Section G), immigrants (Section H), *perceived relationship between government and the individual* (Political Efficacy and Interest, Section I), *participation in school* (Section J), and the *perception of teaching styles in class* (Section N).

- *Actions:* Items concerned with *participation in political discussions, media use for political information* (Section L), and *expected political participation in the future* (Section M).

The scaling process consisted of the following steps:

- Analysis of missing values, distribution of item responses, and exploratory principal component analyses.

- Confirmatory factor analysis (CFA) using SEM of the theoretically expected models based on data from an international sample (200 students per country) and the national samples.

- Selection of scales based on theoretical and empirical grounds.

- Analysis of IRT (partial credit) models for the selected scales based on an international sample (200 students per country).

- Computation of comparative item statistics (item fit and reliabilities) across countries.

- Item adjudication based on a country-by-country item calibration that resulted in excluding items with poor scaling properties from scaling in particular countries.

The final scaling was done using a calibration sample of 500 students from the population of 14-year-old students randomly selected from the weighted country data. Item parameters were estimated for the calibration sample and used as anchors for subsequent scaling of country data sets. Data from the upper secondary students were scaled using the same item parameters as for the 14-year-old students. This was done to make scale scores comparable across populations. Information about the sources of these items in the research literature is found in Torney-Purta, Lehmann, Oswald, & Schulz (2001). In addition, both this volume and Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova (2002) contain extensive analyses of the scales derived from these items by country and gender.

CONFIRMATORY FACTOR ANALYSES (CFA)

SEM was used to confirm theoretically expected dimensions and, if necessary, to re-specify the dimensional structure. SEM takes the measurement error associated with the indicators into account and provides a tool for analyzing the dimensional item structure and the estimation of the correlation between latent factors.

In CFA with SEM, an expected covariance matrix is fitted according to the theoretical factor structure. This can be done due to the possibility of computing the covariances from the estimates in the model. Maximum likelihood estimation (MLE) provides model estimates by minimizing the differences between the expected (Σ) and the observed covariance matrix (S).[1]

Measures for the overall fit of a model are then obtained by comparing the expected matrix with the observed S matrix. If the differences between both matrices are close to zero, then the model "fits the data"; if differences are rather large, the model "does not fit the data" and some re-specification may be necessary or, if this is not possible, the theoretical model has to be rejected. The chi-square test statistic for the null hypothesis of $\Sigma = S$ becomes a rather poor fit measure with larger sample sizes because even small differences between matrices are given as significant deviations.

The assessment of model fit for the CFA of CivEd data was based on the following measures:

- The *root mean square error of approximation* (RMSEA) measures the "discrepancy per degree of freedom for the model" (Browne & Cudeck, 1993, p. 144). A value of .05 and less indicates a close fit, values of .08 and more indicate a reasonable error of approximation, and values greater than 1.0 typically lead to the rejection of a model.

- The *adjusted goodness-of-fit index* (AGFI) is a "traditional" measure of model fit. It indicates the amount of variance in S explained by Σ and should be close to 1.0 to indicate a good model fit.

---

1 For ordinal variables, Jöreskog and Sörbom (1993) recommend using *weighted least square estimation* (WLS) with polychoric correlation matrices and corresponding asymptotic covariance weight matrices because MLE requires normal distribution and continuous variables. However, for the purpose of analysing the dimensional structure of items, the use of covariance matrices and MLE was deemed appropriate.

- The *non-normed fit index* (NNFI) and *comparative fit index* (CFI) are based on the differences between the specified model and a null model. Both NNFI and CFI have the advantage of being less dependent on sample size and providing a better adjustment for model complexity than the AGFI (Hox & Bechger, 1998, p. 364; Gerbing & Anderson, 1993, pp. 56–59). Their values should be close to 1.0.

Additionally, so-called *item reliability* was viewed as another indicator of model fit. If the explained variance for an item in a model is very low, it does not have acceptable scaling properties. Items with low item reliability are usually removed from scaling.

The estimation of a correlation between error terms should generally be avoided unless it is justified by a substantive explanation (Jöreskog, 1993, p. 297). For some of the models presented in this chapter, correlated error terms were specified to show the effects of similar wording or content. This was done for mere illustrative purposes, and correlated error terms were not reflected in subsequent scaling with IRT.

Model estimation was done with LISREL (Jöreskog & Sörbom, 1993) using the STREAMS (Gustafson & Stahl, 2000) interface program, which makes data handling and model specification easier. In view of the considerable amount of missing responses, group means were substituted for missing values.

IRT SCALING FOR CATEGORICAL ITEMS

In the case of categorical items with k categories, the one-parameter (Rasch) model can be generalized to

$$P_{x_i}(\theta) = \frac{\exp\left(\sum_{j=0}^{x}\theta_n - \delta_i + \tau_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{x}\theta_n - \delta_i + \tau_{ij}\right)}, \; x = 0, 1, 2, ..., m_i$$

where $P_{xi}(\theta)$ denotes the probability of person n to score x on item i. Here, τ_{ij} denotes an additional step parameter and θ_n denotes the location of a person on the latent dimension. The item parameter δ_i gives the location of the item on the latent continuum. In the case of attitudinal items, low values denote that an item is relatively easy to agree with, high values that an item is relatively hard to agree with.

The so-called partial credit model (PCM) estimates different step parameters for items on the same scale (Masters & Wright, 1997) whereas the rating scale model (RSM) has the same step parameters for all items on a scale (Andersen, 1997). For Likert-type items where all items on a scale have identical categories, the RSM is theoretically more appropriate and also provides a more parsimonious model. However, analysis of the CivEd data revealed that the model fit for the RSM was generally unsatisfactory, and the PCM was chosen as the scaling model for the categorical items in this study.

Because each step parameter is defined with respect to its corresponding neighbor categories, the steps are not necessarily ordered. This so-called *step disordering* does not indicate that the category definitions are out of sequence

but only that a category defines a very narrow section of the variable. In the case of *category disordering* (i.e., when the means of the latent measure are out of sequence across categories), the need to collapse adjacent categories may arise. However, no cases of category disordering were found in the scaling analyses for this study.

Model fit was assessed using mean square statistics (see Wright & Masters, 1982; also Chapter 7 in this volume). The value of the item fit statistics should be close to 1.0 to indicate a good fit according to the model. Values greater than 1.0 indicate that the item discrimination is less than expected, while values lower than 1.0 indicate that the item discrimination is higher than expected. As the unweighted mean square residual (outfit) statistic may be affected by a small number of outlying observations, the weighted mean square residual (infit) statistic was used as the main criterion for assessing item fit. Rasch parameters were estimated using the ConQuest program (Wu, Adams, & Wilson, 1997).

If the PCM fits the data, person parameters for the latent dimension can be computed and used for subsequent analysis. Though highly correlated with the original raw scores, this method provides a sophisticated scaling method for dealing with missing values because estimates for the latent dimension may be obtained for all respondents who have answered at least one of the items.

It should be noted that measurement error is included in the ML estimates. More sophisticated methods like the use of plausible values and conditioning on background variables can be used to obtain disattenuated population estimates of latent variables (see Adams, Wu, & Macaskill, 1997; Mislevy, 1991). But within the scope and timeline of the IEA Civic Education Study, and in view of the large number of scales, it was not possible to use more complex scaling techniques.

DESCRIPTION OF SCALES AND ITEMS

Concept of Citizenship

Fifteen items on citizenship behavior were administered to the students. Five of these items were discarded from scaling after a preliminary analysis of item dimensionality. Table 8.1 lists the concept items on citizenship, classified along the following dimensions:

- A "conventional" factor (CTCON) loading on items regarding the desirability of a citizen being politically active in conventional forms of participation (voting, parties, information).

- A "social movement" factor (CTSOC) loading on items indicating the desirability of having an active citizenship engaged with new forms of political participation as found in social movements aiming at the defense of human or civil rights, the environment, etc.

Figure 8.1 shows the two-factor solution for these items, which had an acceptable model fit. The estimated correlation between the latent dimensions was quite high (.74). Whereas the item variance explained by CTCON ranged between approximately 20 and 30 percent, the item factor correlation varied

substantively for the second dimension: around 40 percent of the variance in B9 and B11 but only 16 percent of the variance in B5 were explained by CTSOC. The RMSEA was below .50 in only two countries; in a few countries, it was close to .080. Both scales were chosen for scaling and reporting.

Table 8.2 shows the Rasch item parameters for both scales. The location parameters show that students found more passive conventional citizenship characteristics, like voting or reading, easier to agree with than active ones, such as joining a party or engaging in political discussions. Among the social-movement-related citizenship characteristics, the differences between item locations are smaller. Here, item B5 (protest activities) is shown as the item the students were least likely to endorse.

Table 8.3 shows the internal consistency of the scales for both populations. Reliabilities (Cronbach's Alpha) across countries ranged from .52 to .80 for CTCON, and from .51 to .80 for CTSOC.

Table 8.1 Item wording for citizenship items and dimensions

CTCON: Importance of Conventional Citizenship *An adult who is a good citizen . . .*	
B2	votes in every election
B3	joins a political party
B6	knows about the country's history
B8	follows political issues in the newspaper, on the radio, or on TV
B10	shows respect for government representatives [leaders, officials]
B12	engages in political discussions
CTSOC: Importance of Social-movement-related Citizenship *An adult who is a good citizen . . .*	
B5	would participate in a peaceful protest against a law believed to be unjust
B9	participates in activities to benefit people in the community [society]
B11	takes part in activities promoting human rights
B13	takes part in activities to protect the environment

Note: Categories—not important, somewhat unimportant, somewhat important, very important.

Figure 8.1 Two-factor solution for citizenship items
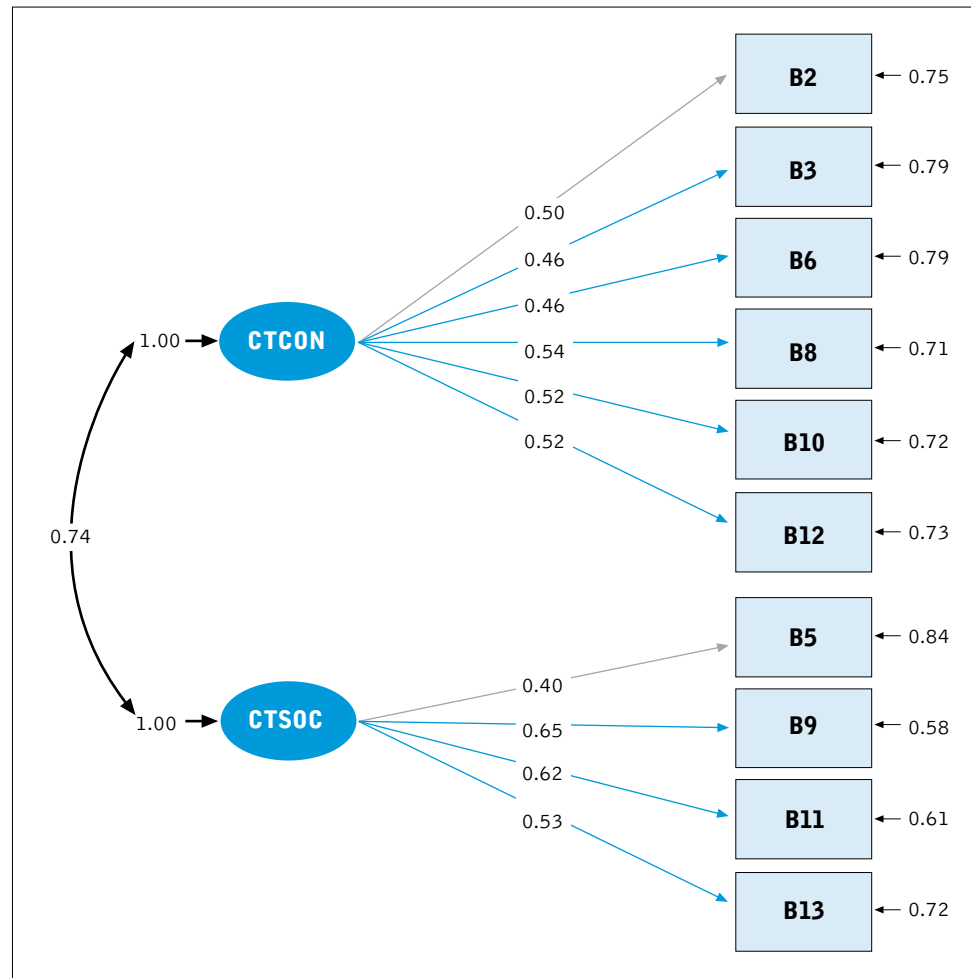
Note: Standardized minimum likelihood estimates for international sample of 14-year-olds. RMSEA = .056, AGFI = .96, NNFI = .91, CFI = .93.

Table 8.2 Item parameters for citizenship scales

CTCON		Location	Tau 1	Tau 2	Tau 3
B2	votes in every election	-0.64	-0.85	-0.35	1.21
B3	joins a political party	0.90	-1.30	0.18	1.12
B6	knows about the country's history	-0.31	-0.72	-0.18	0.90
B8	follows political issues	-0.19	-1.05	-0.48	1.53
B10	shows respect for representatives	-0.29	-1.13	-0.36	1.49
B12	engages in political discussions	0.53	-1.57	0.09	1.48
CTSOC		Location	Tau 1	Tau 2	Tau 3
B5	participates in a peaceful protest	0.52	-0.97	-0.30	1.27
B9	participates in community activities	-0.15	-1.28	-0.47	1.75
B11	activities promoting human rights	-0.29	-1.06	-0.38	1.44
B13	activities to protect the environment	-0.08	-1.04	-0.37	1.41

Table 8.3 Scale reliabilities for concept of citizenship

Country	CTCON		CTSOC	
	14-year-old students	*Upper secondary students*	*14-year-old students*	*Upper secondary students*
Australia	.69		.69	
Belgium (French)	.52		.54	
Bulgaria	.73		.65	
Chile	.60	.65	.51	.63
Colombia	.65	.66	.55	.60
Cyprus	.59	.69	.63	.69
Czech Republic	.63	.66	.62	.62
Denmark	.64	.64	.60	.72
England	.67		.62	
Estonia	.58	.64	.56	.65
Finland	.70		.66	
Germany	.61		.56	
Greece	.62		.62	
Hong Kong (SAR)	.71	.68	.63	.64
Hungary	.63		.60	
Israel		.68		.67
Italy	.60		.61	
Latvia	.57	.62	.53	.61
Lithuania	.68		.58	
Norway	.69	.69	.59	.64
Poland	.65	.69	.58	.68
Portugal	.60	.63	.51	.66
Romania	.63		.57	
Russian Federation	.57	.62	.62	.62
Slovak Republic	.60		.58	
Slovenia	.64	.66	.56	.65
Sweden	.71	.75	.71	.80
Switzerland	.63	.71	.57	.67
United States	.75		.73	
International Sample	**.68**	**.70**	**.63**	**.70**

Scope of Government

Items regarding the scope of government asked students to rate the extent to which the government should have responsibilities for 12 different tasks (see Table 8.4). Two dimensions were expected:

- Economy-related government responsibilities (GOVEC), assigned to those items indicating a desire for government intervention in economic relationships.

- Society-related government responsibilities (GOVSO), assigned to those items indicating more general areas of government control.

Table 8.4 Items on government responsibilities and dimensions

GOVEC: Economy-related Government Responsibilities *What responsibilities should the government have?*	
C1	To guarantee a job for everyone who wants one.
C2	To keep prices under control.
C5	To provide industries with the support they need to grow.
C6	To provide an adequate [decent] standard of living for the unemployed.
C7	To reduce differences in income and wealth among people.
GOVSO: Society-related Government Responsibilities *What responsibilities should the government have?*	
C3	To provide basic health care for everyone.
C4	To provide an adequate [decent] standard of living for old people.
C8	To provide free basic education for all.
C9	To ensure [be sure there are] equal political opportunities for men and women.
C10	To control pollution of the environment.
C11	To guarantee peace and order [stability] within the country.
C12	To promote honesty and moral behavior among people in the country.

Note: Categories—definitely not, probably not, probably, definitely.

The two-factor solution had a satisfactory model fit (see Figure 8.2). Both latent factors were strongly correlated, with $r = .82$. In all countries, the RMSEA was below .080, and in some countries the model showed a close fit to the data.

Here, the decision to choose a two-dimensional approach was taken both on theoretical and empirical grounds. Though the correlation between both factors was high, the two-factor structure showed a better model fit than the alternative one-factor solution. Furthermore, the inclusion of two different scales in this section was seen as theoretically justified and more interesting for the reporting of country differences.

Figure 8.2 Two-factor solution for items on government responsibilities
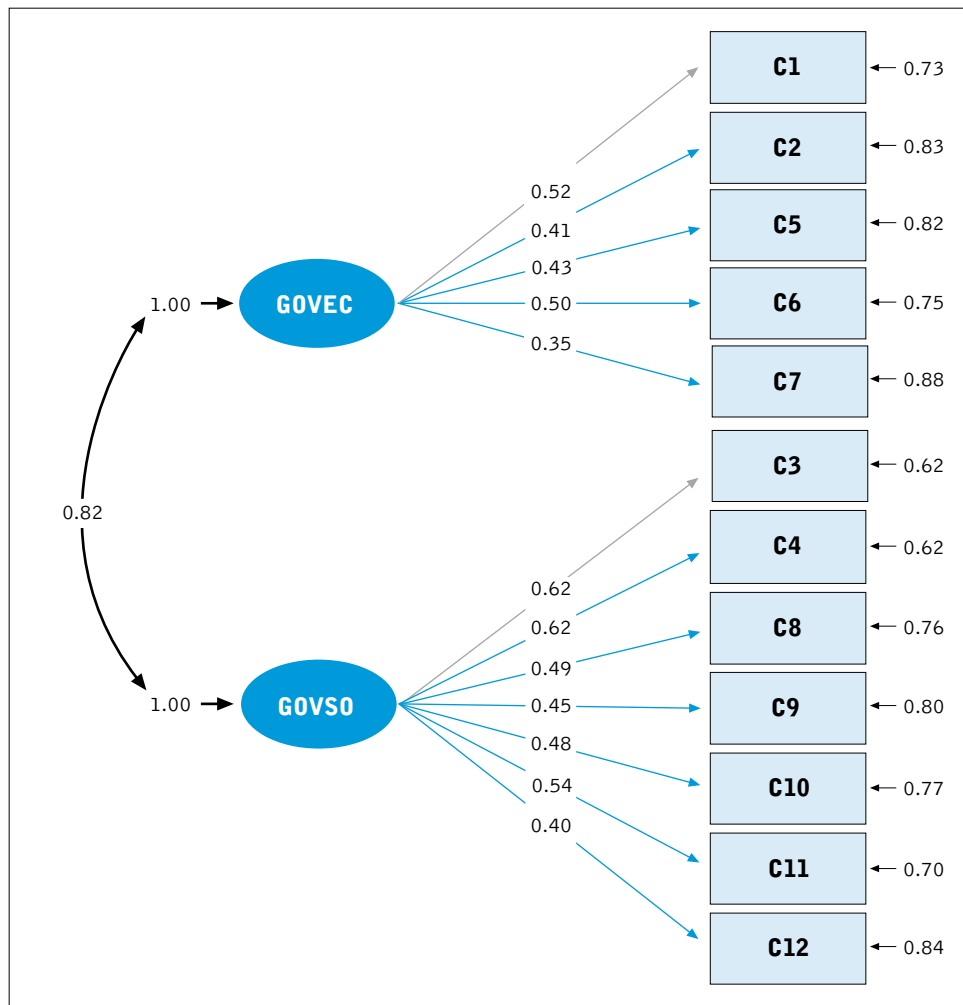
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .046, AGFI = .97, NNFI = .92, CFI = .93.

Table 8.5 shows the item and step parameters used for both scales. As can be seen, most item location parameters are rather close to each other. The item that students were least likely to endorse on GOVEC contained the statement that the government should reduce differences in income and wealth. For several items (C1, C4, C11) step disordering occurred, an indication that some categories described a rather narrow range of the latent continuum.

Table 8.5 Item parameters for scope of government scales

GOVEC		Location	Tau 1	Tau 2	Tau 3
C1	guarantee a job for everyone	-0.11	-0.24	-0.40	0.64
C2	keep prices under control	-0.15	-0.63	-0.49	1.11
C5	provide support for industries	0.05	-1.00	-0.41	1.41
C6	standard of living for unemployed	-0.14	-0.58	-0.44	1.02
C7	reduce differences in income	0.35	-0.74	-0.20	0.94
GOVSO		**Location**	**Tau 1**	**Tau 2**	**Tau 3**
C3	basic health care for everyone	-0.04	-0.27	-0.24	0.51
C4	standard of living for old people	-0.25	-0.47	-0.60	1.07
C8	free basic education for all	-0.04	-0.31	-0.11	0.42
C9	equal political opportunities	0.01	-0.40	-0.29	0.68
C10	control pollution	0.16	-0.67	-0.16	0.83
C11	guarantee peace and order	-0.34	-0.02	-0.24	0.26
C12	promote honesty and moral behavior	0.50	-0.74	-0.29	1.04

Table 8.6 shows the internal consistency for the scale across countries in both populations. Scale reliabilities for GOVEC ranged from .36 to .83, with the average reliability below .60 for both populations. Alpha coefficients for GOVSO ranged from .62 to .87, and the reliability for the pooled international sample was .71 for the 14-year-old students and .69 for the upper secondary students.

Table 8.6 Scale reliabilities for scope of government scales

	GOVEC		GOVSO	
Country	14-year-old students	Upper secondary students	14-year-old students	Upper secondary students
Australia	.63		.78	
Belgium (French)	.52		.73	
Bulgaria	.83		.87	
Chile	.52	.54	.73	.66
Colombia	.58	.60	.64	.66
Cyprus	.50	.61	.65	.76
Czech Republic	.54	.49	.70	.61
Denmark	.45	.57	.63	.62
England	.59		.71	
Estonia	.61	.61	.73	.70
Finland	.64		.75	
Germany	.36		.62	
Greece	.59		.67	
Hong Kong (SAR)	.63	.59	.77	.77
Hungary	.54		.71	

table contd. on next page

Table 8.6 contd. from previous page

Country	GOVEC		GOVSO	
	14-year-old students	Upper secondary students	14-year-old students	Upper secondary students
Israel		.64		.73
Italy	.54		.69	
Latvia	.51	.53	.70	.73
Lithuania	.55		.66	
Norway	.52	.54	.69	.67
Poland	.58	.54	.77	.65
Portugal	.53	.52	.66	.63
Romania	.61		.75	
Russian Federation	.55	.63	.73	.72
Slovak Republic	.53		.67	
Slovenia	.55	.56	.66	.68
Sweden	.51	.65	.68	.70
Switzerland	.50	.43	.64	.59
United States	.61		.72	
International Sample	**.57**	**.58**	**.71**	**.69**

Attitudes of Trust

For the items on trust in institutions, two dimensions were assumed: trust in government-related institutions (TRUST) (including political parties) and trust in media (MEDIA). Additional items regarding trust in the United Nations, schools, and the people living in the country were seen as unrelated to these latent dimensions. Table 8.7 lists the items and their corresponding dimensions.

Table 8.7 Items on trust in institutions and dimensions

TRUST: Trust in Government-related Institutions *How much of the time can you trust each of the following institutions?*	
D1	The national [federal] government
D2	The local council or government of town or city
D3	Courts
D4	The police
D8	Political parties
D11	National Parliament [Congress]
MEDIA: Trust in Media *How much of the time can you trust each of the following institutions?*	
D5	News on television
D6	News on the radio
D7	News in the press [newspapers]

Note: Categories—never, only some of the time, most of the time, always.

The resulting two-factor solution had a moderate fit (RMSEA = .067). The model fit improved after a correlation between the error terms for Items 3 (courts) and 4 (police) was added to the model. As both institutions probably are perceived as very close to each other and trust or distrust in a country's judicial system might affect both police and courts, a common residual variance between these two items is plausible. The final model in Figure 8.3 had a close fit, and the dimensional structure was also confirmed across country sub-samples. Only TRUST was retained for further scaling and inclusion in the international report.

Table 8.8 shows the item parameters for TRUST. The item locations indicate that students were least likely to endorse "political parties" (Item D8) as trustworthy, but were more likely to rate the courts (Item D3) and the police (Item D4) as trustworthy. The step parameters are spread over the continuum, an indication that each category defines a broader range on the TRUST dimension. Scale reliabilities were highly satisfactory for both populations, ranging between .63 and .83 (see Table 8.9).

Figure 8.3 Two-factor solution for trust items
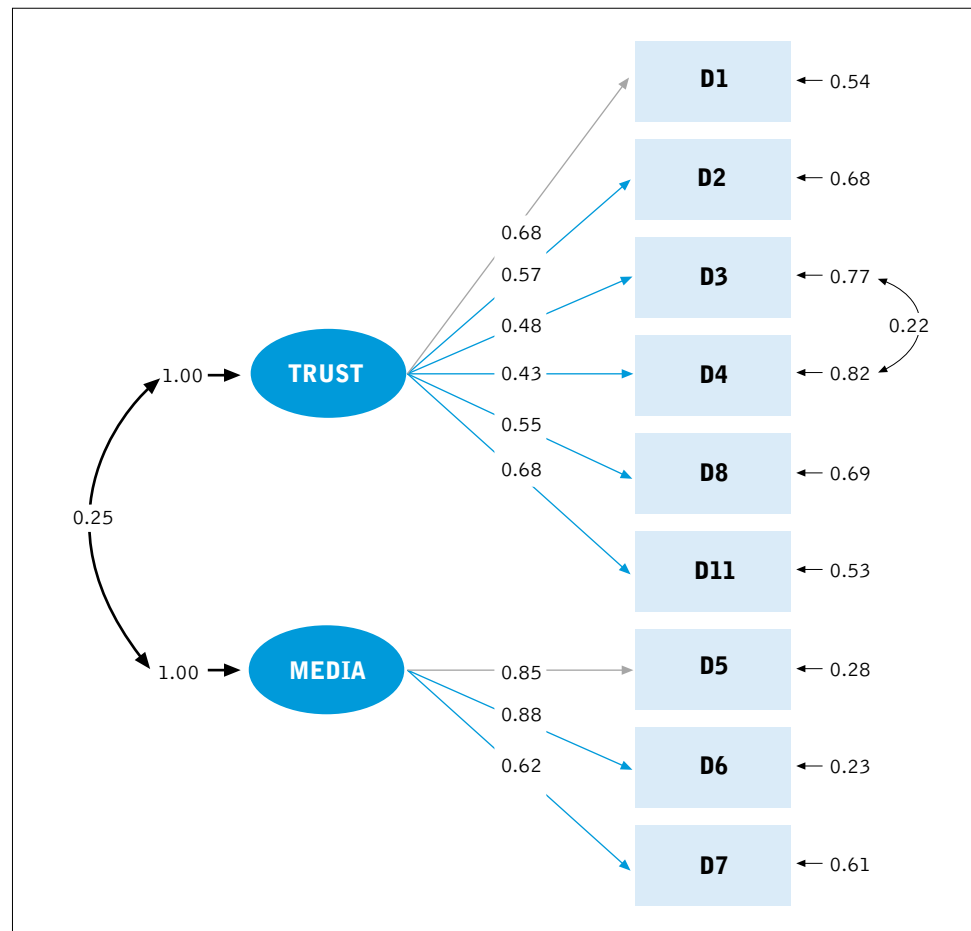
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .046, AGFI = .98, NNFI = .97, CFI = .98.

Table 8.8 Item parameters for TRUST

TRUST		Location	Tau 1	Tau 2	Tau 3
D1	Government	0.14	-2.22	0.04	2.18
D2	Local council	-0.10	-2.32	-0.04	2.35
D3	Courts	-0.50	-1.76	-0.08	1.84
D4	Police	-0.52	-1.48	-0.10	1.58
D8	Parties	0.90	-2.09	0.22	1.87
D11	Parliament	0.08	-1.82	-0.11	1.93

Table 8.9 Scale reliabilities for TRUST

Country	14-year-old Students	Upper Secondary Students
Australia	.79	
Belgium (French)	.76	
Bulgaria	.78	
Chile	.69	.74
Colombia	.77	.78
Cyprus	.70	.76
Czech Republic	.75	.76
Denmark	.76	.71
England	.78	
Estonia	.72	.78
Finland	.78	
Germany	.73	
Greece	.75	
Hong Kong (SAR)	.80	.73
Hungary	.84	
Israel		.71
Italy	.63	
Latvia	.74	.74
Lithuania	.78	
Norway	.76	.76
Poland	.75	.77
Portugal	.69	.69
Romania	.76	
Russian Federation	.72	.72
Slovak Republic	.73	
Slovenia	.76	.75
Sweden	.79	.83
Switzerland	.71	.73
United States	.80	
International Sample	**.77**	**.79**

Attitudes toward the Nation

Initially, all 12 items measuring students' attitudes toward their nation had been assumed to be uni-dimensional, but the model fit for the corresponding one-factor solution was rather unsatisfactory (RMSEA = .088, AGFI = .90, NNFI = .80, CFI = .84). After the preliminary analyses, four items were dropped, and the remaining eight items were assigned to the following dimensions (see Table 8.10):

• Protective feelings toward one's nation (PROTC), relating to those items indicating perceived threats and the need to defend the country against other countries.

• Positive attitudes toward one's nation (PATRI), relating to items indicating the importance of national symbols and emotional affection towards the country.

The two-factor solution (see Figure 8.4) had a close fit to the sample data, and the model fit was acceptable for all countries. The correlation of .58 between the two latent factors was a clear indication that a one-dimensional model was not appropriate. Although the two-dimensional solution was confirmed, the item reliabilities for items loading on the first factor were very low. Only items loading on the second factor (PATRI) were retained for scaling and inclusion in the international report.

Table 8.10 Items on national identity

PROTC: Protective Feelings toward One's Nation	
E1	To help protect jobs in this country [name of country] we should buy products made in this country [name of country].
E2	We should keep [prevent] other countries from trying to influence political decisions in this country [name of country].
E4	We should always be alert and stop threats from other countries to this country's [name of country] political independence.
E12	We should stop outsiders from influencing this country's [name of country] traditions and culture.
PATRI: Positive Attitudes toward One's Nation	
E3	The flag of this country [name of country] is important to me.
E7	I have great love for this country [name of country].
E9	This country [name of country] should be proud of what it has achieved.
E11*	I would prefer to live permanently in another country.

Notes: Categories—strongly disagree, disagree, agree, strongly agree. *Reversed item.

Figure 8.4 Two-factor solution for national identity items
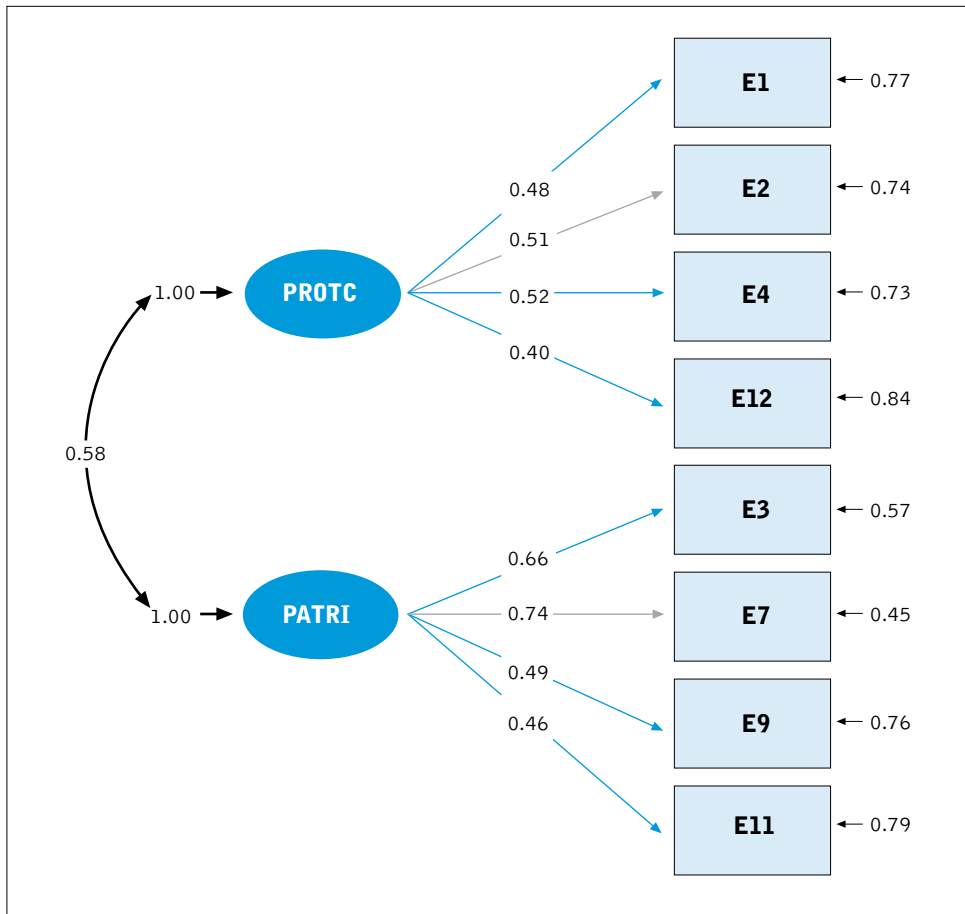
Notes: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA =. 044, AGFI = .98, NNFI = .95, CFI = .96.

Table 8.11 shows the item parameters for PATRI. Item E11 (preference to live abroad) is a reversed item and its location parameter is higher than for the other three items. The item that students were most likely to endorse was E7 (love for the country).

Table 8.12 shows the scale reliabilities for PATRI. The internal consistency of around .70 was satisfactory for both populations. The low reliability for Hong Kong (SAR) may have been a result of students having a different perception of national identity due to Hong Kong's past as a former British Crown colony and its current status as a special administrative region of China.

Table 8.11 Item parameters for PATRI

PATRI		Location	Tau 1	Tau 2	Tau 3
E3	Flag important	-0.03	-0.82	-0.48	1.30
E7	Love for this country	-0.22	-0.72	-0.72	1.44
E9	Country should be proud	-0.08	-1.10	-0.92	2.02
E11*	Prefer to live abroad	0.34	-0.76	-0.54	1.29

Note: *Reversed item.

Table 8.12 Scale reliabilities for PATRI

Country	14-year-old students	Upper secondary students
Australia	.69	
Belgium (French)	.68	
Bulgaria	.64	
Chile	.61	.65
Colombia	.59	.59
Cyprus	.54	.61
Czech Republic	.66	.67
Denmark	.65	.67
England	.70	
Estonia	.65	.68
Finland	.73	
Germany	.77	
Greece	.67	
Hong Kong (SAR)	.47	.52
Hungary	.63	
Israel		.80
Italy	.65	
Latvia	.67	.72
Lithuania	.65	
Norway	.70	.71
Poland	.67	.64
Portugal	.56	.61
Romania	.55	
Russian Federation	.57	.62
Slovak Republic	.65	
Slovenia	.64	.64
Sweden	.73	.70
Switzerland	.72	.81
United States	.68	
International Sample	**.69**	**.70**

Attitudes toward Women, Minorities, and Anti-Democratic Groups

Three dimensions were assumed for items measuring attitudes toward social groups: desired rights or opportunities for women (WOMRT), minorities (MINOR), and anti-democratic groups (ADGR). Table 8.13 lists the items in this section for each of these dimensions.

Figure 8.5 shows the model with a three-dimensional factor structure (WOMRT, MINOR, ADGR). The structure had a satisfactory model fit for the pooled international sample and for all country sub-samples. Estimates of the correlation between the three latent factors showed that WOMRT and MINOR were positively correlated whereas "intolerance for anti-democratic groups" had negative, but considerably lower, correlations with both WOMRT and MINOR. Only WOMRT was retained for scaling and reporting in the first international report.

Table 8.13 Items on desired opportunities for women, minorities, and anti-democratic groups

WOMRT: Attitudes toward Women's Political and Economic Rights	
G1	Women should run for public office [a seat in the legislature] and take part in the government just as men do.
G4	Women should have the same rights as men in every way.
G6*	Women should stay out of politics
G9*	When jobs are scarce, men [should] have more right to a job than women.
G11	Men and women should get equal pay when they are in the same jobs [occupations].
G13*	Men are better qualified to be political leaders than women.
MINOR: Attitudes toward Opportunities for Minorities	
G2	All ethnic [racial or national] groups should have equal chances to get a good education in this country.
G5	All ethnic [racial or national] groups should have equal chances to get good jobs in this country.
G8	Schools should teach students to respect members of all ethnic [racial or national] groups.
G12	Members of all ethnic [racial or national] groups should be encouraged to run in elections for political office.
ADGR: Attitudes toward Political Rights for Anti-Democratic Groups	
G3	Members of anti-democratic groups [groups that are against democracy] should be prohibited from hosting a television show talking about these [their] ideas.
G7	Members of anti-democratic groups [groups that are against democracy] should be prohibited from organizing peaceful [non-violent] demonstrations or rallies.
G10	Members of anti-democratic groups [groups that are against democracy] should be prohibited from running in an election for political office.
G14	Members of anti-democratic groups [groups that are against democracy] should be prohibited from making public speeches about these [their] ideas.

Notes: Categories-strongly disagree, disagree, agree, strongly agree. *Reversed items.

Figure 8.5 Three-factor solution for opportunity items
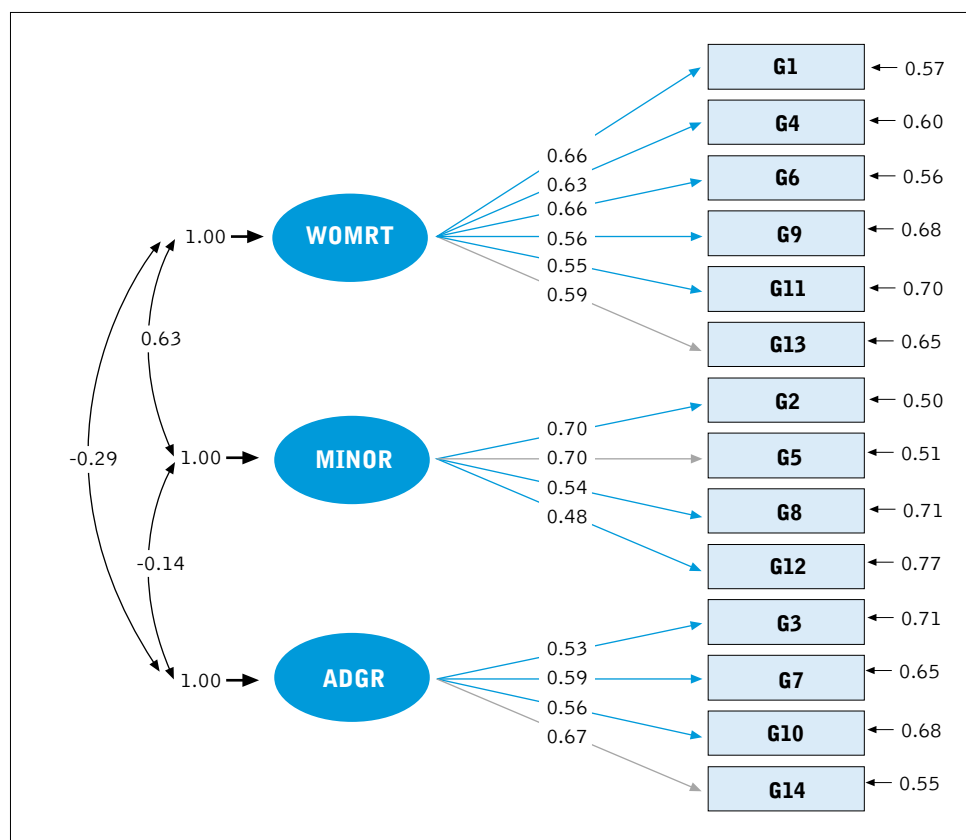
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .052, AFGI = .96, NNFI = .93, CFI = .94.

Table 8.14 shows the item parameters for WOMRT. The reversed item G13 (men are better qualified to be political leaders) was least likely to be endorsed, whereas G11 (equal pay for the same job) had, on average, the highest probability of endorsement. Scale reliabilities as shown in Table 8.15 were satisfactory across countries in both populations and ranged between .62 and .84.

Table 8.14 Item parameters for WOMRT

WOMRT		Location	Tau 1	Tau 2	Tau 3
G1	Run for office	0.07	-0.56	-1.26	1.82
G4	Have the same rights	-0.44	-0.51	-0.79	1.30
G6*	Stay out of politics	-0.08	-0.50	-0.69	1.19
G9*	Men more right to a job	0.36	-0.95	-0.39	1.35
G11	Get equal pay	-0.48	-0.56	-0.77	1.33
G13*	Men better qualified	0.57	-1.15	-0.22	1.37

Note: *Reversed items.

Table 8.15 Scale reliabilities for WOMRT

Country	14-year-old Students	Upper Secondary Students
Australia	.81	
Belgium (French)	.74	
Bulgaria	.69	
Chile	.62	.68
Colombia	.63	.68
Cyprus	.79	.83
Czech Republic	.77	.78
Denmark	.83	.79
England	.82	
Estonia	.75	.78
Finland	.84	
Germany	.84	
Greece	.81	
Hong Kong (SAR)	.79	.81
Hungary	.80	
Israel		.81
Italy	.80	
Latvia	.70	.75
Lithuania	.72	
Norway	.83	.83
Poland	.77	.79
Portugal	.71	.73
Romania	.72	
Russian Federation	.67	.73
Slovak Republic	.66	
Slovenia	.78	.80
Sweden	.76	.81
Switzerland	.81	.81
United States	.82	
International Sample	**.78**	**.80**

Attitudes toward Immigration

The student questionnaire included eight items measuring attitudes toward immigration that were assumed to be uni-dimensional. Table 8.16 shows the item wording and indicates the two items that had to be reversed for scaling.

The assumed one-factor solution for all these items was not entirely supported by the results of the CFA. The resulting model fit was rather poor and, in many countries, the RMSEA fit index was above .08. Item H7 was discarded because of its low item reliability. This was probably due to the wording, which did not necessarily reflect favorable or unfavorable attitudes toward immigrants and so might have been viewed as a somewhat factual statement.

The model fit for the one-factor solution without Item H7 was further improved by introducing a correlation between the error terms for Items H1 and H4. Both items refer to the cultural integration of immigrants. To illustrate the effect of wording on the item responses, a correlation between the errors for these items was estimated in the final model. The resulting model had a close fit (see Figure 8.6).

Table 8.16 Items on immigration

	IMMIG: Positive Attitudes toward Immigrants
H1	Immigrants should have the opportunity [option] to keep [continue speaking] their own language.
H2	Immigrants' children should have the same opportunities for education that other children in the country have.
H3	Immigrants who live in a country for several years should have the opportunity to vote in elections.
H4	Immigrants should have the opportunity [option] to keep [continue] their own customs and lifestyle.
H5	Immigrants should have all the same rights that everyone else in a country has.
H6*	Immigrants should be forbidden to engage in political activity.
H7*	Having many immigrants makes it difficult for a country to be united and patriotic.
H8	All countries should accept refugees who are trying to escape from wars or political persecution in other countries.

Notes: Categories—strongly disagree, disagree, agree, strongly agree. *Reversed items.

Figure 8.6 One-factor solution for immigrant items
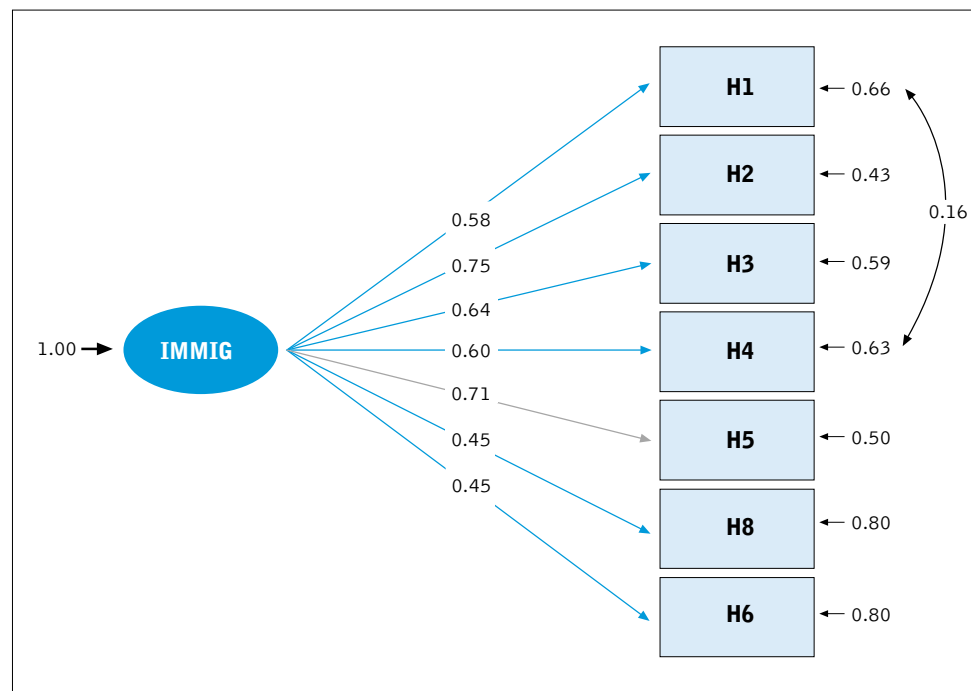
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .046, AGFI = .98, NNFI = .98, CFI = .99.

Estimating an IRT partial credit model for the remaining seven items showed that Items H6 and H8 had an unsatisfactory item fit. These items also had very low item reliabilities in the CFA. Only those five items with satisfactory scaling properties were retained for further scaling and reporting.

Table 8.17 shows the item parameters for the resulting scale on positive attitudes toward immigrants. Item H3 (same education for immigrant children) was the item that students were most likely to agree with whereas Item H1 (continue to speak their language) was the one they were least likely to endorse.

Scale reliabilities for both populations were satisfactory on the national and international levels and ranged between .68 and .90 (see Table 8.18).

Table 8.17 Item parameters for IMMIG

Item		Location	Tau 1	Tau 2	Tau 3
H1	Keep their language	0.38	-1.35	-1.03	2.38
H2	Same education	-0.59	-1.08	-1.20	2.28
H3	Vote in elections	0.23	-1.62	-0.78	2.40
H4	Keep customs and lifestyle	0.09	-1.49	-0.91	2.40
H5	Have the same rights	-0.11	-1.58	-0.69	2.27

Table 8.18 Scale reliabilities for IMMIG

Country	14-year-old Students	Upper Secondary Students
Australia	.88	
Belgium (French)	.81	
Bulgaria	.79	
Chile	.68	.72
Colombia	.74	.77
Cyprus	.73	.80
Czech Republic	.80	.75
Denmark	.85	.83
England	.90	
Estonia	.75	.81
Finland	.88	
Germany	.89	
Greece	.75	
Hong Kong (SAR)	.84	.84
Hungary	.79	
Israel		.58
Italy	.79	
Latvia	.74	.78
Lithuania	.77	
Norway	.88	.89

table contd. on next page

Table 8.18 contd. from previous page

Country	14-year-old Students	Upper Secondary Students
Poland	.80	.82
Portugal	.67	.70
Romania	.67	
Russian Federation	.73	.80
Slovak Republic	.76	
Slovenia	.78	.74
Sweden	.90	.89
Switzerland	.86	.83
United States	.85	
International Sample	**.82**	**81**

School Participation

For the items on school participation, a single dimension was assumed. Table 8.19 shows the wording of these items.

The one-dimensional solution, however, was not confirmed by the CFA. The model fit was poor (RMSEA = .117, AGFI = .89, NNFI = .77, CFI = .84), and in most countries the model fit was unsatisfactory. An alternative, two-dimensional model was estimated with the following two dimensions:

- General confidence in school participation (CONFS), relating to items indicating whether students think that student activities in school may be effective.

- Self-confidence in school participation (SCON), relating to items indicating whether students themselves are willing to act.

Items J6 and J7 related to SCON, and Items J1, J2, J3, and J5 related to CONFS. Item J4 loaded on both factors because it related to the self-concept of a student and the concept of what students can achieve through student organization.

Figure 8.7 shows the two-factor solution for these items. The model has a very good fit for the international sample; in most countries, the RMSEA was below .05. The two-factor structure for this set of items was clearly supported by the data.

Table 8.19 Items on school participation

	CONFS: Confidence in Participation at School
J1	Electing student representatives to suggest changes in how the school is run [how to solve school problems] makes schools better.
J2	Lots of positive changes happen in this school when students work together.
J3	Organizing groups of students to state their opinions could help solve problems in this school.
J4*	If members of my class felt they were unfairly treated, I would be willing to go with them to speak to the teacher.
J5	Students acting together [in groups] can have more influence on what happens in this school than students acting alone [by themselves].
J6*	I am interested in participating in discussions about school problems.
J7*	When school problems are being discussed I usually have something to say.

Note: Categories—strongly disagree, disagree, agree, strongly agree. *Items not included in the final scaling.

Figure 8.7 Two-factor solution for items on school participation
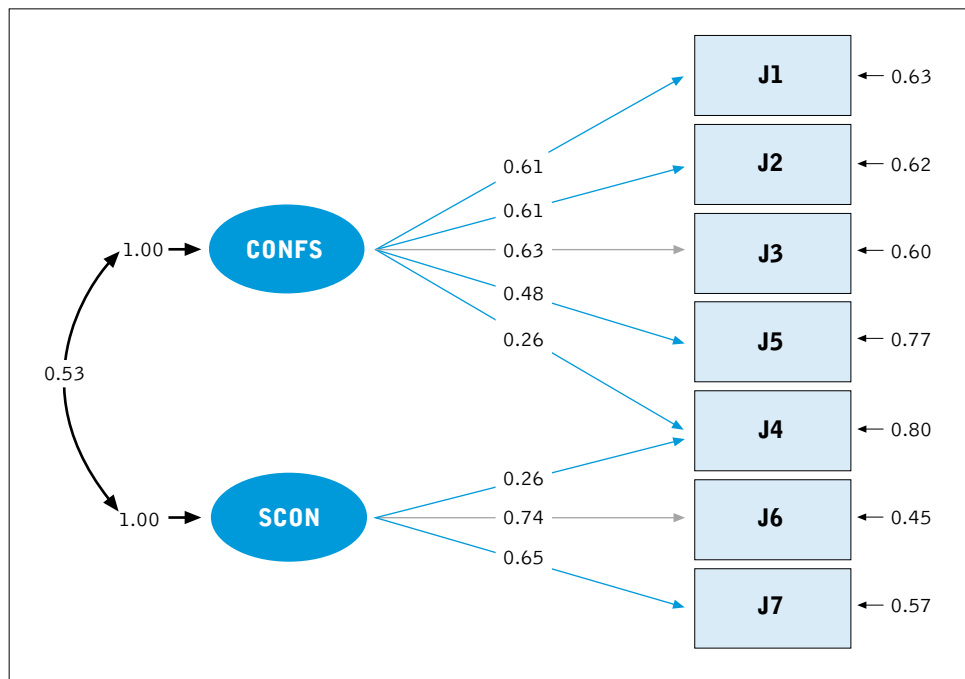
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .032, AGFI = .99, NNFI = .98, CFI = .99.

Because the second factor consisted of three items only, with J4 loading also on both factors, only the confidence in participation in school (CONFS) scale was selected. Item J4 was discarded due to its double loading on both factors.

Table 8.20 shows the item parameters for these four items. The location parameters are generally very close to each other, showing that they had similar probabilities of being endorsed by the 14-year-old students.

Scale reliabilities for both populations were moderate to good in most countries and ranged from .56 to .80 (see Table 8.21).

Table 8.20 Item parameters for CONFS

Item		Location	Tau 1	Tau 2	Tau 3
J1	Electing representatives	0.23	-1.03	-1.09	2.12
J2	Students working together	-0.12	-1.33	-0.93	2.26
J3	Organizing groups	-0.03	-1.46	-0.98	2.44
J5	Students acting together	-0.09	-1.12	-0.90	2.02

Table 8.21 Scale reliabilities for CONFS

Country	14-year-old Students	Upper Secondary Students
Australia	.76	
Belgium (French)	.67	
Bulgaria	.78	
Chile	.57	.64
Colombia	.59	.61
Cyprus	.67	.77
Czech Republic	.57	.56
Denmark	.72	.76
England	.75	
Estonia	.69	.80
Finland	.72	
Germany	.68	
Greece	.69	
Hong Kong (SAR)	.75	.77
Hungary	.64	
Israel		.80
Italy	.61	
Latvia	.65	.67
Lithuania	.73	
Norway	.75	.78
Poland	.75	.77
Portugal	.69	.70
Romania	.62	
Russian Federation	.56	.65
Slovak Republic	.60	
Slovenia	.65	.64
Sweden	.75	.77
Switzerland	.66	.69
United States	.79	
International Sample	**.69**	**.73**

Expected Political Participation

The student questionnaire included 12 items asking about the students' expected political participation. Five of these items related to conventional participation (CONV), four to unconventional or social movement-related participation (UNCONV), and three to protest activities (PROTE). Table 8.22 lists these items and their corresponding theoretical dimensions.

A CFA showed that this initial three-factor solution had an unsatisfactory model fit and was not supported by the data. Discarding five of these items resulted in a two-factor solution (see Figure 8.8) with political activities (POLAT) and protest activities (PROTE), which had a good model fit for the international sample and the country sub-samples. Only POLAT was retained for scaling and reporting.

Table 8.22 Items on political participation and expected dimensions

Item		CONV	UNCONV	PROTE
M1*	Vote in national elections	X		
M2*	Get information about candidates before voting in an election	X		
M3	Join a political party	X		
M4	Write letters to a newspaper about social or political concerns	X		
M5	Be a candidate for a local or city office	X		
M6*	Volunteer time to help [benefit] [poor or elderly] people in the community		X	
M7*	Collect money for a social cause		X	
M8*	Collect signatures for a petition		X	
M9*	Participate in a non-violent [peaceful] protest march or rally		X	
M10*	Spray-paint protest slogans on walls			X
M11*	Block traffic as a form of protest			X
M12*	Occupy public buildings as a form of protest			X

Notes: Categories—I will certainly not do this; I will probably not do this; I will probably do this; I will certainly do this.

*Items not included in the final scaling.

Figure 8.8 Two-factor solution for expected participation
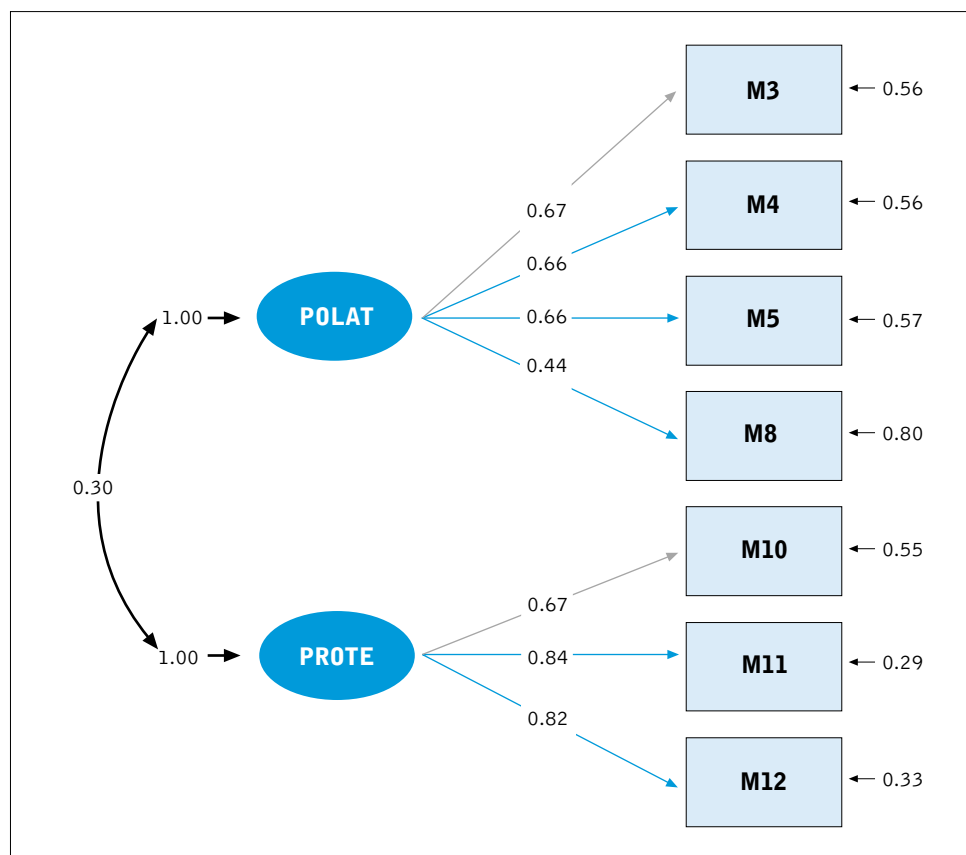
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .039, AGFI = .99, NNFI = .98, CFI = .99.

The IRT scaling analysis showed that although all of these items had satisfactory fit statistics for the pooled international sample, Item M8 had unsatisfactory scaling properties in all but two countries and so was discarded.

Table 8.23 shows the item parameters for the three items measuring political activities (POLAT). The location parameters are very close to each other, indicating a similar level of endorsement. The scale reliabilities were satisfactory in a majority of countries for both populations (14-year-old students and upper secondary students) (see Table 8.24).

Table 8.23 Item parameters for POLAT

Item	Location	Location	Tau 1	Tau 2	Tau 3
M3	Join political party	0.00	-1.89	0.32	1.57
M4	Write letters	-0.08	-2.05	0.37	1.68
M5	Be a candidate	0.08	-1.77	0.29	1.48

Table 8.24 Scale reliabilities for POLAT

Country	14-year-old Students	Upper Secondary Students
Australia	.78	
Belgium (French)	.72	
Bulgaria	.76	
Chile	.69	.75
Colombia	.65	.70
Cyprus	.68	.72
Czech Republic	.71	.75
Denmark	.76	.76
England	.77	
Estonia	.75	.81
Finland	.74	
Germany	.72	
Greece	.71	
Hong Kong (SAR)	.82	.80
Hungary	.65	
Israel		.77
Italy	.72	
Latvia	.72	.73
Lithuania	.79	
Norway	.72	.79
Poland	.73	.78
Portugal	.62	.67
Romania	.68	
Russian Federation	.72	.78
Slovak Republic	.72	
Slovenia	.69	.73
Sweden	.76	.76
Switzerland	.77	.75
United States	.74	
International Sample	**.73**	**.75**

Teaching Styles

The student questionnaire included 12 items on teaching style that were designed to measure two different dimensions: open climate for classroom discussion (CCLIM), reflecting a teaching that encourages free and controversial discussions and respects diverging opinions, and lecturing style (LECTR), indicating a teaching style that places emphasis on lecturing and testing achievement. Table 8.25 shows the item wording and assignment to the latent constructs.

Table 8.25 Items on teaching style

CCLIM: Open Climate for Classroom Discussion	
N1	Students feel free to disagree openly with their teachers about political and social issues during class.
N2	Students are encouraged to make up their own minds about issues.
N3	Teachers respect our opinions and encourage us to express them during class.
N5	Students feel free to express opinions in class even when their opinions are different from most of the other students.
N7	Teachers encourage us to discuss political or social issues about which people have different opinions.
N8	Teachers present several sides of [positions on] an issue when explaining it in class.
N9*	Students bring up current political events for discussion in class.
LECTR: Lecturing Style	
N4*	Teachers place great importance [stress, emphasis] on learning facts or dates when presenting history or political events.
N6*	Teachers require students to memorize dates or definitions.
N10*	Memorizing dates and facts is the best way to get a good grade [mark] from teachers in these classes.
N11*	Teachers lecture and the students take notes.
N12*	Students work on material from the textbook.

Notes: Categories—strongly disagree, disagree, agree, strongly agree. *Items not included in the final scaling.

The initial two-factor solution for this set of items had an acceptable fit of RMSEA = .067. Inspection of squared correlations showed that the item reliabilities for Items N11 and N12 were below .10. An alternative and slightly modified model was specified without items N11 and N12 (see Figure 8.9). Here, Item 4 (emphasis on dates and facts) also loaded on CCLIM, and correlated error terms for Items N1-2 and N7-9 were included. Items N1 and N2 have a very similar meaning (disagreement, own minds), and Items N7 and N9 both emphasize discussion of issues, similarities that are reflected in the common residual variance terms for these pairs of items. Figure 8.9 shows the final two-factor solution, which had a satisfactory model fit. Given that the lecturing style construct was measured with only three items, one of them clearly related to both factors, a decision was made to retain only open climate for classroom discussion (CCLIM) with the six items that had the highest factor loadings.

Table 8.26 shows the IRT item parameter for the teaching style scale. Item N5 (students feel free to express their opinion) was, on average, the item students were most likely to agree with, whereas Item N7 (teachers encourage us to discuss political and social issues) was the one students found hardest one to endorse. Table 8.27 shows that scale reliabilities were highly satisfactory for both international samples (.77 for the 14-year-old students and .79 for the upper secondary students); the range was between .68 and .83 across the countries.

Figure 8.9 Two-factor solution for classroom items
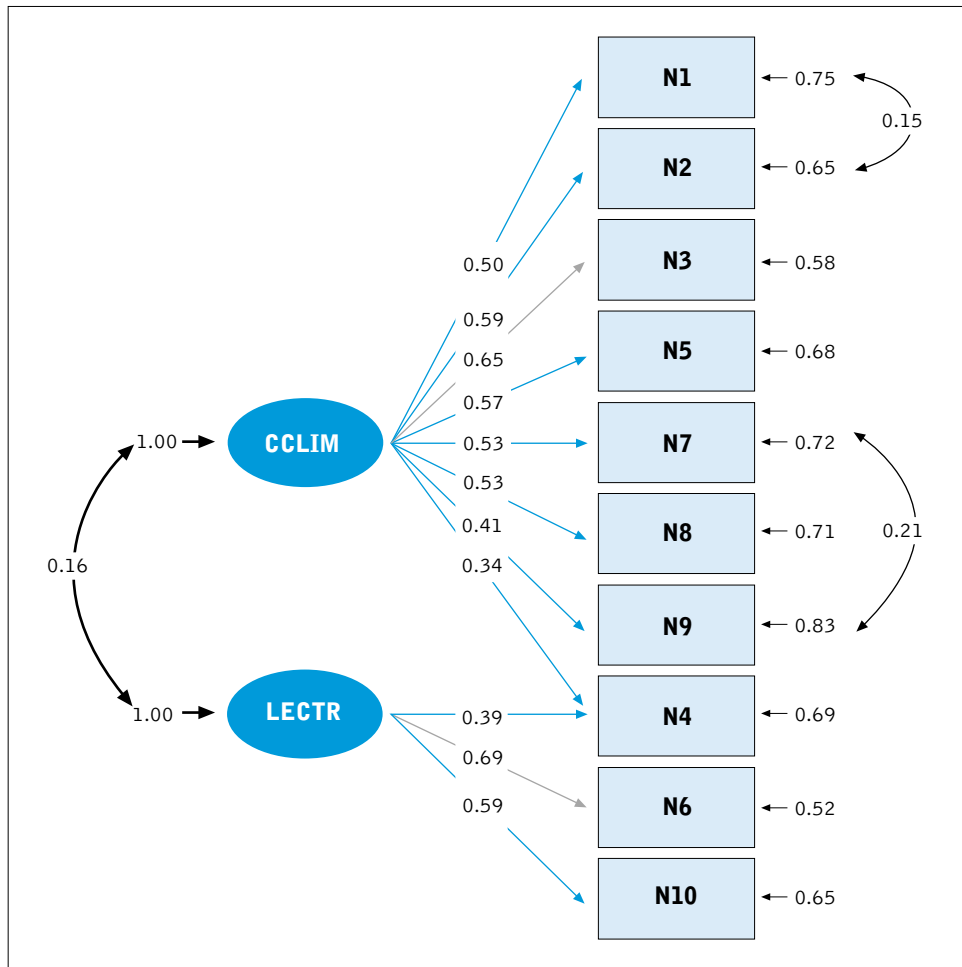
Note: Standardized maximum likelihood estimates for international sample of 14-year-olds. RMSEA = .042, AGFI = .99, NNFI = .96, CFI = .97.

Table 8.26 Item parameters for CCLIM

Item		Location	Tau 1	Tau 2	Tau 3
N1	Students disagree openly	0.13	-0.95	-0.23	1.18
N2	Students are encouraged	-0.24	-1.04	-0.18	1.23
N3	Teachers respect opinions	-0.17	-0.89	-0.19	1.08
N5	Students feel free	-0.35	-1.13	-0.17	1.31
N7	Teachers encourage	0.65	-1.25	-0.30	1.54
N8	Teacher presents several sides	-0.02	-1.29	-0.35	1.64

Table 8.27 Scale reliabilities for CCLIM

Country	14-year-old Students	Upper Secondary Students
Australia	.81	
Belgium (French)	.78	
Bulgaria	.72	
Chile	.73	.78
Colombia	.69	.76
Cyprus	.68	.76
Czech Republic	.78	.75
Denmark	.82	.77
England	.80	
Estonia	.75	.79
Finland	.80	
Germany	.78	
Greece	.71	
Hong Kong (SAR)	.79	.81
Hungary	.71	
Israel		.78
Italy	.79	
Latvia	.72	.78
Lithuania	.70	
Norway	.79	.80
Poland	.82	.82
Portugal	.73	.78
Romania	.69	
Russian Federation	.75	.78
Slovak Republic	.75	
Slovenia	.73	.72
Sweden	.79	.83
Switzerland	.79	.82
United States	.82	
International Sample	**.77**	**.79**

THE PROCESS OF ITEM ADJUDICATION

Item parameters and fit statistics for each country's data sets were estimated for the item adjudication. The adjudication was based on the resulting item statistics. The following conditions were set as an indication of problematic scaling properties that could lead to the discarding of items from scaling:

- The weighted mean square residual of an item in any country is an indication of item misfit.

- Items should have a satisfactory item-total correlation.

- No *category disordering* should be observed, that is, means of latent estimates should not be out of sequence across categories.

If an item had poor scaling properties in more than 30 percent of all countries, this item was discarded from further scaling for all countries.

Table 8.28 shows the items that were discarded for individual countries due to unsatisfactory scaling properties in the survey of the 14-year-old students. Table 8.29 shows the results of the item adjudication for the upper secondary students.

Table 8.28 Items excluded from scaling after item adjudication (14-year-old students)

Country	CTCON	CTSOC	GOVEC	GOVSO	TRUST	PATRI	WOMRT	IMMIG	CONFS	POLAT	CCLIM
Australia											
Belgium (French)							G9				
Bulgaria			C7	C12		E11	G13				
Chile							G6				
Colombia							G13				
Cyprus											
Czech Republic		B5									
Denmark	B10										N8
England											
Estonia											
Finland				C12							
Germany					D2		G9				
Greece				C12							
Hong Kong (SAR)		B5		C12	D4, D8						
Hungary											N1
Italy											
Latvia							G9				
Lithuania											
Norway				C12							
Poland											
Portugal											
Romania											
Russian Federation											
Slovak Republic							G9				
Slovenia		B5									
Sweden						E11	G9				
Switzerland											
United States		B5				E11	G9				

Table 8.29 Items excluded from scaling after item adjudication (upper secondary students)

Country	CTCON	CTSOC	GOVEC	GOVSO	TRUST	PATRI	WOMRT	IMMIG	CONFS	POLAT	CCLIM
Chile											
Colombia							G13				
Cyprus											
Czech Republic		B5									
Denmark											
Estonia											
Hong Kong (SAR)				C12							
Israel						E11		H3			
Latvia				C12			G9				
Norway											
Poland											
Portugal											
Russian Federation							G9				
Slovenia		B5									
Sweden						E11	G9				
Switzerland							G9				

STUDENT SCORES

After the item adjudication, international item parameters were estimated for each scale based on a calibration sample of 500 randomly selected students per country (14,000 students from the population of 14-year-old students). These item parameters were used to compute ML estimates for each sub-sample. On the basis of the results of the item adjudication, some items were excluded from scaling.

The person parameters (logits) were transformed to the international metric with an international mean of 10 and a standard deviation of 2. The transformation was achieved by applying the formula

$$\theta'_n = 10 + 2\left(\frac{\theta_n - \bar{\theta}}{\sigma_\theta}\right)$$

where θ'_n are the student scores in the international metric, θ_n the original logit scores, $\bar{\theta}$ is the international mean of student logit scores with equally weighted country sub-samples, and σ_θ is the corresponding international standard deviation.

For the upper secondary students, the same item parameters derived from the population of 14-year-old students were used to scale the data. The resulting student scores were transformed to the same international metric. Table 8.30 shows the means and standard deviations of student logit scores (14-year-olds tested in 1999) used for the transformation into the international metric.

Table 8.30 International means and standard deviations for IRT scales (logits) for 14-year-old students tested in 1999

Scale	Mean	Standard Deviation
CTCON	0.35	1.09
CTSOC	1.13	1.37
GOVSOC	1.00	1.10
GOVEC	1.54	1.24
TRUST	0.05	1.39
PATRI	1.32	1.44
WOMRT	1.57	1.54
IMMIG	1.28	1.86
CONFS	1.42	1.64
POLAT	-1.43	1.79
CCLIM	0.76	1.32

SUMMARY

Eleven scales were constructed for the Likert-type items on the students' concepts, attitudes, and actions. Generally, the expected dimensional structures were confirmed in the CFA. In some cases, a re-specification was deemed appropriate and some of the initially hypothesized dimensions were not supported by the data. Despite the relative shortness of most scales, their reliabilities across countries were mostly satisfactory.

The IRT partial credit model was used for the scaling of Likert-type items. Use of this methodology allowed a better assessment of item fit and an elegant way of dealing with the problem of missing responses. ML estimates were computed and then transformed into international scores, giving international averages of 10 and standard deviations of 2.

REFERENCES

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, volume II: Implementation and analysis—primary and middle school years* (pp. 111–145). Chestnut Hill, MA: Boston College.

Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries.* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Andersen, E. B. (1997). The rating scale model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 67–84). New York/Berlin/Heidelberg: Springer.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & S. J. Long (Eds.), *Testing structural equation models* (pp. 36–162). Newbury Park/London: Sage Publications.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & S. J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park/London: Sage Publications.

Gustafsson, J. E & Stahl, P. A. (2000). *STREAMS user's guide, Version 2.5 for Windows*. Mölndal, Sweden: MultivariateWare.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review, 11*(4), 234–373.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & S. J. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park/London: Sage Publications.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

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

Mislevy, R. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*(2), 177–196.

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

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

Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software* (computer program). Camberwell, Vic.: Australian Council for Educational Research.

Chapter 9:

MAPPING STUDENT SCORES TO ITEM RESPONSES

Wolfram Schulz

OVERVIEW

In the IEA Civic Education Study (CivEd), categorical items from the student questionnaire were scaled using item response theory (IRT) modeling. Maximum likelihood (ML) estimates (logits) for the latent dimension were transformed to a scale with an international average of 10 and a standard deviation of 2. These scores can be interpreted by comparing individual scores or group average scores to the international mean, but they do not reveal anything about the actual item responses. Nor is it possible to determine from scale score values the extent to which students endorsed the items used for the measurement of the latent variable.

This chapter describes how the properties of the IRT scaling for categorical items were used to describe item characteristics and to map scale scores against expected item responses. The scale *positive attitudes toward women's political and economic rights* (see details on scaling in Chapter 8) and particularly the (negatively phrased) Item G9 (When jobs are scarce, men [should] have more rights to a job than women) are chosen to illustrate methodology and procedure (see details on scaling in Chapter 8). As the scale was designed to measure positive attitudes, and the example item was negatively phrased, it had to be reversed so that its four categories were coded 0 (strongly agree), 1 (agree), 2 (disagree), and 3 (strongly disagree).

PROPERTIES OF THE PARTIAL CREDIT MODEL

The partial credit model (see Masters & Wright, 1997; Wright & Masters, 1982) used in the IEA Civic Education Study is an extension of the Rasch one-parameter model for dichotomous items (Rasch, 1960) and can be written as

$$P_{x_i}(\theta) = \frac{\exp\left(\sum_{j=0}^{x} \theta_n - \delta_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{k} \theta_n - \delta_{ij}\right)} \quad, x = 0, 1, 2, ..., m_i$$

where $P_{x_i}(\theta)$ denotes the probability of person n to score x on item i, δ_{ij} gives the location of step j for item i on the latent dimension, and θ_n denotes the location of person n on the latent dimension.[1]

---

1 Note that the form of the equation is slightly different from that given in Chapter 8, where an item parameter δ_i and step parameter τ_{ij} are used. The "delta" parameters for an item are equal to the sums of its location parameter and its step parameters, that is, $\delta_{ij} = \delta_i + \tau_{ij}$.

From the item parameters, it is possible to derive the probabilities for choosing each item response. Figure 9.1 shows the item characteristic curves for Item G9. Each of the four curves represents the probability of choosing a category depending on the person's score on the latent dimension θ. The four curves are not empirical curves but depend entirely on the estimated model. If the item fits the model, one can predict the probabilities to select a certain category for each scale score.

Figure 9.1 Category characteristic curves for Item G9
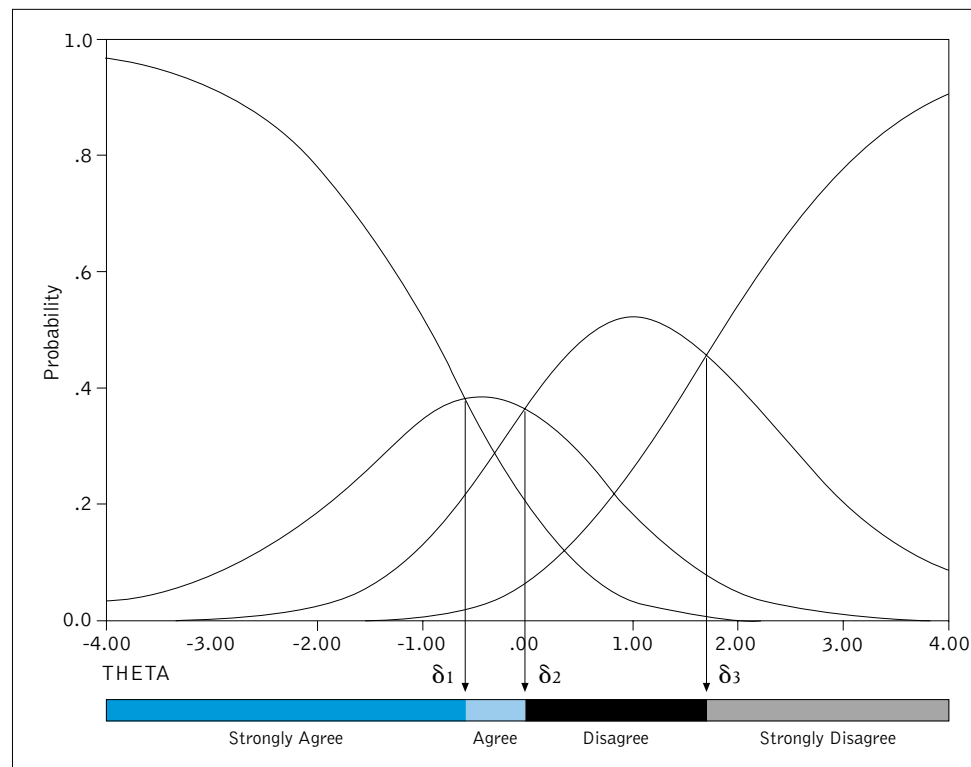
The four curves can be computed based on the parameters δ_1, δ_2, and δ_3. The δ parameters indicate points of intersection for each category curve on the latent dimension. δ_1 corresponds to the logit at the intersection between the lowest category 0 and the next category 1, δ_2 is determined by the intersection between scores 1 and 2, and δ_3 by the intersection between categories 2 and 3. This means that, for logit scores below δ_1, the lowest category (strongly agree) is the one most likely to be chosen, between δ_1 and δ_2 is the second category (agree) most likely to be chosen, and so on.

For some items, due to reversed δ parameters, it may happen that a category is less likely to be chosen than others at any point of the latent continuum. However, this does not mean that this category is not chosen; it may still have a reasonable probability of response for a range of scores.

A second possibility for describing the item characteristics as derived from the partial credit model is summing the probabilities for curves, that is, computing the odds of scoring higher than a particular category. The results for Item G6 are displayed in Figure 9.2. The three vertical lines denote the points on the latent continuum where it becomes more likely to score >0, >1, or >2. These locations Γ_k—so-called Thurstonian thresholds—can be obtained through an iterative procedure.

Figure 9.2 Summed category probabilities for Item G9
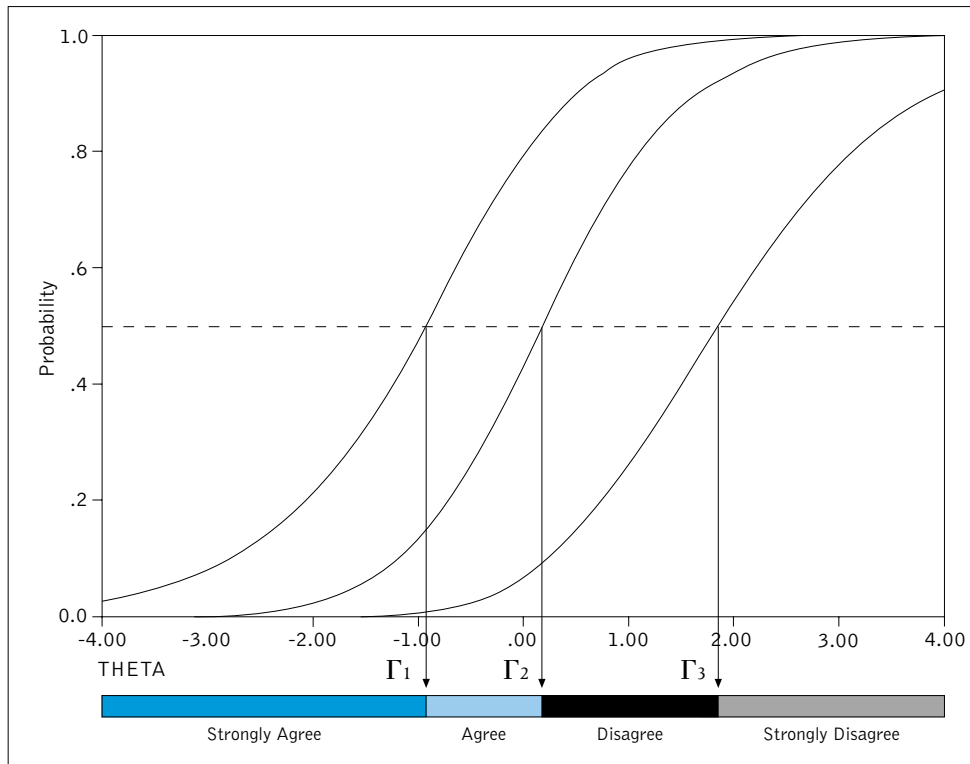
Summed probabilities are not identical with expected item scores. They have to be understood in terms of the odds of scoring *at least* a particular category. For example, Γ_1 is the point on the latent dimension where scoring *1 or more* becomes more likely than scoring the lowest category 0.

A third way of describing item properties using the partial credit model is the computation of expected item scores. The expected item scores can be calculated as the sum of the products of each category probability P_{ijh} with its corresponding category score h:

$$E_x = \sum_{h=0}^{m_i} h P_{ijh}$$

Figure 9.3 shows the expected item scores as a function of the latent variable θ. The lowest category 0 (strongly agree) may be defined as the range of attitudes for which the expected score is between 0 and 0.5. Likewise, the probability of choosing the second category (agree) would be highest for the range of score points with an expected scale score from 0.5 to 1.5, and so on.

Based on this information, it is possible to map the latent score against the expected item responses. In order to provide such a mapping, it is necessary to find those points on the latent variable at which $E_x = 0.5$, 1.5, etc. The thresholds ω_k that give the corresponding locations on θ can be estimated with an iterative procedure that calculates expected scores for each (decimal) point on the latent variable.

Figure 9.3 Expected item score for Item G9
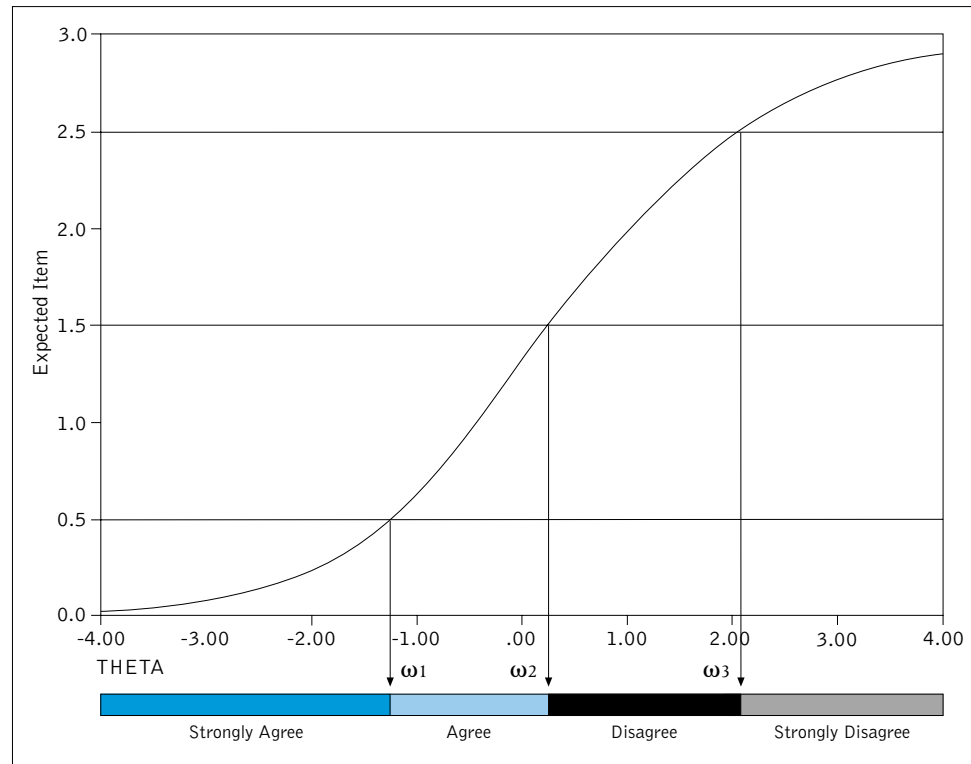
ITEM-BY-SCORE MAPS

Category characteristic curves, summed category probabilities, and *expected item scores* are different but equally legitimate ways of describing the characteristics of the partial credit (or rating scale) model. However, using expected item scores was deemed the most appropriate way of describing the CivEd scales derived from Likert-type items and of communicating the mapping of scale scores to item categories to the reader.

After transforming the locations ω_k (which indicate the expected ranges for each category) to the international metric of the scale, bar charts can be drawn that indicate which item responses can be expected for each score on the international scale. In the IEA Civic Education Study, this approach was used to provide so-called *item-by-score maps* for Likert-type items. Figure 9.4 shows the item-by-score map for the scale on positive attitudes toward women's economic and political rights.

The vertical lines indicate for each of the scale scores which response a student is most likely to give. If, for example, a respondent has a score of 10 on the women's rights scale, he or she is likely to strongly agree with the second and fifth item but only to agree with the first item. On the other hand, he or she would probably disagree with the (inverted) fourth and sixth items and strongly disagree with the (inverted) third item.

This mapping also demonstrates that even with a scale score of 8 (that is, one standard deviation below the international mean), students can still be expected to agree with all three statements supporting women's rights and to disagree with the statement that women should stay out of politics. A look at the

international metric might suggest that students with this score will hold more negative attitudes toward women's rights.

Figure 9.4 Item-by-score map for women's economic and political rights

International Item Frequencies (percent)	strongly disagree	disagree	agree	strongly agree	Sum
...public office	4	7	48	41	100
...same rights	3	6	32	58	100
...politics	52	33	9	6	100
...job	40	35	17	8	100
...equal pay	3	6	33	58	100
...political leaders	36	36	19	9	100

Note: The bars indicate the expected response of an item for a given scale score on the horizontal axis. International item frequencies are based on all 28 equally weighted country data.

Source: *IEA Civic Education Study*, Standard Population of 14-year-olds tested in 1999.

SUMMARY

IRT scores cannot be interpreted with regard to their content unless they are mapped to the expected item responses. *Item-by-score maps* illustrate that for some scales even scores of one standard deviation below the international average may still indicate rather positive responses. This situation would not have been revealed by the international score values indicating the position relative to the international average. The item-by-score maps used in the reports of the IEA Civic Education Study enable readers of reports to interpret means and mean differences of IRT scale scores between countries or groups of students.

REFERENCES

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

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielsen & Lydiche.

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

Chapter 10:

REPORTING STUDENT AND TEACHER DATA[1]

Wolfram Schulz and Rainer Lehmann

OVERVIEW

This chapter describes how the international student and teacher data were reported. Test and questionnaire results were compared across countries, and country averages were compared with the international average. Reporting required an appropriate treatment of missing values, procedures to estimate the sampling variance, the computation of significance tests for mean differences, and a correction for multiple comparisons. The chapter also describes the analysis of the effects of background and school variables on civic knowledge and civic engagement, and the reporting of data from the teacher survey.

TREATMENT OF MISSING VALUES

In most types of survey research, large amounts of missing data often cause problems with data analysis. For the reporting of cognitive items used in the IEA Civic Education Study (CivEd), missing responses were generally treated as incorrect responses unless the respective item had not been administered or had been excluded from scaling. However, so-called "not reached items", that is, items at the end of the test that students had been unable to attempt due to lack of time, were treated as missing responses for the international item calibration (see Chapter 7 for details).

For background and Likert-type items on students' concepts, attitudes, and actions, higher amounts of missing values needed to be addressed. A variety of methods are available to deal with item non-response (see Winglee, Kalton, Rust, & Kaspryzk, 2001). Imputation methods can be considered, such as *EM algorithms* (using an iterative maximum likelihood procedure providing estimates of means and variance-covariance matrices based on all available data for each respondent); so, too, can *available case analysis* (using all available data from students for regression analysis in a pair-wise fashion), or *hot-deck imputation* (where missing values are actually substituted based on information available for classes of respondents). The use of imputation methods, however, is computationally extensive and would have required additional resources and time for data analysis and management that were not available prior to the publication of the two first international reports. Nonetheless, secondary analysis might reconsider imputation methods as ways of dealing with the considerable amount of missing data in some parts of the questionnaire data.

---

1 Parts of this chapter are based on Chapter 12 in the *TIMSS Technical Report* on the estimation of sampling variance (see Gonzalez & Foy, 2000) and Chapter 16 in the same volume on the reporting of student achievement (see Gonzalez & Gregory, 2000).

For the reporting of the non-cognitive components of the CivEd data, missing responses, not administered items, and "don't know" responses generally were not included in the calculation of percentages and means. In some cases, a different treatment for "don't know" responses might have been considered,[2] but for consistency reasons such responses were excluded from reporting. IRT scaling of the Likert-type items used in this study provided an elegant way of reducing the amount of missing information. This technique allowed us to obtain student scores for all students who had responded to at least one item in a scale (see Chapter 8 for details).

ESTIMATION OF SAMPLING VARIANCE

Student samples were obtained through two-stage cluster sampling procedures. During the first stage, schools were sampled from a sampling frame with a probability proportional to their size; during the second stage, intact classrooms were randomly sampled within schools (see Chapter 4 on the CivEd sampling design). Cluster sampling techniques permit an efficient and economic data collection, but given that the aforementioned samples were not simple random samples, the usual formula to obtain standard errors for population estimates was not appropriate.

Replication techniques provide tools to estimate the correct sampling variance on population estimates (Gonzalez & Foy, 2000; Wolter, 1985). For the IEA Civic Education Study, the jackknife repeated replication technique (JRR) was used to compute standard errors for population means, percentages, and regression coefficients.

The JRR method for stratified samples generally requires pairing sampled schools as the primary sampling units (PSUs) into pseudo-strata or *sampling zones*. The assignment of schools to sampling zones therefore needed to be consistent with the sampling frame from which they were sampled. In countries where explicit stratification had been used, sampling zones were constructed within explicit strata. In the case of odd numbers of schools within an explicit stratum or of no explicit stratification within the sampling frame, the remaining schools were randomly divided into two halves to form pseudo-schools. In Cyprus, where all schools were tested and two classrooms within each school had been sampled, schools were defined as sampling zones and classrooms as sampling units. Table 10.1 shows the number of sampling zones used in each participating country.

---

2 In the case of items on media use, for example, it could be assumed that students who selected the "don't know" category were not doing any such activities. Here, the percentages of students reporting to read newspapers or watch TV news more frequently that included students with "don't know" responses might have provided better population estimates.

Table 10.1 Range of sampling zones used in each country

Country	14-year-old Students	Upper Secondary Students
Australia	75	
Belgium (French)	56	
Bulgaria	74	
Chile	75	75
Colombia	73	
Cyprus	61	41
Czech Republic	75	75
Denmark	75	71
England	64	
Estonia	75	72
Finland	73	
Germany	75	
Greece	75	
Hong Kong (SAR)	75	
Hungary	73	
Israel		75
Italy	75	
Latvia	66	63
Lithuania	75	
Norway	75	62
Poland	75	75
Portugal	75	75
Romania	74	
Russian Federation	56	56
Slovak Republic	73	
Slovenia	75	73
Sweden	70	75
Switzerland	75	35
United States	40	

Within each of these sampling zones, one school was randomly assigned a value of 2 while the other school received a value of 0. *Jackknife replicates* were formed by giving one of the paired schools a contribution of zero and the other school a double contribution; all other schools outside the sampling zone had a contribution of one.

The *replicate weights* were then obtained by simply multiplying student weights with the jackknife indicators of each replicate. As a result, for each jackknife replicate a weight was added to the data file where, for one sampling zone at a time, one school received a double weight and one other school a zero weight. Table 10.2 illustrates this procedure with a simple example of 24 students from six different schools (A–F) paired into three sampling zones.

Table 10.2 Example for computation of replicate weights

ID	Student Weight	School	Sampling Zone	Jackknife Indicator	Replicate Weight 1	Replicate Weight 2	Replicate Weight 3
01	5.2	A	1	0	0.0	5.2	5.2
02	5.2	A	1	0	0.0	5.2	5.2
03	5.2	A	1	0	0.0	5.2	5.2
04	5.2	A	1	0	0.0	5.2	5.2
05	9.8	B	1	2	19.6	9.8	9.8
06	9.8	B	1	2	19.6	9.8	9.8
07	9.8	B	1	2	19.6	9.8	9.8
08	9.8	B	1	2	19.6	9.8	9.8
09	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	0.0	7.2
14	7.2	D	2	0	7.2	0.0	7.2
15	7.2	D	2	0	7.2	0.0	7.2
16	7.2	D	2	0	7.2	0.0	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	0.0
22	8.2	F	3	0	8.2	8.2	0.0
23	8.2	F	3	0	8.2	8.2	0.0
24	8.2	F	3	0	8.2	8.2	0.0

For each country sample, 75 replicate weights were computed regardless of the number of sampling zones, allowing for 150 schools per country. For countries where more than 150 schools had been sampled, schools were collapsed to form larger pseudo-schools in order to keep the total number to 75. In countries with fewer sampling zones, the remaining replicate weights were equal to the original sampling weight.

To compute the sampling variance for the statistic t, the statistic is estimated once for the original sample S and then for each of the jackknife replicates. The JRR variance is computed using the formula

$$Var_{jrr}(t) = \sum_{h=1}^{H} [t(J_h) - t(S)]^2$$

where H is the number of sampling zones, $t(S)$ the statistic t estimated for the population using the original sampling weights, and $t(J_h)$ the same statistic estimated using the weights for the hth jackknife replicate. The standard error for t is

$$\sigma(t) = \sqrt{Var_{jrr}(t)}$$

The computation of JRR variance can be obtained for any statistic. Standard statistical software generally does not include any procedures for replication techniques. For the IEA Civic Education Study, SPSS macros were used to estimate JRR variance for means and percentages, and the software WesVarPC (Weststat Inc., 1997) provided standard errors for the coefficients of the regression models presented in both international reports.

REPORTING OF MEAN DIFFERENCES ACROSS COUNTRIES

The aim of the international reports of the IEA Civic Education Study was to compare test and survey results across participating countries, that is, means of scales and percentages were compared in graphs and tables. Each population estimate was accompanied by its standard error. In addition, tests of significance for the difference between estimates were provided to describe the probability that differences were just a result of sampling error.

The following types of significance tests were reported:

- differences in population estimates between countries;
- differences in population estimates between countries and the international mean; and
- differences in population estimates between sub-groups within countries.

Multiple comparison charts allowed the comparison of population estimates between one country and other participating countries. The significance tests included an adjustment for multiple comparison using a Bonferroni adjustment. This was necessary, as the probability of erroneously stating significant differences (the so-called Type I error) increases with the number of simultaneous comparisons.

To test the significance between two means at the .95 level, a critical value of 1.96 is used for the test statistics. Any value higher than the critical value indicates that there is a .95 probability that this difference is not the result of sampling error. Conversely, there is a .05 chance that a difference is found that does not exist. When several means are compared with one another at the same time, the probability of making a Type I error is the product of the probabilities for each comparison. Thus, the chance of making such an error increases with the number of comparisons.

For multiple comparisons in the IEA Civic Education Study, a Dunn-Bonferroni adjustment was used that increased the critical value for significance tests when multiple comparisons were made (Dunn, 1961). This meant that for the standard population of 14-year-old students in the 28 participating countries, the number of simultaneous comparisons for each country was 27. The resulting critical value for the 14-year-old students adjusted for 27 simultaneous comparisons was 3.11295, and the critical value for the upper secondary students adjusted for 13 simultaneous comparisons was 2.89052.

Differences between country means were considered significant when the test statistic t was greater than the critical value. t is calculated by dividing the difference by its standard error as given in the formula

$$SE_{dif\_ij} = \sqrt{SE_i^2 + SE_j^2}$$

where SE_{dif_ij} is the standard error of the difference and SE_i and SE_j are the sampling standard errors of the compared countries i and j.

When comparing the country means with the international average, it was necessary to take into account that the respective country had contributed to the international standard error. This was achieved by calculating the standard error SE_{dif_ic} of the difference between the international average and the country mean as

$$SE_{dif\_ic} = \sqrt{\frac{((N-1)^2 - 1)SE_i^2 + \sum_{k=1}^{N} SE_k^2}{N}}$$

where SE_c is the sampling standard error for country c and SE_k for country k, and N is the number of participating countries. This formula was used for determining the statistical significance of differences between countries and the international averages for cognitive and attitudinal scales throughout the reports.

COMPARING DATA FROM DIFFERENT POPULATIONS

We stated at the beginning of this report that the national samples that were drawn for the study of the upper secondary students did not conform to a uniform international definition of the target population. For pragmatic reasons, countries were left free to choose a grade or study program that appeared important as a subject of investigation and for which funding could be obtained. The implications of such a strategy were clear from the very beginning: comparisons of results obtained for different age groups would be very difficult to defend, both within and across countries.

In principle, this situation applied to all measures reported in the descriptive reports for the upper secondary students no matter whether test results or student attitudes were concerned, and it was due to the fact that differences between samples were confounded by the following:

- effects related to the chosen grade;
- effects related to the age of the students within the chosen grade; and
- effects related to the nature of the study programs of the chosen grade, in particular to the coverage of the respective age group.

Table 10.3 illustrates the main differences between the upper secondary samples assessed in the IEA Civic Education Study. Whereas in some countries like Norway or Poland all upper secondary students were included, other countries like Denmark included a rather selective population definition. These differences were due not only to the way of defining target populations for this particular study, but also to the differences between the educational systems of the participating countries. The coverage index, calculated as the ratio of the estimated target population size divided by the estimated size of the corresponding age group (see Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002, p. 34), illustrates the extent to which the data can be seen as representative for age groups in a country. It ranged from .39 in Switzerland (German-speaking cantons) to .99 in Norway. The table also shows that the grade difference between the population of 14-year-old students and the

population of upper secondary students varied across countries in terms of between two to four years of formal education. An additional difference between populations was evident for Switzerland because only German-speaking schools were included in the sample.

Table 10.3 Characteristics of samples and populations in the survey of upper secondary students

Country	Age	Grade	Grade Difference with 14-year-olds	Coverage Index	Study Program(s)
Chile	17.9	12	+4	.64	Last year of upper secondary
Cyprus	17.7	12	+3	.67	Last year of upper secondary
Czech Republic	17.9	12	+4	.78	Third year of upper secondary (excl. special education)
Denmark	19.4	12	+4	.55	Last year of upper secondary (excl. vocational education)
Estonia	18.2	12	+4	.49	Last year of upper secondary (excl. special education)
Israel	16.8	11		.83	Second year of upper secondary (excl. private and special schools)
Latvia	16.6	12	+2	.89	First year of upper secondary (excl. special schools)
Norway	18.1	12	+4	.99	Second year of upper secondary
Poland	17.6	11	+3	.90	Second year of upper secondary (excl. special schools)
Portugal	17.6	11	+3	.76	Second year of upper secondary
Russian Federation	17.0	11	+2	.50	Last year of upper secondary (excl. special education)
Slovenia	18.4	12	+4	.68	Last year of upper secondary
Sweden	18.9	12	+4	.84	Last year of upper secondary
Switzerland (German)	17.9	12	+4	.39	Third year of upper secondary (excl. special education)

It is obvious that the effects caused by differences in these upper secondary samples can interact in complex patterns. Moreover, their impact (and even their direction) can differ between the various levels of aggregation (country, school type, etc.) present in the data. It is not possible to discuss all of these ramifications here. However, we can show how the test results (total scores as well as sub-scores for content knowledge and interpretative skills) were affected by grade, mean age of the sample, and the coverage of the age cohort (see Amadeo et al., 2002, p. 51ff; Torney-Purta, Lehmann, Oswald, & Schulz, 2001, p. 56). Table 10.4 presents the inter-correlations for these measures, with students and countries as levels of analysis.

Table 10.4 Inter-correlations between cognitive scales and student and sample characteristics in 14 countries

	Total Civic Knowledge	Content Knowledge	Interpretative Skills	Grade Tested	Student Age	Cohort Coverage
Total Civic Knowledge		.893	.948	.492	.755	-.152
Content Knowledge	.892		.713	.388	.592	-.046
Interpretative Skills	.908	.664		.553	.790	-.227
Grade Tested	.178	.116	.206		.761	-.136
Student Age	.112	.070	.129	.510		-.214

Note: Lower diagonal: between students; upper diagonal: between countries.

The table shows that the correlation coefficients for countries as the unit of analysis (upper diagonal) were generally higher than those for the students (lower diagonal), especially in the cases of "grade tested" and "student age". This effect, which is called "aggregation bias", is attributable to neglecting the between-students variance within countries, and it renders any attempt to adjust statistically for sample differences between countries subject to criticism.

Given the inadequacies of attempts to adjust for differences in the defined target populations at the upper secondary level, comparisons between the 14-year-old students and the upper secondary students could only be presented with the caveat that it was impossible to control for the potentially confounding effects of age and cohort coverage. This led to choosing a graphical presentation that included information on the observed achievement levels in the two national samples, the mean age for the two samples, and the coverage of the age cohort in the upper secondary sample (Amadeo et al., 2002, pp. 64, 66ff). Moreover, any attempt to estimate standard errors of achievement differences between the two samples in each country would have suggested a level of precision that, under the given circumstances, simply could not have been attained.

REPORTING OF GENDER DIFFERENCES

The IEA Civic Education Study reported the differences in civic knowledge between females and males. This was done by applying the formula

$$SE_{dif\_ij} = \sqrt{SE_i^2 + SE_j^2}$$

where SE_{dif_ij} is the standard error of the difference and SE_i and SE_j are the sampling standard errors of males (i) and females (j). Simple t-tests with a critical value corrected for multiple comparisons were applied to determine statistical significance.

The formula used to calculate the standard error of the difference assumes that sampling standard errors on the means have been calculated for independent samples, which was clearly not the case for males and females because both groups were parts of the same school samples. Computing the correct standard errors using replication methods was not feasible due to time constraints when preparing the first international report.

Table 10.5 shows the reported standard errors on the difference in civic knowledge between males and females in the population of 14-year-old students compared to those derived using the JRR method. In most countries, the jackknifed standard errors of gender differences in civic knowledge were smaller than the reported ones. Similar figures (see Tables 10.6, 10.7, and 10.8) were obtained when comparing the reported standard errors and those obtained through JRR estimation for gender differences in content knowledge, interpretative skills, and economic literacy for the population of upper secondary students.

Table 10.5 Comparison of standard errors for gender differences in civic knowledge for 14-year-old students

Country	Difference between Males and Females	Reported Standard Errors of Difference	JRR Standard Errors of Difference
Australia	2.37	1.40	1.18
Belgium (French)	4.52	1.68	1.46
Bulgaria	2.24	1.96	1.20
Chile	-1.68	1.14	0.83
Colombia	0.36	1.70	1.65
Cyprus	0.25	0.91	0.81
Czech Republic	-1.68	1.30	0.78
Denmark	-2.53	0.98	0.85
England	-0.12	1.28	1.32
Estonia	1.40	0.90	0.67
Finland	1.59	1.15	0.91
Germany	-1.43	0.92	0.84
Greece	2.07	1.18	0.70
Hong Kong (SAR)	1.08	1.80	1.32
Hungary	0.86	1.04	0.78
Italy	2.10	1.41	1.30
Latvia	3.92	1.32	0.83
Lithuania	2.32	1.08	0.59
Norway	-0.52	0.92	0.86
Poland	2.72	2.62	1.48
Portugal	-1.44	1.16	0.72
Romania	0.42	1.35	0.74
Russian Federation	-0.25	2.05	1.14
Slovak Republic	-0.18	1.14	0.70
Slovenia	3.75	0.83	0.72
Sweden	1.14	1.34	1.10
Switzerland	-2.34	1.22	0.60
United States	1.86	1.80	0.94

Table 10.6 Comparison of standard errors for gender differences in content knowledge for upper secondary students

Country	Difference between Males and Females	Reported Standard Errors of Difference	JRR Standard Errors of Difference
Chile	-1.78	0.43	0.42
Cyprus	0.12	0.89	0.66
Czech Republic	-1.35	0.53	0.52
Denmark	-2.65	0.32	0.30
Estonia	-1.16	0.50	0.38
Israel	-0.88	0.60	0.50
Latvia	0.12	1.05	0.78
Norway	-2.24	0.83	0.74
Poland	-1.92	0.74	0.65
Portugal	-2.65	0.47	0.40
Russian Federation	-1.46	0.85	0.43
Slovenia	-0.50	0.98	0.71
Sweden	0.30	0.59	0.50
Switzerland	-1.53	1.13	1.04

Table 10.7 Comparison of standard errors for gender differences in interpretative skills for upper secondary students

Country	Difference between Males and Females	Reported Standard Errors of Difference	JRR Standard Errors of Difference
Chile	-1.98	1.20	1.26
Cyprus	1.58	2.12	1.28
Czech Republic	-1.70	1.23	1.27
Denmark	-6.08	0.88	0.85
Estonia	-1.59	1.43	1.00
Israel	0.41	1.50	1.21
Latvia	1.36	2.07	1.61
Norway	-3.06	1.88	1.68
Poland	-0.80	1.63	1.38
Portugal	-4.40	0.93	0.76
Russian Federation	-2.68	1.73	1.06
Slovenia	-0.85	2.00	1.50
Sweden	0.75	1.32	1.16
Switzerland	-3.23	3.29	1.75

Table 10.8 Comparison of standard errors for gender differences in economic literacy for upper secondary students

Country	Difference between Males and Females	Reported Standard Errors of Difference	JRR Standard Errors of Difference
Chile	-5.28	0.98	0.94
Cyprus	0.24	2.03	1.40
Czech Republic	-5.23	1.17	1.19
Denmark	-8.16	1.14	1.05
Estonia	-6.40	1.08	0.73
Israel	-3.64	1.28	1.01
Latvia	-1.82	1.94	1.44
Norway	-5.91	1.57	1.37
Poland	-2.90	1.34	1.20
Portugal	-6.42	0.86	0.62
Russian Federation	-5.19	1.88	1.04
Slovenia	-4.61	1.66	1.27
Sweden	-4.09	1.18	1.10
Switzerland	-9.22	3.19	2.65

REPORTING OF MULTIPLE REGRESSION ANALYSIS

The reporting of the results from the IEA Civic Education Study not only describe knowledge, concepts, and attitudes in the participating countries but also provide a first analysis of potential explanatory variables for student outcomes. Multiple regression analysis was used to describe the effects of background and school variables on civic knowledge and civic engagement (as measured by the students' expectation to vote as an adult).

Effects of predictors on the criterion variables in these models should generally not be interpreted as causal. Some of the significant effects might well be due to a non-recursive relationship. It is also recognized that for some of these factors school-level effects might be hypothesized that would have required a multi-level analysis. However, this type of regression analysis was deemed appropriate to give a first indication of variables associated with civic knowledge and civic engagement, and to estimate how much of the variance at the student level could be accounted for by these factors.

For the 14-year-old students, a simple path model with manifest variables was estimated for the calibration sample consisting of 500 randomly selected students per country, with both civic knowledge and civic engagement included as endogenous variables. Then, for each participating country, two different multiple regression models were estimated, each using the same sets of predictor variables. Model A was the multiple regression model for civic knowledge (as measured by the total civic knowledge scale), and Model B was the model for civic engagement (as measured by a Likert-type item on the students' expectation to vote as an adult).

For the reporting on the upper secondary students, separate regression models rather than a path model were estimated for civic knowledge and civic engagement. Some additional predictor variables were also included.

Table 10.9 lists all variables that were included in the multiple regression models A and B. (For a more detailed description of the variables, see Amadeo et al., 2002, p. 145ff; Torney-Purta et al., 2001, p. 148.) Many of the variables used in these models were non-continuous categorical variables, and the OLS estimates for these kinds of data can be biased. This concern was particularly problematic in the case of the indicator for civic engagement, where a non-continuous Likert-type item was used as the criterion variable. The (methodologically more appropriate) use of logistical regression modeling, however, would have rendered rather complex results that would have been difficult to communicate to the reader. Furthermore, using the same type of modeling for both dependent variables provided a consistent way of reporting the results of these analyses.

All multiple regression models were checked for multi-collinearity. For the 14-year-olds, the highest correlation between predictors was found for expected education (in years) and home literacy (number of books at home) with $r = .29$ (international calibration sample from 28 countries); for the upper secondary students it was $r = .32$ for students' reports on reading news in the newspaper and watching TV news (international calibration sample from 14 countries).

Table 10.9 Variables in the multiple regression models

Variables	Type	Model	Population*
Criterion			
Civic knowledge	IRT score	A	2 & 3
Expectation to vote	Likert-type item	B	2 & 3
Predictors			
Gender (female)	dichotomous	A & B	2 & 3
Number of people at home	continuous	A & B	3
Number of books at home	categorical	A & B	2 & 3
Educational level of parents	dichotomous	A & B	3
Expected years of education	categorical	A & B	2 & 3
Open classroom climate	IRT score	A & B	2 & 3
Having learned to vote	Likert-type item	B	2 & 3
Participation in student council	dichotomous	A & B	2 & 3
Civic knowledge	IRT score	B	2 & 3
Reading news in newspaper	Likert-type item	A & B	3
Watching TV news	Likert-type item	A & B	2 & 3
Political interest	Likert-type item	A & B	3
Spending evenings outside	categorical	A & B	2 & 3

Note: *Population 2 = 14-year-old students; Population 3 = upper secondary students.

Listwise-exclusion of missing values was chosen for these analyses. This choice led to a considerable reduction of cases in some countries, particularly for Model B. The Swiss data were not included in the regression models for the upper secondary students because the relatively small sample size would have been even further reduced by the number of missing values. Table 10.10 shows the weighted percentages of cases excluded from the analysis for each regression model.

The JRR method was applied to determine the standard errors of the unstandardized regression coefficients and the significance of each coefficient. Significance tests were adjusted for multiple comparisons using the Dunn-Bonferroni correction. To simplify the interpretation of the results within each country, standardized regression coefficients were used and non-significant coefficients were omitted from the tables in the report.

Table 10.10 Weighted percentage of missing cases for multiple regression analyses

Country	14-year-old Students		Upper Secondary Students	
	Model A	*Model B*	*Model A*	*Model B*
Australia	13	24		
Belgium (French)	18	29		
Bulgaria	12	30		
Chile	6	19	4	12
Colombia	8	13		
Cyprus	4	11	7	13
Czech Republic	3	16	5	13
Denmark	13	30	12	17
England	16	30		
Estonia	5	23	9	16
Finland	5	22		
Germany	8	21		
Greece	5	14		
Hong Kong (SAR)	13	26		
Hungary	4	7		
Israel			18	22
Italy	3	22		
Latvia	9	28	15	31
Lithuania	8	30		
Norway	12	27	35	39
Poland	4	13	8	12
Portugal	7	18	7	11
Romania	4	15		
Russian Federation	4	15	18	22
Slovak Republic	2	8		
Slovenia	5	17	6	12
Sweden	13	31	17	23
Switzerland	11	30		
United States	13	23		
International Mean	**8**	**21**	**12**	**19**

However, the use of standardized coefficients was not appropriate for the comparison of regression coefficients across countries because the differences in variances between countries would have affected the standardization. Tables 10.11 and 10.12 show the unstandardized coefficients and their respective standard errors for the multiple regression models for civic knowledge and civic engagement in the 28 countries that participated in the survey of 14-year-old students. Tables 10.13 and 10.14 present the unstandardized coefficients and JRR standard errors of the regression models for the survey of upper secondary students in 13 countries. The regression coefficients and the standard errors for the effect of civic knowledge on civic engagement (Model B, Tables 10.12 and 10.14) were multiplied by 10 to avoid too small numbers in these tables. The effect coefficients thus indicated the increase in the score of the criterion variable with half a standard error of the civic knowledge score, which was equal to 10 score points.

Table 10.11 Unstandardized coefficients and standard errors of multiple regression models for civic knowledge (14-year-old students)

Country	Constant	Gender (Female)	Home Literacy	Expected Education	Classroom Climate	School Council	Evenings Outside	Television News
Australia	71.67	.36	2.60	2.95	.71	5.58	-2.85	2.81
	(2.63)	(1.01)	(0.34)	(0.42)	(0.18)	(0.97)	(0.49)	(0.46)
Belgium (French)	56.76	2.24	3.09	2.96	1.51	1.10	-1.96	2.32
	(2.42)	(1.23)	(0.42)	(0.37)	(0.22)	(1.18)	(0.41)	(0.60)
Bulgaria	67.01	-.25	3.02	3.76	.84	.47	-.90	1.31
	(2.89)	(0.98)	(0.60)	(0.51)	(0.25)	(1.74)	(0.48)	(0.55)
Chile	57.45	-1.90	3.33	3.03	.99	-1.36	-.18	1.71
	(1.70)	(0.59)	(0.23)	(0.15)	(0.15)	(1.04)	(0.22)	(0.32)
Colombia	69.66	-.24	2.25	1.74	.52	1.71	-.78	.56
	(1.94)	(1.23)	(0.47)	(0.30)	(0.27)	(0.92)	(0.26)	(0.55)
Cyprus	77.74	-2.64	1.32	5.46	.64	6.81	-1.19	.89
	(2.68)	(0.82)	(0.30)	(0.32)	(0.21)	(0.72)	(0.43)	(0.45)
Czech Republic	68.80	-4.60	2.23	7.29	.78	3.01	-2.02	.26
	(2.63)	(0.62)	(0.31)	(0.29)	(0.24)	(1.06)	(0.40)	(0.51)
Denmark	61.60	-3.53	2.12	4.72	1.99	4.29	-3.34	1.79
	(2.57)	(0.71)	(0.33)	(0.33)	(0.17)	(0.72)	(0.40)	(0.48)
England	71.96	-1.67	3.86	2.52	1.16	3.51	-3.56	2.06
	(2.11)	(0.99)	(0.23)	(0.34)	(0.18)	(0.97)	(0.36)	(0.33)
Estonia	65.25	-1.10	2.18	3.14	1.32	.58	-3.76	1.37
	(2.48)	(0.63)	(0.33)	(0.24)	(0.19)	(0.82)	(0.31)	(0.36)
Finland	80.93	-1.54	1.86	6.54	.23	.84	-2.64	2.81
	(2.70)	(0.87)	(0.33)	(0.36)	(0.25)	(0.87)	(0.47)	(0.49)
Germany	67.50	-3.05	3.69	3.50	1.21	3.41	-2.55	1.87
	(2.58)	(0.66)	(0.25)	(0.32)	(0.19)	(1.08)	(0.32)	(0.36)
Greece	71.06	-1.38	1.88	6.55	1.19	5.60	-2.28	1.26
	(2.72)	(0.65)	(0.35)	(0.41)	(0.24)	(0.66)	(0.56)	(0.54)

Note: JRR estimates of standard errors are in brackets.

table contd. on next page

Table 10.11 contd.from previous page

Country	Constant	Gender (Female)	Home Literacy	Expected Education	Classroom Climate	School Council	Evenings Outside	Television News
Hong Kong (SAR)	79.61 (2.73)	-1.21 (1.04)	-.21 (0.30)	2.59 (0.27)	.90 (0.20)	2.32 (0.83)	-5.60 (0.47)	6.33 (0.49)
Hungary	65.95 (1.99)	-1.78 (0.61)	2.32 (0.27)	6.44 (0.36)	.53 (0.14)	2.32 (0.77)	-2.04 (0.34)	.55 (0.39)
Italy	71.12 (2.36)	-1.21 (0.95)	2.76 (0.31)	3.42 (0.35)	1.25 (0.23)	.53 (1.02)	-1.96 (0.34)	2.37 (0.46)
Latvia	60.44 (2.38)	1.40 (0.75)	2.10 (0.39)	3.18 (0.30)	1.24 (0.19)	2.30 (1.37)	-1.76 (0.41)	1.38 (0.37)
Lithuania	63.96 (1.99)	.14 (0.58)	1.86 (0.31)	3.60 (0.27)	1.15 (0.20)	2.71 (0.97)	-1.24 (0.35)	1.57 (0.34)
Norway	63.71 (2.60)	-3.03 (0.74)	2.96 (0.34)	4.69 (0.40)	1.30 (0.21)	5.42 (0.86)	-2.76 (0.53)	2.66 (0.51)
Poland	69.10 (2.85)	-1.98 (1.08)	2.43 (0.37)	6.89 (0.77)	.20 (0.18)	2.80 (1.50)	-1.49 (0.41)	2.74 (0.47)
Portugal	71.94 (2.29)	-3.65 (0.61)	2.55 (0.37)	2.70 (0.17)	1.32 (0.21)	-.51 (0.61)	-2.17 (0.33)	.91 (0.37)
Romania	75.29 (2.88)	-1.46 (0.80)	1.22 (0.36)	2.77 (0.41)	.62 (0.28)	1.54 (1.18)	-1.67 (0.45)	1.26 (0.42)
Russian Federation	61.67 (4.78)	-3.44 (1.00)	3.63 (0.63)	1.77 (0.34)	2.23 (0.36)	3.09 (1.53)	-2.34 (0.75)	1.20 (0.61)
Slovak Republic	62.51 (2.45)	-1.93 (0.54)	2.21 (0.33)	4.72 (0.35)	1.39 (0.19)	1.73 (2.17)	.01 (0.29)	1.92 (0.37)
Slovenia	64.42 (2.08)	.27 (0.62)	1.57 (0.27)	5.33 (0.26)	1.34 (0.23)	4.79 (0.87)	-1.29 (0.32)	1.32 (0.39)
Sweden	58.56 (2.83)	-3.14 (1.00)	3.23 (0.42)	4.57 (0.41)	1.69 (0.26)	3.58 (0.83)	-1.54 (0.43)	.86 (0.64)
Switzerland	69.88 (2.62)	-3.94 (0.63)	3.05 (0.34)	3.99 (0.39)	1.11 (0.23)	-.75 (1.48)	-1.24 (0.39)	.81 (0.45)
United States	64.49 (2.65)	-1.49 (0.79)	3.93 (0.46)	4.97 (0.37)	1.16 (0.19)	2.88 (1.00)	-2.49 (0.54)	1.63 (0.60)

Note: JRR estimates of standard errors are in brackets.

Table 10.12 Unstandardized coefficients and standard errors of multiple regression models for expected vote (14-year-old students)

Country	Constant	Gender (Female)	Home Literacy	Expected Education	Classroom Climate	Learned to Vote	School Council	Evenings Outside	Television News	Civic Knowledge
Australia	0.56 (0.13)	0.03 (0.03)	0.03 (0.01)	0.04 (0.02)	0.06 (0.01)	0.12 (0.02)	0.08 (0.03)	-0.07 (0.02)	0.18 (0.03)	0.12 (0.01)
Belgium (French)	-0.32 (0.21)	0.17 (0.05)	0.03 (0.02)	0.08 (0.02)	0.05 (0.02)	0.17 (0.03)	-0.03 (0.08)	-0.06 (0.03)	0.19 (0.03)	0.15 (0.02)
Bulgaria	0.05 (0.19)	0.04 (0.05)	0.04 (0.02)	0.04 (0.02)	0.06 (0.02)	0.13 (0.03)	0.12 (0.06)	0.01 (0.02)	0.19 (0.03)	0.09 (0.01)
Chile	0.23 (0.15)	-0.04 (0.04)	0.03 (0.02)	0.02 (0.01)	0.06 (0.01)	0.31 (0.02)	0.04 (0.04)	0.00 (0.02)	0.13 (0.02)	0.09 (0.01)
Colombia	0.61 (0.15)	0.05 (0.04)	0.02 (0.02)	0.03 (0.01)	0.08 (0.01)	0.20 (0.02)	-0.01 (0.06)	-0.01 (0.02)	0.17 (0.03)	0.08 (0.01)
Cyprus	2.43 (0.16)	0.01 (0.03)	-0.02 (0.01)	0.04 (0.01)	0.03 (0.01)	0.06 (0.02)	0.01 (0.03)	-0.03 (0.02)	0.13 (0.02)	0.04 (0.01)
Czech Republic	-0.46 (0.20)	-0.10 (0.04)	0.05 (0.02)	0.10 (0.02)	0.00 (0.01)	0.27 (0.03)	0.09 (0.06)	-0.02 (0.02)	0.16 (0.04)	0.16 (0.01)
Denmark	1.09 (0.12)	0.08 (0.03)	0.05 (0.01)	0.06 (0.01)	0.04 (0.01)	0.12 (0.02)	0.03 (0.03)	0.02 (0.02)	0.14 (0.02)	0.08 (0.01)
England	0.73 (0.17)	-0.04 (0.04)	0.06 (0.02)	0.06 (0.02)	0.04 (0.01)	0.16 (0.02)	0.09 (0.04)	-0.03 (0.02)	0.17 (0.03)	0.10 (0.01)
Estonia	-0.25 (0.17)	0.12 (0.04)	0.04 (0.02)	0.05 (0.01)	0.04 (0.01)	0.24 (0.02)	0.06 (0.05)	-0.04 (0.02)	0.14 (0.02)	0.15 (0.01)
Finland	1.04 (0.16)	0.06 (0.03)	0.04 (0.02)	0.12 (0.02)	0.03 (0.01)	0.27 (0.02)	0.07 (0.04)	0.01 (0.02)	0.08 (0.03)	0.05 (0.01)
Germany	-0.27 (0.15)	-0.09 (0.03)	0.09 (0.02)	0.02 (0.01)	0.04 (0.01)	0.24 (0.03)	0.14 (0.06)	0.00 (0.02)	0.12 (0.02)	0.15 (0.01)
Greece	0.95 (0.17)	0.12 (0.03)	-0.01 (0.01)	0.03 (0.01)	0.08 (0.01)	0.09 (0.02)	0.08 (0.04)	-0.03 (0.02)	0.17 (0.02)	0.08 (0.01)
Hong Kong (SAR)	0.46 (0.12)	0.03 (0.03)	0.00 (0.01)	0.03 (0.01)	0.07 (0.01)	0.21 (0.02)	0.05 (0.03)	-0.04 (0.02)	0.19 (0.02)	0.08 (0.01)
Hungary	1.08 (0.14)	0.03 (0.02)	0.05 (0.02)	0.04 (0.02)	0.03 (0.01)	0.16 (0.02)	0.09 (0.03)	-0.01 (0.02)	0.15 (0.02)	0.08 (0.01)
Italy	0.49 (0.14)	-0.03 (0.03)	0.00 (0.02)	0.06 (0.02)	0.03 (0.01)	0.17 (0.02)	-0.07 (0.05)	-0.05 (0.02)	0.17 (0.03)	0.12 (0.01)
Latvia	-0.21 (0.26)	0.18 (0.05)	0.02 (0.03)	0.03 (0.02)	0.07 (0.01)	0.24 (0.03)	-0.02 (0.06)	-0.05 (0.03)	0.25 (0.03)	0.11 (0.02)
Lithuania	0.82 (0.12)	0.08 (0.04)	0.02 (0.02)	0.04 (0.01)	0.04 (0.01)	0.08 (0.02)	0.06 (0.04)	-0.04 (0.02)	0.21 (0.03)	0.10 (0.01)
Norway	0.78 (0.12)	0.01 (0.03)	0.02 (0.01)	0.06 (0.02)	0.03 (0.01)	0.13 (0.02)	0.10 (0.03)	0.03 (0.02)	0.18 (0.03)	0.10 (0.01)
Poland	0.85 (0.18)	0.12 (0.03)	-0.01 (0.01)	0.08 (0.01)	0.03 (0.01)	0.13 (0.02)	0.04 (0.03)	-0.02 (0.01)	0.19 (0.04)	0.09 (0.01)
Portugal	1.05 (0.14)	-0.06 (0.03)	0.00 (0.01)	0.04 (0.01)	0.04 (0.01)	0.13 (0.02)	0.02 (0.04)	-0.01 (0.02)	0.13 (0.02)	0.11 (0.01)

Notes: JRR estimates of standard errors below coefficients are in brackets. Coefficients and standard errors for civic knowledge are multiplied by 10.

Table 10.12 contd. from previous page

Country	Constant	Gender (Female)	Home Literacy	Expected Education	Classroom Climate	Learned to Vote	School Council	Evenings Outside	Television News	Civic Knowledge
Romania	0.93 (0.19)	-0.01 (0.04)	0.03 (0.02)	0.03 (0.02)	0.01 (0.02)	0.18 (0.02)	0.06 (0.04)	0.00 (0.03)	0.15 (0.03)	0.11 (0.02)
Russia	0.52 (0.14)	0.10 (0.04)	0.04 (0.02)	0.01 (0.01)	0.07 (0.01)	0.14 (0.04)	0.05 (0.04)	0.03 (0.03)	0.21 (0.02)	0.07 (0.01)
Slovak Republic	1.45 (0.12)	0.07 (0.03)	0.03 (0.02)	0.06 (0.02)	0.03 (0.01)	0.12 (0.02)	0.01 (0.06)	0.00 (0.01)	0.16 (0.02)	0.07 (0.01)
Slovenia	1.11 (0.15)	0.06 (0.04)	0.00 (0.02)	0.07 (0.01)	0.03 (0.01)	0.17 (0.02)	0.05 (0.04)	-0.06 (0.02)	0.16 (0.02)	0.08 (0.01)
Sweden	-0.26 (0.18)	0.16 (0.05)	0.09 (0.03)	0.04 (0.02)	0.05 (0.02)	0.23 (0.03)	0.04 (0.05)	-0.06 (0.03)	0.16 (0.03)	0.13 (0.01)
Switzerland	0.20 (0.20)	-0.12 (0.04)	0.11 (0.02)	0.09 (0.02)	0.01 (0.01)	0.23 (0.03)	0.07 (0.08)	-0.06 (0.03)	0.07 (0.04)	0.10 (0.02)
United States	0.37 (0.14)	0.08 (0.03)	0.08 (0.01)	0.06 (0.02)	0.06 (0.01)	0.20 (0.02)	0.11 (0.03)	0.01 (0.02)	0.12 (0.02)	0.08 (0.01)

Notes: JRR estimates of standard errors below coefficients are in brackets. Coefficients and standard errors for civic knowledge are multiplied by 10.

Table 10.13 Unstandardized coefficients and standard errors of multiple regression models for civic knowledge (upper secondary students)

Country	Constant	Gender (Female)	People in Household	Home Literacy	Parental Education	Expected Education	Classroom Climate	School Council	Reading Newspaper	Television News	Interest in Politics	Evenings Outside
Chile	86.707 (2.080)	-3.570 (.773)	-.731 (.119)	2.206 (.262)	6.240 (.735)	3.702 (.248)	.223 (.143)	.006 (.726)	1.745 (.396)	.882 (.383)	1.404 (.227)	-.332 (.225)
Cyprus	100.094 (4.686)	-.142 (1.324)	-1.481 (.748)	1.026 (.438)	5.810 (1.232)	5.306 (.840)	.032 (.249)	6.088 (1.017)	.422 (.686)	1.574 (.635)	1.710 (.467)	-2.537 (.840)
Czech Republic	96.862 (3.279)	-4.694 (.887)	-.723 (.309)	2.429 (.345)	3.271 (.857)	6.036 (.380)	.306 (.235)	.122 (.928)	.362 (.528)	.025 (.519)	2.395 (.428)	-.617 (.393)
Denmark	110.300 (2.644)	-7.547 (.715)	-.440 (.294)	2.061 (.310)	3.406 (.929)	3.188 (.327)	.402 (.195)	.636 (.707)	.745 (.437)	-.813 (.584)	3.871 (.458)	-1.584 (.512)
Estonia	101.977 (4.329)	-3.232 (.930)	-.117 (.345)	2.457 (.442)	5.605 (1.393)	.702 (.671)	.631 (.215)	.725 (1.026)	.973 (.629)	1.222 (.754)	2.339 (.610)	-4.388 (.631)
Israel	101.290 (2.614)	-2.668 (.894)	-2.339 (.225)	1.678 (.297)	6.950 (.749)	3.253 (.335)	.745 (.179)	1.440 (.814)	3.263 (.418)	-.119 (.494)	1.196 (.372)	-1.270 (.441)
Latvia	66.961 (3.528)	-1.972 (1.327)	-1.018 (.304)	1.597 (.617)	5.340 (1.312)	3.849 (.467)	1.903 (.289)	-1.646 (1.253)	.995 (.529)	1.133 (.613)	3.008 (.652)	-2.734 (.541)
Norway	83.915 (5.441)	-7.820 (1.391)	-.877 (.502)	2.989 (.488)	.785 (1.007)	5.327 (.573)	.928 (.338)	.044 (1.175)	1.937 (1.069)	1.071 (1.082)	4.264 (.648)	-1.134 (.775)
Poland	97.722 (2.408)	-3.859 (1.056)	-1.264 (.237)	2.522 (.384)	5.818 (1.174)	4.776 (.455)	.050 (.190)	-3.090 (1.141)	1.879 (.541)	1.358 (.511)	1.160 (.487)	-2.855 (.402)
Portugal	100.911 (3.445)	-6.145 (.814)	-1.002 (.335)	1.489 (.312)	4.465 (1.152)	3.856 (.313)	.882 (.246)	-.203 (.799)	-.697 (.583)	.693 (.631)	2.856 (.405)	-1.739 (.398)
Russian Federation	73.869 (6.376)	-4.403 (1.005)	.175 (1.080)	1.471 (.741)	7.252 (1.346)	1.581 (.583)	1.676 (.388)	3.073 (1.253)	-.386 (.699)	1.978 (1.021)	2.735 (.920)	-2.533 (.733)
Slovenia	76.001 (3.719)	-1.379 (1.448)	-.610 (.296)	2.634 (.556)	4.573 (1.252)	2.232 (.640)	1.544 (.296)	4.491 (1.949)	1.319 (.563)	4.650 (.857)	-.611 (.717)	-.119 (.672)
Sweden	90.462 (6.466)	-4.743 (1.015)	-2.623 (.686)	4.045 (.556)	5.114 (1.278)	1.473 (.796)	1.902 (.343)	2.988 (.678)	2.041 (.923)	.429 (1.212)	4.071 (.655)	-1.839 (.602)

Note: JRR estimates of standard errors are in brackets.

Table 10.14 Unstandardized coefficients and standard errors of multiple regression models for expectation to vote (upper secondary students)

Country	Constant	Gender (Female)	People in Household	Home Literacy	Parental Education	Expected Education	Classroom Climate	Learned to Vote	School Council	Civic Knowledge	Reading Newspaper	Television News	Interest in Politics	Evenings Outside
Chile	.508 (.126)	.139 (.031)	.003 (.007)	-.009 (.012)	.083 (.031)	.028 (.013)	.011 (.009)	.236 (.015)	.094 (.033)	.050 (.007)	.113 (.023)	.090 (.020)	.277 (.016)	.012 (.012)
Cyprus	1.957 (.245)	.083 (.038)	.033 (.013)	.000 (.022)	-.006 (.042)	.009 (.023)	.029 (.012)	.089 (.029)	.104 (.039)	.036 (.013)	.053 (.022)	.112 (.039)	.091 (.018)	-.079 (.028)
Czech Republic	-.046 (.176)	-.027 (.031)	.013 (.017)	.037 (.019)	.054 (.032)	.094 (.014)	.026 (.011)	.229 (.020)	.096 (.041)	.105 (.010)	.062 (.026)	.051 (.025)	.201 (.025)	-.021 (.020)
Denmark	2.024 (.110)	-.004 (.022)	.009 (.007)	.014 (.010)	.051 (.024)	.006 (.010)	.007 (.006)	.127 (.015)	.031 (.018)	.029 (.006)	.010 (.016)	.082 (.023)	.177 (.013)	.016 (.014)
Estonia	.731 (.180)	.099 (.039)	.005 (.012)	.039 (.026)	.093 (.026)	.011 (.023)	.017 (.007)	.233 (.021)	.015 (.041)	.064 (.010)	.082 (.027)	.090 (.026)	.087 (.020)	-.036 (.018)
Israel	1.270 (.135)	.080 (.027)	-.024 (.007)	.002 (.012)	.060 (.028)	.007 (.011)	.031 (.006)	.051 (.013)	.092 (.022)	.075 (.008)	.192 (.021)	.115 (.023)	.043 (.017)	.043 (.016)
Latvia	-.265 (.282)	.047 (.058)	.015 (.018)	.053 (.028)	.030 (.041)	.031 (.017)	.067 (.010)	.223 (.031)	-.004 (.051)	.077 (.016)	.039 (.041)	.263 (.062)	.088 (.027)	-.022 (.026)
Norway	1.243 (.244)	.051 (.038)	-.030 (.015)	.015 (.015)	.066 (.035)	.074 (.022)	.033 (.010)	.176 (.031)	-.040 (.045)	.031 (.010)	.107 (.040)	.082 (.042)	.152 (.019)	.033 (.031)
Poland	1.618 (.145)	.125 (.026)	.010 (.007)	.004 (.011)	.027 (.028)	.044 (.015)	.020 (.006)	.136 (.015)	-.002 (.037)	.044 (.006)	.072 (.019)	.125 (.022)	.066 (.017)	-.020 (.013)
Portugal	1.668 (.168)	.096 (.034)	.008 (.012)	-.006 (.015)	.050 (.043)	.027 (.014)	.002 (.013)	.156 (.018)	.006 (.029)	.041 (.009)	.059 (.021)	.104 (.030)	.171 (.023)	-.018 (.018)
Russian Federation	1.694 (.219)	-.032 (.045)	-.035 (.039)	.026 (.025)	-.055 (.033)	.003 (.022)	.029 (.008)	.165 (.030)	.059 (.042)	.048 (.009)	.029 (.034)	.116 (.042)	.072 (.030)	-.001 (.024)
Slovenia	1.331 (.173)	.051 (.040)	.017 (.013)	.023 (.016)	-.068 (.047)	.001 (.015)	.034 (.013)	.133 (.030)	.155 (.047)	.048 (.011)	.109 (.024)	.110 (.029)	.052 (.018)	-.016 (.018)
Sweden	.567 (.177)	.064 (.045)	-.018 (.027)	.056 (.020)	.034 (.043)	.022 (.020)	.038 (.012)	.128 (.030)	.017 (.038)	.097 (.012)	.093 (.035)	.055 (.032)	.086 (.024)	.002 (.020)

Notes: JRR estimates of standard errors are in brackets. Coefficients and standard errors for civic knowledge are multiplied by 10.

REPORTING OF TEACHER DATA (14-YEAR-OLD STUDENTS)

The assessment and survey of the 14-year-old students were accompanied by a teacher survey. In a loosely bounded curricular field such as civic education, it is a particular challenge to draw a sample of teachers that is comparable across countries. From the case study data obtained during Phase I of the IEA Civic Education Study, it was known that in some countries civic education is affiliated with history and in other countries with mother tongue, or that it also may be taught as a specific subject. For some countries, civic education is even lodged in the domain of religious instruction, while for others it is developed as a specific amalgamated school subject called social studies that draws teachers from multiple social science disciplinary backgrounds. Civic education may also be found in some countries as an encompassing cross-curricular concern of the whole school. In this case, teachers from all disciplinary backgrounds may teach in the field.

To ensure a comparable sample across countries, a *subject allocation grid* (see Appendix C) was composed that listed the topics from which items for the cognitive part of the student questionnaire were drawn. The national research coordinators were asked to identify which teachers, teaching which subjects, were primarily responsible for covering these topics in their countries (see Chapter 3 for details). Each sampled school was asked to administer the teacher questionnaire to three such teachers. Schools were asked to choose their teachers in this sequence:

1. Three teachers of civic education-related subjects teaching the tested class of students.
2. If three teachers could not be selected this way, then other teachers of civic education-related subjects of a parallel, previous, or following grade within the school.

The second condition applied to almost all countries. Selected teachers who declined to participate were not substituted. Thus, the questionnaire was administered to both teachers "linked" and "not linked" to the tested class. Reporting of teacher data, however, was restricted to teachers who reported that they were teaching the tested class.

As the selection procedure for the teacher questionnaire was based on participating students, the sampled teachers did not necessarily represent all teachers from civic-related subjects in a country but rather the teachers of the representative samples of *students sampled for this study*. Consequently, the reporting of teacher data in the IEA Civic Education Study was linked to the student sample, and the teacher results were weighted with the number of students they were teaching. The unit of analysis for all results was the student, not the teacher. If a student was taught by more than one teacher, the student's weight was distributed equally between the teachers who were teaching the student. Teacher data from Colombia and the United States were omitted from the international report due to country-specific problems regarding the linkage between students and teachers. However, all teacher data have been included in the international database.

Table 10.15 shows a fictitious example of how teachers were linked to student weights. In this example, the weights of students 1 to 10 have been assigned to teachers A, B, and C. Whereas teacher C teaches only one group of students (students 6 to 10), teachers A and B teach the same group of students (students 1 to 5). Thus, student weights for the latter group of students have been divided by the number of teachers per students (in this case two) and assigned to both teachers. The sum of weights in the resulting weighted teacher file was therefore always equal to the sum of weights in the student file (as shown in the example).

Table 10.15 Example for the assignment of student weights to teacher data

Teacher	Student	Original Student Weight	Assigned Weight
A	1	7.2	3.6
B	1		3.6
A	2	7.2	3.6
B	2		3.6
A	3	7.2	3.6
B	3		3.6
A	4	7.2	3.6
B	4		3.6
A	5	7.2	3.6
B	5		3.6
C	6	4.9	4.9
C	7	4.9	4.9
C	8	4.9	4.9
C	9	4.9	4.9
C	10	4.9	4.9
Sum of Weights		**60.5**	**60.5**

Data based on the teacher survey were generally presented as percentages of students whose teachers reported characteristics, attitudes, or instructional strategies, or as "student-weighted" averages. Standard errors were estimated using the JRR method, as was done for the reporting of student data.

It should be noted that data from a representative teacher survey based on a random sample of the same target population in a given country might produce results different from those obtained in the IEA Civic Education Study. The approach used here, however, is consistent with the goal of this study, which was to collect data on the instructional context of civic education for 14-year-old students.

REFERENCES

Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries.* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56,* 52–64.

Gonzalez, E. J, & Foy, P. (2000). Reporting student achievement in mathematics and science. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999: Technical report* (pp. 279–295). Chestnut Hill, MA: Boston College.

Gonzalez, E. J, & Gregory, P. (2000). Estimation of sampling variance. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999: Technical report* (pp. 203–222). Chestnut Hill, MA: Boston College.

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

Weststat, Inc. (1997). *A user's guide to WesVarPC.* Rockville, MD: Author.

Winglee, M., Kalton, G., Rust, K., & Kasprzyk, D. (2001). Handling item non-response in the U.S. component of the IEA reading literacy study. *Journal of Educational and Behavioral Statistics, 26*(3), 343–359.

Wolter, K. M. (1985). *Introduction to variance estimation.* New York: Springer.

APPENDIX A: ACKNOWLEDGMENTS

The IEA Civic Education Study began Phase 1 in 1994, transitioning in 1997 into Phase 2. The results of case studies carried out during Phase 1 were published in *Civic Education Across Countries: Twenty-four National Case Studies from the IEA Civic Education Project* (IEA, 1999). The results of a survey of 14-year-olds were published in *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen* (IEA, 2001), and the results of a survey of upper secondary students in *Civic Knowledge and Engagement: An IEA Study of Upper Secondary Students in Sixteen Countries* (IEA, 2002). Web information on this study can be found at > http://www.wam.umd.edu/ ~iea<.

The study has been a truly collaborative effort among hundreds of individuals around the world. Staff from the national research centers in each participating country, the International Association for the Evaluation of Educational Achievement (IEA), the International Coordinating Center for Phase 1 at the University of Maryland, the International Coordinating Center for Phase 2 at the Humboldt University in Berlin, the IEA Data Processing Center, the members of the International Steering Committee, advisors, and funding agencies worked closely together to develop and implement the IEA Civic Education Study. Its success would not have been possible without the tireless efforts of all involved.

Individuals and organizations participating in this study are acknowledged for their contributions. Given that the study spanned more than eight years, the list may not include all who have contributed throughout the life of this project. Any omission is inadvertent. The IEA Civic Education Study also would like to thank the 140,000 students who responded to the IEA Civic Education surveys of 14-year-old and upper secondary students. Without the cooperation of school administrators, teachers, and students, we would not have been privileged to have this glimpse into the civic world of young people across three continents.

FUNDING AGENCIES

Since 1998, the major funder for the international coordination of Phase 2 (reported in this volume) has been the Deutsche Forschungsgemeinschaft (DFG or German Science Association) through a grant to the Humboldt University of Berlin to support the International Coordinating Center. Since 2000, substantial additional funding for analysis and publication has been provided by the William T. Grant Foundation through a grant to the University of Maryland at College Park (Department of Human Development). Other funding and support in Germany came from the Bundeszentrale für Politische Bildung (Federal Agency for Civic Education), and in the United States from the National Center for Education Statistics of the United States Department of Education (for quality control and scaling) and the Carnegie Corporation of New York (for instrument preparation during the transition from Phase 1 to Phase 2). The Pew Charitable Trusts of Philadelphia funded Phase 1 of the study.

MANAGEMENT AND OPERATIONS

The IEA Civic Education Study was conducted under the auspices of the IEA. Although the study was centrally managed by the International Coordinating Centers at the University of Maryland during Phase 1 and at the Humboldt University of Berlin during Phase 2, important activities were carried out in centers around the world. The IEA Data Processing Center (DPC) in Hamburg was responsible for data processing, the implementation of the CivEd sample design and the calculation of sampling weights.

The IEA Secretariat in Amsterdam provided funds and support throughout both phases of the study, and managed matching funds from the participating countries. Barbara Malak-Minkiewicz was the liaison between the research centers and the Secretariat. Hans Wagemaker (Executive Director), Alejandro Tiana Ferrer (Chair), and Leendert Dijkhuizen (Financial Officer) provided important support throughout this study.

Paula Wagemaker provided thoughtful editing of this volume that was always responsive to authors' concerns, and Becky Bliss designed the publication with skill.

INTERNATIONAL COORDINATING CENTER (PHASE 1)

Judith Torney-Purta, *International Coordinator and Chair of International Steering Committee*
Jo-Ann Amadeo, *Senior Researcher*

INTERNATIONAL COORDINATING CENTER (PHASE 2)

Rainer Lehmann, *International Coordinator*
Donald Gray, *Associate International Coordinator (1997–1998)*
Wolfram Schulz, *Associate International Coordinator (1998–2001)*
Vera Husfeldt, *Associate International Coordinator (2001–2002)*
Roumiana Nikolova, *Associate International Coordinator (2002–2003)*

IEA DATA PROCESSING CENTER

Heiko Sibberns, *Senior Researcher*
Pierre Foy, *Senior Researcher*
Angelica Barajas, *Research Assistant*
Jens Brockmann, *Research Assistant*
Ursula Itzlinger, *Research Assistant*
Kristin Habeck, *Research Assistant*

NATIONAL RESEARCH COORDINATORS

The IEA Civic Education Study national research coordinators (NRCs) and their staff had the enormous task of implementing the study design in their countries. They had to obtain funding, attend international meetings, contribute to the development of the international study, adapt international procedures and instruments to their national context, work with schools to arrange test and questionnaire sessions, arrange the data entry, prepare data files for submission to the IEA Data Processing Center, and prepare national reports. The commitment and competence of the NRCs over nearly seven years made this collaborative effort both successful and enjoyable. They participated in every part of Phase 2, from reviewing instruments, to sampling, testing, and supervising data entry, and, finally, reviewing analysis and this volume. Their thoughtful comments at meetings and as reviewers have enriched all participants' understanding of this subject area and its cross-national variations.

Addresses of National Research Coordinators

AUSTRALIA

Kerry Kennedy
University of Canberra, University Drive
Bruce ACT 2616
Email: kerryk@adminserver.canberra.edu.au

BELGIUM (FRENCH)

Annette Lafontaine and Christiane Blondin
Service de Pedagogie Experimentale
5 Bld du Rectorat
B4000 Liege (Sart-Tilman)
[http://www.ulg.ac.be/pedaexpe/SPE98.html#A3.1]
Emails: a.lafontaine@ulg.ac.be
cblondin@ulg.ac.be

BULGARIA

Roumen Nikolov
Sofia University
Faculty of Math & Informatics
Department of Information Technology
5, James Bouchier St., Sofia 1126
Sofia 1126
[http://www-it.fmi.uni-sofia.bg/~roumen/]
Email: roumen@fmi.uni-sofia.bg

CHILE

Leonor Cariola
Ministerio de Educación
UCE
Alameda Bernardo O'Higgins # 1146, Sector B, Piso 8
Santiago
[http://www.mineduc.cl]
Email: lcariola@mineduc.cl

COLOMBIA

Carlos A. Pardo and Jose Guillermo Ortiz
SNP-ICFES
Calle 17 Nr. 3-40 piso 10
Bogotá
[http://www.icfes.co]
Emails: cpardo@hemeroteca.icfes.gov.co
profab400@acuario.icfes.gov

CYPRUS

Constantinos Papanastasiou and Mary Koutselini
University of Cyprus
Department of Education
P.O. Box 20537
CY-1678 Nicosia
[http://www.ucy.ac.cy/faculty/cp/]
Emails: edpapan@ucy.ac.cy
edmaryk@ucy.ac.cy

CZECH REPUBLIC

Jana Strakova and Ivana Krizova
UIV/IIE
Senovazne nam.26
P.O. BOX c.1
11006 Praha 1
[http://www.uiv.cz/]
Emails: janastr@alfa.uiv.cz
ivana@alfa.uiv.cz

DENMARK

Lars-Henrik Schmidt
Danmarks Pædagogiske Universitet
Emdrupvej 101
DK-2400 København NV
Email: schmidt@dpu.dk

ENGLAND

David Kerr
National Foundation for Educational Research (NFER)
The Mere
Upton Park, Slough, Berkshire, SLI 2DQ
[http://www.nfer.ac.uk/risheets/iec.htm]
Email: d.Kerr@nfer.ac.uk

ESTONIA

Anu Toots
Tallin University of Educational Sciences
Department of Government
Narva mnt. 25
10120 Tallinn
[http://www.tpu.ee/]
Email: anuto@tpu.ee

FINLAND

Sakari Suutarinen
University of Jyväskylä
Institute for Education Research
P.O. Box 35, 40351
Jyväskylä
[http://www.jyu.fi/ktl/civics.htm]
Email: suutarin@piaget.edu.jyu.fi

GERMANY

Jürgen Baumert and Detlef Oesterreich
Max-Planck-Institut für Bildungsforschung
Lentzeallee 94
D-14195 Berlin
[http://www.mpib-berlin.mpg.de/EuB/program/areas/projectI-4.htm]
Emails: sekbaumert@mpib-berlin.mpg.de
Oest@mpib-berlin.mpg.de

GREECE

Georgia Kontogiannopoulou-Polydorides
University of Athens
Department of Pre-school Education, Neo Chimio 3rd Floor
13 A Navarinou St.
Athens 10680
[http://www.cc.uoa.gr/]
Email: gpol@cc.uoa.gr

HONG KONG (SAR)

Lee Wing On
Hong Kong Institute of Education
Centre for Citizenship Education
10 Lo Ping Road, Tai Po, New Territories
[http://www.ied.edu.hk/sfe/cce/eindex.htm]
Email: WOLEE@ied.edu.hk

HUNGARY

Zsuzsa Matrai
National Institute of Public Education
Center for Evaluation Studies
Dorottya u. 8
1051 Budapest
[http://www.oki.hu/1e.htm]
Email: matraiz@mail.matav.hu

ISRAEL

Orit Ichilov
Tel Aviv University
School of Education
69978 Tel Avivi
[http://www.tau.ac.il/education/nsm.html]
Email: ichilov@post.tau.ac.il

ITALY

Bruno Losito

National Center for the Evaluation of the Education System (CEDE)

Villa Falconieri

I-00044 Frascati (Roma)

[http://www.cede.it/cede/default.htm]

Email: blosito@cede.it

LATVIA

Andris Kangro

University of Latvia

Faculty of Education and Psychology

Jurmalas gatve 74/76

Riga LV-1083

[http://www.eduinf.lu.lv]

Email: kangro@eduinf.lu.lv

LITHUANIA

Irena Zaleskiene

Institute of Pedagogy

Department of Social Sciences

M. Katkaus 44

2600 Vilnius

[http://www.elnet.lt/pi/e-about.htm]

Email: zaleskiene@pi.elnet.lt

NORWAY

Rolf Mikkelsen

University of Oslo

Institute for Teacher Education and School Development

PO Box 1099 Blindern

0316 Oslo

[http://www.ils.uio.no/civic/]

Email: rolf.mikkelsen@ils.uio.no

POLAND

Adam Fraczek

University of Warsaw

Faculty of Education

Mokotowska Str.16/20

00-561 Warsaw

[http://www.iss.uw.edu.pl/]

Email: adamfra@samba.iss.uw.edu.pl

PORTUGAL

Isabel Menezes

Inst. De Inovacao Educacional

Tr. Das Terras de Sant 'Ana 15

1250 Lisboa Codex

[http://www.iie.min-edu.pt/eng/iieproj.htm]

Email: imenezes@psi.up.pt

ROMANIA

Gheorghe Bunescu
Institute for Education Sciences
Str. Stirbei Voda, nr. 37
70732 Bucharest
Email: ise@acc.usis.ro

RUSSIAN FEDERATION

Galina Kovalyova
Russian Academy of Education
Center for Evaluating the Quality of Education
Institute for General Secondary Education
Pogodinskay st. 8, rooms 201, 202, 204
119905 Moscow
Emails: gkovalev@aha.ru
centeroko@ioso.iip.net

SLOVAK REPUBLIC

Maria Capova
SPU-National Institute for Education
Pluhova 8
SK - 830 +Bratislava
Email: spu@spu.sanet.sk

SLOVENIA

Marjan Simenc
Educational Research Institute, Pedagoski Institut
Gerbiceva 62, p.p. 76
61111 Ljubljana
Email: marjan.simenc@guest.arnes.si

SWEDEN

Sverker Hard and Kersti Blidberg
Swedish National Agency for Education
Skolverket
SE-106 20 Stockholm
[http://www.skolverket.se/]
Emails: sverker.hard@skolverket.se
Kersti.blidberg@skolverket.se

SWITZERLAND

Fritz Oser
University of Fribourg
Department of Education
Rue Faucigny 2
CH-17+Fribourg
[http://www.unifr.ch/pedg/]
Email: Fritz.Oser@unifr.ch

UNITED STATES OF AMERICA
Carole Hahn
Emory University
Division of Educational Studies
1784 Decatur Road, Suite 240
Atlanta, GA 30322
[http://nces.ed.gov/surveys/cived/]
Email: chahn@emory.edu

CIVIC EDUCATION STUDY ADVISORY COMMITTEES

The International Coordinating Center was supported by the International Steering Committee and the IEA Technical Executive Group. These committees brought several disciplinary perspectives, as well as viewpoints from different parts of the world. The Steering Committee prepared the study's overall design in the proposal, wrote and reviewed items for the instruments, planned and interpreted data analysis, led groups at meetings, and reviewed this volume in detail. The IEA Technical Executive Group, also represented at the Steering Committee through its chair, Ray Adams, provided guidance on issues related to design, sampling, instrument construction, analysis, and reporting, and ensured that the methodologies used in this study were technically sound. Heinrich Mintrop was consultant on the development and later analysis of the teacher questionnaire for the standard population (14-year-olds).

International Steering Committee

Judith Torney-Purta (Chair), United States
Ray Adams (*ex-officio*), Australia
Barbara Fratczak-Rudnicka, Poland
Barbara Malak-Minkiewicz, IEA Secretariat
Georgia Kontogiannopoulou-Polydorides, NRC Greece
Bruno Losito, NRC Italy
Ingrid Munck, Sweden
Hans Oswald, Germany
John Schwille, United States
Gita Steiner-Khamsi, United States
Lee Wing-On, NRC Hong Kong (SAR)

Technical Executive Group

Ray Adams (Chair), Australia
Pierre Foy, Canada (now IEA Data Processing Center)
Larry Hedges, United States
Eugene Johnson, United States
Michael O. Martin, United States

Consultant

Heinrich Mintrop, United States

APPENDIX B: TEST AND QUESTIONNAIRE DESIGN

B.1 Content Outline Serving as Guidelines for the IEA Civic Education Study Test/Survey Issued November 1998

The first version of these guidelines was developed in May of 1996, based primarily on the core international framing questions developed through international consultation in early 1995. The revisions made in 1997 incorporated changes made after meetings on test development held in December 1996, February 1997, March 1997, and May 1997. They also included additional quotations from Phase 1 documents.

The material enclosed in " " below is taken directly from the Phase 1 reports; other material from these documents is included in more general phraseology. The materials regarding status refer to the types of items in the database where the topic fits best.

A pre-pilot was undertaken in 1997 in 20 countries. Some item reclassifications were made on advice from the national research coordinators (NRCs) based on their pre-pilot experience. A few new items were written by NRCs at the March 1998 meeting and included in the pilot.

These guidelines are stated primarily in terms useful in the formulation of Type 1 items (knowledge). Some topics are also suitable for Column 2 skills items (which often include some text or cartoon material to which knowledge is to be applied). The guidelines have also been very useful in developing Column 3 items, Concepts of Democracy, Concepts of Citizenship, and Concepts of the Scope of Government.

Use is not made of sub-scores by content domain or sub-domain, but rather of constructed separate sub-scores for knowledge and skills (as well as a total).

B.2 Content Guidelines Organized by Domains I—III

Domain I: Democracy: *What does democracy mean and what are its associated institutions and practices (including rights and responsibilities of citizens)?*

A) What are the defining characteristics of democracy as a form of social and political organization?

　1. Recognizes defining characteristics of democracy especially in its leadership, including "representative democracy", direct or "unanimous democracy", "benevolent paternalistic democracy", and constitutional monarchy (in a generalized way, but using other terms and also in school settings). Some statements about citizens may be included here if the focus is on defining the characteristics of democracy by defining the role played by citizens, especially understanding "sovereignty of the people" and "institutionalized power given to citizens through active participation".

2. Identifies or recognizes defining characteristics and examples of limited and unlimited government (e.g., totalitarian and authoritarian regimes, including dictatorships, and how they can pose threats to democracy; "demagogues" and their threats; "different types of relations between subordinates and authorities"). Because of difficulties in translating some of these specifics, the term "undemocratic governments" or "non-democratic governments" is used. Understands how the rule of law can guard against excesses of "governments of persons".

3. Understands or evaluates significant strengths of democratic systems ("peace, order and good government", "pursuit of equality and justice", support given to "public good" or "positive evolution of public life", role of real and presumed citizen efficacy, connections with economic reform and weaknesses of democratic systems (past and present), including indecisiveness, abuses of power and corruption. [Note: some items categorized under 1 might fit here as well.]

4. Applies criteria for democracy to specific instances, problems, and scenarios; in particular, recognizes that there are sometimes gaps or conflict between the ideals of democracy and current governmental and political realities (see also ways of dealing with unresponsive authorities in section B). Recognizes how democratic processes may be both fostered and adapted for use in schools or informal group settings such as school-community connections and in the "micro-politics of everyday life" (in addition to national or local contexts). This may include appeal boards, youth movements, "pedagogical boards", action projects, community service, student organizing committees, "student governments or parliaments". Understands the "implementation of democracy in various sectors of life". [Note: writing col. 1/2 multiple-choice questions about micro-politics of schools and other associations proved difficult because of differences in the nature of these organizations across nations.]

5. Identifies situations where there are incentives to participate in political life (especially within a "dynamic view of society") or to develop one's political identity, in factors such as "political communication", thoughtful self-interest, and concerns for concrete issues such as the environment. Also identifies "factors that undermine democracy", including "fears to participate", skepticism or apathy, "wariness of challenging the status quo", "willingness to accept benevolent paternalism", "non-fulfillment or lack of responsibility among citizens to their obligations", "corruption".

6. Identifies the meaning of "political pluralism" and the role of individuals and individuals participating in civic groups or "non-governmental organizations" who hold influence or power in democracy; also the role of élites, experts or highly educated or cultured persons, business groups. [Note: this is dealt with under other categories such as IA1.]

7. Identifies or recognizes problems in transitions between types of government, especially during rapid democratization or "evolutions in public life"; changes in institutions, symbols, leaders; how institutions and citizens cope with these changes, especially when there is also a "period of transition to a market economy"; "how to build on earlier periods of democracy"; and how to deal with "transitions in schools' curricula".

B) What are important democratic institutions and practices?

1. Identifies or recognizes defining characteristics, functions, and consequences of democratic elections as ways of selecting leaders (e.g., contested elections, "free elections", elections that are "secret and regularly scheduled", role of political parties, election campaigns as periods when candidates present positions and seek support, role of media, procedures for ensuring fair or "clean" elections, "freedom of expression exemplified in a variety of parties").

2. Identifies "qualifications of candidates for governmental positions". Compares and evaluates particular programs or appeals of hypothetical candidates or parties (use proposed newspaper format or cartoons) and shows how one might make up one's mind in an election. Identifies important characteristics to look for in leaders and how to balance strengths of different candidates and connections between policy positions.

3. Identifies the ways that elected officials are viewed in their role as political authorities. Understands what it means to have a "healthy critical attitude which holds officials accountable"; also, specific ways they can be held accountable for their actions in democracy "through consultation", "media attention"; "movements for reform". Recognizes that sometimes candidates once elected fail to live up to their promises or obligations by sacrificing public interests to small group or private interests, "treating subordinates inappropriately", "financial scandals", "failing to deal with problems like unemployment", corruption, "conflicts of interest", "lack of concern for the common good in the business of politics"; "failure to establish a 2-way communication between government and people". How citizens can effectively obtain information from media and other sources and deal with "non-responsiveness of officials".

4. Identifies or recognizes "the basic characteristics of parliament, the law and judicial system", "foundation and evolution of law", "goals and tasks of the legal system", provisions for establishing laws and also regulations (the "rules and regulations of democratic life"); "importance of law to communities"; also how laws are "reinforced" (made known), interpreted and enforced (to include legislative, judicial, and executive processes); also dealing with police authority and with "members of administrative bureaucracies". Understands what "civilian control of the military" requires and why it is important and "role of police as guardians of public order".

5. Identifies the function of constitutions and the types of provisions usually included in them, in particular the "rules and regulations that permit democratic decision-making" and "principles of delegation to regional governments".

6. Understands what it means for power to be balanced or distributed among branches of the government and what function this serves; also federal versus unitary governments (national items in countries where this is the structure). [Note: this was not stressed in the majority of countries, and there is a great deal of national specificity in the meaning of balance of power; this was to be dealt with, if at all, in the national items.]

7. Understands basic economic issues and "their political implications", "recognizes the relationship between the government and the market", especially in an "information society", what it means to "shift resources from the private to the public sector", "state ownership and its implications", "how the government pays for services it provides through taxation and borrowing", "welfare state", what some of the reasons are for unemployment and the policies that may be enacted that influence it.

C) What are the rights and obligations of citizenship in democracy?

1. Identifies generally the rights and qualifications of citizens "common in and essential to" democracies, especially political or "civil rights" (e.g., "vote" and "association") and also "worship, right to fair trial". Recognizes the extent to which voting is a right and also in some countries an "obligation of citizens".

2. Identifies rights usually guaranteed nationally, especially a citizen's participation or "right to express dissenting opinions", "way that voting or speaking can fight repression", "right to associate with others" and "speak without fear". Understands "acceptable forms of protest", both conventional and unconventional, "official and unofficial", including civil disobedience without violence and "petitions, demonstrations, environmental actions, strikes". Understands "right to voice and promote their own opinions and oversee the work of political decision-makers" and "why the government needs to know people's opinions and why people have to make known their opinions and grievances". Recognizes importance of "individual motivation and persistence in exercising self-determination and participation, including criticism and civil disobedience". Understands some of the "channels for forming and expressing public opinion", "how to approach state institutions with suggestions", "productive ways of expressing disagreement with political decisions", "how to search for non-violent solutions", and "how to be an active, critical, and responsible member of society".

3. Identifies the obligations, responsibilities, or "civic duties" of citizens common in democracies, including "willingness to participate in attending to common affairs" and "defend democratic values" and to whom these obligations are referenced (e.g., "society, the government, the family, elders"). Has skill in understanding how "rights orient

obligations which lead to behavior" and "readiness for continuous analysis of information and critical thinking used in making judgments" about the citizen's rights and responsibilities. Includes as examples "being law abiding, respecting others' rights, supporting civic activities which support the common good", "civility", paying taxes, "interest in constructive democratic reform", being informed about public issues, monitoring public officials, and voting after considering all candidates carefully. Understands "how civic action sustains a democracy".

4. Understands the role of media (press) in "presenting opinions to the public", "right to freely present political opinion in an independent press", and the "right to non-censored information". Understands "what is necessary for the public to have confidence in the mass media", problems when "freedom of the media conflicts with other goals", such as anti-racism. Recognizes arguments for censorship or government press control, and why they may be appealing but should be resisted in a democracy.

5. Identifies the network of private, advisory, and voluntary associations (including those in schools) within which citizen obligations and rights may be encouraged and enacted, "right to create and join non-governmental interest-based associations" or "pressure groups which tell about the needs of ordinary citizens". Recognizes the importance of "practices exerting an influence on various causes and issues in society …, especially in local decision-making mechanisms". Recognizes the diversity of political opinion on most issues. Develops the ability to "influence the communication environment". Recognizes the special status of "environmental issues as social issues around which group and individual actions are centered".

6. Identifies the relationship between "human rights" as defined in international documents, including UN documents, such as the Universal Declaration of Human Rights and the Convention on Rights of the Child and rights as defined within the country (usually in the Constitution). Also "respect for human dignity", the need to "demand human rights for oneself and defend them for others", principle of non-discrimination in human rights, "human rights issues in immigrant societies", and "human rights during international collisions".

7. Identifies rights in the economic sphere, especially in relation to owning property, labor unions, strikes, value of competition (and problems of unfair competition), rights to social security, and "right to join a union". Recognizes that some countries support a "right to a minimum standard of living". [Note: this includes primarily ideas having to do with individuals interacting with the economic system.]

8. Demonstrates awareness of trade-offs between individual rights, the rights of others, and the good of the community (locally and on a broader level). Recognizes the existence of debates over rights such as "clean environment" or programs to aid those in poverty.

Domain II: National Identity and Regional and International Relationships: *How can the sense of national identity or national loyalty among young people be described and how does it relate to sub-national and supranational organizations?*

A) What are the characteristics of membership in a national community or society (sometimes referred to as "citizenry")?

1. Recognizes the difference between ascribed citizen status (by birth, by residence, by naturalization) versus citizenship as an acquired status (by loyalty to the state or acceptance of common norms). Recognizes terms such as "native country".

2. Recognizes the series of common memberships that are often associated with these statuses (e.g., speaking one of a selected small number of languages (and the lower status of certain languages)). "Aware of ways in which low status and lack of privilege may be associated."

3. Recognizes the importance of "solidarity", "unity", or a sense of "collective identity" with the "native country" (e.g., having pride in your country, "in special features such as natural beauty, education, culture, or science", and in being a _____; respecting the symbols of the country, such as the flag or statues of national heroes, documents, or "community traditions"; knowing national narratives about "patriots, national heroes or learned persons", interpretations of historical events such as the "founding of the country", or "struggles for national existence"; "identifying with the democratic constitution"; believing certain "values and thoughts" exemplified in common literature or culture; serving in the military; defending the country from unjust criticism; being respectful of government leaders/trusting the government; being part of a sense of collective memory. In some countries may include "filial piety extended to the government" and in others may include "conflicts between unity and diversity". Recognizes the importance of patriotism and the problems associated with excesses of national identity (e.g., "chauvinistic nationalism").

4. Recognizes that every nation has some events in its history of which it is justifiably proud. Also recognizes that every nation has some events of which it is not proud in its "social and its political history", and knows something about what those events are for their country. Recognizes how nations generally sometimes deal with these issues, including "responsibility for developing countries as compensation for colonial history".

5. Recognizes the problems associated with national identity in "different historical periods", including times of transition (e.g., in rapid democratization) or of "dual citizenship".

B) What does national identity or loyalty mean in relation to extra-national and supranational relationships and organizations?

1. Recognizes the nature of the role the country has played and continues to play in global and regional spheres of influence.

2. Recognizes countries or groups of countries past, present, or future that have been allies or enemies, and the nature of these "alliances or dependencies" or sense of "common traditions", or enmities. Recognizes "neutral or mediating roles" among nations when "peace is threatened". Understands "contributions for assuring the development of a modern nation" and "the role of international trade and everyday contact".

3. Recognizes major supranational regional structures (in most cases within Europe) or international organizations (e.g., the United Nations) or organizations "concerned with finding cooperative solutions to world issues" (e.g., Greenpeace) that may have a place in awareness, identity, or loyalty; "transforming shock and indignation over global inequalities into understanding and action".

4. Recognizes particular sub-national groups (e.g., ethnic or religious groups) that may have a place in the young person's awareness, identity, or loyalty.

Domain III: Social Cohesion and Diversity: *What is the make-up of the nation in terms of diversity? What issues of social cohesion are raised, and how are they dealt with?*

A) Which groups within the nation are seen as set apart or disenfranchised and in what ways?

1. Recognizes groups viewed as subject to discrimination in contemporary society (as defined, for example, by ethnicity, race, immigrant status, refugee status, mother tongue, social class, religion, gender).

2. Recognizes instances of past and present oppression or discrimination of certain groups in many societies and "racism as a violation against human rights"; this includes differences in opportunities for political participation or in the extent to which certain groups serve in positions of political leadership ("under-representation"). "Recognizes the importance of the schools equipping males and females for equal rights and responsibilities." Recognizes the work of leaders and citizens who have fought against oppression. Recognizes the seeking of enhanced political power by some of these groups.

3. Recognizes tension in their society and many others between perceptions of the need for social cohesion and the need to recognize and value the cultural, social, and political contributions of these groups. Recognizes ways of mediating conflict between these groups and "knows about provisions to guard rights of minorities".

4. Understands the "fundamental values of community life" and the importance of attitudes and behaviors of respect and tolerance between these groups on the part of individuals. Recognizes "world cultural plurality", "interaction between different cultures", and need for "equality of all cultural groups". "Recognizes gaps in public discourse between friendliness and hostility."

Table B.1 Domain Content Categories and Short Titles for Items in Final Test

I A: Democracy and Its Defining Characteristics

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter[1]
Identify defining characteristics of democracy	1	…necessary feature of democratic government	91	80
Identify limited and unlimited government, undemocratic regimes	6	…what makes a government non-democratic	80	106
Identify problems in transitions of government from non-democratic to democratic	14	…most convincing action to promote democracy	71	106
	28	…what direct democracy is	70	111

I B: Institutions and Practices in Democracy

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter[1]
Identify characteristics and functions of elections and parties	9	…function of periodic elections (also IEA, 1971)	54	123
	18	…meaning of coalition government	71	106
	20	…main message of cartoon about elections	60	119
Identify qualifications of candidates for positions and making up one's mind during elections	11	…which party issued political leaflet	80	97
	12	…what issuers of leaflet think about taxes	83	91
	13	…which policy issuers of leaflet likely to favor	75	100
Identify a healthy critical attitude toward officials and their accountability	30	…example of corruption in national minister	80	100
Identify basic character of parliament, judicial system, law, police	17	…part of legislative branch	68	112
Identify provisions of constitution	16	…what countries' constitutions contain	86	99

Table B.1 (continued)

I C: Citizenship: Rights and Duties

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter[1]
Identify general rights, qualifications, and obligations of citizens in democracies	8	…a political right	90	85
Identify citizens' rights to participate	5	…violation of civil liberties in democracy (also IEA 1971)	72	107
Understand the role of mass media in democracy	7	…result if large publisher buys many newspapers	75	103
	43	…influence of media	75	101
Identify network of associations and differences of political opinion	3	…why organizations are important in democracy	85	93
Identify the human rights defined in international documents	2	… purpose of Universal Declaration of Human Rights	92	86
Identify rights in the economic sphere	4	…purpose of labor unions	86	98

Table B.1 (continued)

II A: National Identity

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter[1]
Recognize sense of collective identity source of conflict	25	...national or ethnic difference as the source of conflict	68	110
Recognize that every nation has events in its history of which it is not proud	10	...main message of cartoon about history textbooks	67	110

II B: International Relations

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter[1]
Recognize major inter-governmental organizations and understand diplomatic relations between countries	23	...reasons for recalling an ambassador	67	113

III A: Social Cohesion and Diversity

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter[1]
Recognize groups subject to discrimination	15	...an example of discrimination in pay equity	68	108
	29	...main message of a cartoon about excluding women candidates	71	110
	31	...persistence of gender discrimination at work	70	109

1 The reported item parameters are based on the joint scale of civic knowledge (which includes the sub-dimensions content knowledge and interpretative skills). The item parameters mean that students with that level of proficiency had a 65 percent probability of providing a correct response.

Table B.2 Domain Content Categories and Short Titles for Items in Final Test

I A: Democracy and Its Defining Characteristics

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter
Identify defining characteristics of democracy	12	...who ought to govern in democracy (also IEA, 1971)	71	88
	19	...necessary feature of democratic government	65	96
Identify limited and unlimited government, undemocratic regimes	17	...what makes a government non-democratic	53	106
Evaluate strengths and weaknesses of democratic systems	14	...main message of cartoon about democracy	61	100
Identify incentives to participate in the form of factors undermining democracy	9	...most serious threat to democracy	72	90
Identify problems in transitions of government from non-democratic to democratic	29	...most convincing action to promote democracy	54	106

Table B.2 (continued)

I B: Institutions and Practices in Democracy

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter
Identify characteristics and functions of elections and parties	11	...function of having more than one political party	75	88
	22	...function of periodic elections (also IEA, 1971)	42	113
Identify qualifications of candidates for positions and making up one's mind during elections	23	...which party issued political leaflet	65	97
	24	...what issuers of leaflet think about taxes	71	91
	25	...which policy issuers of leaflet likely to favor	58	100
Identify a healthy critical attitude toward officials and their accountability	30	...example of corruption in national legislature	66	96
	33	...main message of cartoon about political leader	77	84
Identify basic character of parliament, judicial system, law, police	2	...an accurate statement about laws	78	84
	13	...main task of national legislature	67	94
Identify provisions of constitution	28	...what countries' constitutions contain	62	99
Understand basic economic issues and their political implications	27	...essential characteristic of market economy	47	110
	38	...a fact (not an opinion) about taxes	49	109

Table B.2 (continued)

I C: Citizenship: Rights and Duties

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter
Identify general rights, qualifications, and obligations of citizens in democracies	3	...a political right	78	85
Identify citizens' rights to participate and express criticism and their limits	10	...illegal activity for a political organization	59	101
	15	...violation of civil liberties in democracy (also IEA, 1971)	53	107
Identify obligations, civic duties of citizens in democracy	1	...role of citizen in democratic country	79	83
Understand the role of mass media in democracy	4	...which of a reporter's rights was violated	70	92
	18	...result if large publisher buys many newspapers	57	103
Identify network of associations and differences of political opinion	7	...why organizations are important in democracy	69	93
	34	...main point of article about factory being shut	35	121
Identify the human rights defined in international documents	6	... purpose of Universal Declaration of Human Rights	77	86
	20	...what is in Convention on Rights of the Child	77	84
Identify rights in the economic sphere	8	...purpose of labor unions	64	98
Demonstrate awareness of tradeoffs	35	...economic objections to factory being shut	67	93

II A: National Identity

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter
Recognize sense of collective identity	32	...an opinion (not a fact) about flags	66	95
Recognize that every nation has events in its history of which it is not proud	36	...main message of cartoon about history textbooks	58	102

Table B.2 (continued)

II B: International Relations

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter
Recognize international economic issues and organizations (other than inter-governmental) active in dealing with matters with economic implications	21	...who owns multinational businesses	47	110
	31	...an opinion (not a fact) about the environment	53	106
Recognize major inter-governmental organizations	16	...major purpose of United Nations (also IEA, 1971)	85	77

III A: Social Cohesion and Diversity

Domain Content Category	Item #	Short Titles for Items	Correct Answers (in %)	Item Parameter
Recognize groups subject to discrimination	5	...an example of discrimination in employment	65	97
	26	...an example of discrimination in pay equity	50	108
	37	...a fact (not an opinion) about women and politics	72	89

APPENDIX C: TEST AND OPERATION FORMS

Form 1:   Translation Deviation Form

Form 2:   Translation Verification Report Form

Form 3:   Subject Allocation Grid

Form 4:   Class Sampling Form

Form 5:   Teacher-Student Linkage Form

Form 6:   Test Administration Form

Form 7:   Student Tracking Form

Form 8:   Teacher Tracking Form

Form 1: Translation Deviation Form

To be completed by the National Centre. To be completed for each item set, if applicable.

CIVIC Participant:

(a) Part: _____

(b) Questionnaire:
☐ Student
☐ Teacher
☐ School

(c) Translator Name and Contact

_____

_____

_____

_____

Page Number	Question Number	Change/ Vocabulary	Change/ Content	Other translation problems	(5) Description of Deviation
(1)	(2)	(3)	(4)	(5)	

Form 2: Translation Verification Report Form

Country _____

Language _____

Translator´s Name _____

Instrument: _____ Student Questionnaire _____ Teacher Questionnaire

_____ School Questionnaire

(1) Page Number	(2) Question Number	(3) Severity Code	(4) Description of Deviation	(5) Suggested Change

Form 3: Subject Allocation Grid

Country:					
	Subject:	Subject:	Subject:	Subject:	Subject:
Item 1					
Item 2					
Item 3					
Item 4					
Item 5					
Item 6					
Item 7					
Item 8					
Item 9					
Item 10					
Item 11					
Item 12					
Item 13					
Item 14					
Item 15					
Item 16					
Item 17					
Item 18					
Item 19					
Item 20					
Item 21					
Item 22					
Item 23					
Item 24					
Item 25					
Item 26					
Item 27					
Item 28					
Item 29					
Item 30					
Item 31					
Item 32					
Item 33					
Item 34					
Item 35					
Item 36					
Item 37					
Item 38					
Sum of ticks					

Form 4: Class Sampling Form

CIVIC Participant: _____ Stratum: _____

School Name: _____

[a] School ID	[b] Grade	[c]Minimum Cluster Size	[d] Random Start	[e] Sampling Interval*	[f] First Selected Class	[g] Second Selected Class*
			0 ___ . ___	___ . ___		

(1) Class ID	(2) Class Name	(3) Number of Students	(4) Line Number	(5) Sampled

* Necessary only if sampling more than one class per grade.

Form 5: Teacher-Student Linkage Form

CIVIC Participant: _____ Stratum: _____

School Name: _____

[a] School ID	[b] Grade	[c] Class ID	[d] Class name

(1)	(2)	(3)	(4)	(5)		
Student Name or Number	Excl. Code	Date of Birth (mmyy)	Sex	Subject 1: Teacher: Class Name: Link No:	Subject 2: Teacher: Class Name: Link No:	Subject 3: Teacher: Class Name: Link No

Form 6: Test Administration Form

Complete one form for each testing session

1. School identification code:

2. School name:

3. Class name:

4. School Coordinator:

5. Test Administrator:

6. Test Administrator's position:

   [   ]-National Center Staff

   [   ]-Teacher from school but not teacher of the selected class

   [   ]-Teacher of the selected class

   [   ]-Other, please describe:

7. Type of testing session: [   ]-Regular   [   ]-Makeup

8. Date of testing:

Scheduled starting time:

Actual schedule of the testing sessions		
Start time	End time	
10a	10b	Administration Session 1 (preparation of students, reading of instructions, distribution of student envelopes, etc.)
11a	11b	Session 1 (Part 1 of the *Student Questionnaire*)
12a	12b	Session 1 (Part 2 of the *Student Questionnaire*)
13a	13b	Administration Session 2 (preparation of students, redistribution of the *Student Questionnaires* etc.)
14a	14b	Session 2. If the *Student Questionnaire* (Session 2) is administered on a different date, indicate the date here:

15. Where there any special circumstances or unusual events during the session?

[   ]-No   [   ]-Yes - Explain:

16. Did students have any particular problems with the testing (for example, tests too difficult, not enough time provided, tiring, confusing)?

[   ]-No   [   ]-Yes - Explain:

17. Were there any problems with the testing materials (for example, errors or omissions in the Student Tracking Forms, incorrect test booklet assignments, insufficient booklets)?

[   ]-No   [   ]-Yes - Explain:

Form 7: Student Tracking Form

CIVIC Participant: _____   Stratum: _____

School Name: _____

[a] School ID	[b] Class ID	[c] Class Name	[d] Grade

(1) Student Name or Number	(2) Student ID	(3) Excl.	(4) Date of Birth (MMYY)	(5) Sex	Participation Status			
					(6) Session		(7) Make-up	
					1	2	1	2

Form 8: Teacher Tracking Form

CIVIC Participant: _____

School ID: _____

Stratum: _____

School Name: _____

(1) Teacher Name	(2) Teacher ID*	(3) Link No.	(4) Subject	(5) Grade	(6 Class Name	(7) Sampled Class ID	(8) Eligible Students	(9) Questionnaire	(10) Participation Status

* A teacher should be listed once for each subject he or she teaches to the selected students, each time with the same Teacher ID, but a different Link No.

APPENDIX D: SAMPLE CHARACTERISTICS

D.1 CIVED STANDARD POPULATION (14-YEAR-OLDS)

Introduction

The following material describes the target population definition and necessary coverage and exclusions. It also shows the sample design for each country within the standard population participating in the IEA Civic Education Study (CivEd).

Australia

Target Population
In Australia, the national desired target population was Grade 9.

Coverage and Exclusions
Exclusions within the national defined target population consisted of very small schools, distant schools, and special schools for the functionally and mentally disabled with atypical educational systems.

Sample Design
- Explicit stratification was organized by states and territories and school type, for a total of 20 explicit strata.
- No implicit stratification was used.

Table D.1 Allocation of school sample in Australia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
NSW Government schools	32	0	26	5	1
NSW Catholic schools	11	0	5	3	3
NSW Independent schools	5	0	2	3	0
VIC Government schools	22	0	17	4	1
VIC Catholic schools	8	0	5	1	2
VIC Independent schools	5	0	5	0	0
QLD Government schools	20	0	17	3	0
QLD Catholic schools	5	0	5	0	0
QLD Independent schools	5	0	4	1	0
SA Government schools	8	0	7	1	0
SA Catholic schools	2	0	2	0	0
SA Independent schools	2	0	2	0	0
WA Government schools	11	0	7	3	1
WA Catholic schools	3	0	2	1	0
WA Independent schools	2	0	2	0	0
TAS Government schools	3	0	3	0	0
TAS Non-government schools	1	0	0	1	0
NT All schools	2	0	1	1	0
ACT Government schools	2	0	2	0	0
ACT Non-government schools	1	0	1	0	0
Total	**150**	**0**	**115**	**27**	**8**

Belgium (French)

Target Population

In Belgium, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems.

Sample Design

- No explicit stratification was used.

- Implicit stratification was organized by type of education (academic and vocational).

Table D.2 Allocation of school sample in Belgium (French)

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Belgium (French)	150	0	84	28	38
Total	**150**	**0**	**84**	**28**	**38**

Bulgaria   Target Population
In Bulgaria, the national desired target population was Grade 8.

Coverage and Exclusions
School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems, schools for students with criminal behavior, and very small schools (MOS<9).

Sample Design
- Explicit stratification was organized by school size (large and small schools), for a total of two explicit strata.
- No implicit stratification was used.
- Schools in the "small schools" stratum selected with equal probabilities.

Table D.3 Allocation of school sample in Bulgaria

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Large schools	150	0	130	0	20
Small schools	22	0	18	0	4
Total	**172**	**0**	**148**	**0**	**24**

Chile

Target Population

In Chile, the national desired target population was Grade 8.

Coverage and Exclusions

Exclusions within the national defined target population considered two different criteria: distant schools and small size schools (urban and rural schools).

Sample Design

• Explicit stratification was organized by school type (municipal, subsidized, and private paid), for a total of three explicit strata.

• Implicit stratification was organized by urbanization (urban and rural schools) and geographical area (North, Metropolitan, and South), for a total of six implicit strata.

Table D.4 Allocation of school sample in Chile

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Municipal schools	60	0	59	1	0
Subsidized schools	60	0	59	1	0
Private paid schools	60	0	60	0	0
Total	**180**	**0**	**178**	**2**	**0**

Colombia

Target Population

In Colombia, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of special schools for the physically and mentally disabled and very small schools.

Sample Design

- Explicit stratification was organized by regions (Region Atlántica, Region Bogotá, Region Occidental, Region Centro-Oriente and Orinoquía-Amazonía) and school type (public and private), for a total of 10 explicit strata.

- Implicit stratification was organized by the population of the municipality and school performance, for a total of 16 implicit strata.

Table D.5 Allocation of school sample in Colombia

Explicit Stratum		Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
				Sampled	Replacement	
Region Atlántica	Public	18	0	13	3	2
	Private	16	0	10	2	4
Region Bogotá	Public	17	0	9	6	2
	Private	16	0	15	1	0
Region Occidental	Public	16	0	10	6	0
	Private	16	0	11	4	1
Region Centro-Oriente	Public	16	0	13	3	0
	Private	16	0	12	2	2
Region Orinoquía-Amazonía	Public	16	0	9	4	3
	Private	16	0	11	3	2
Total		**163**	**0**	**113**	**34**	**16**

Cyprus

Target Population

In Cyprus, the national desired target population was Grade 8.

Coverage and Exclusions

All existing 61 schools were included and sampled.

Sample Design

- Explicit stratification was organized by districts.

- Implicit stratification was organized by school type (public and private), for a total of two implicit strata.

Table D.6 Allocation of school sample in Cyprus

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Cyprus	61	0	61	0	0
Total	**61**	**0**	**61**	**0**	**0**

Czech Republic

Target Population

In the Czech Republic, the national desired target population was Grade 8.

Coverage and Exclusions

Exclusions within the national defined target population were special schools for the functionally and mentally disabled with atypical educational systems and small schools.

Sample Design

- Explicit stratification was organized by school type, for a total of two explicit strata.

- Implicit stratification was organized by urbanization, for a total of five implicit strata.

Table D.7 Allocation of school sample in Czech Republic

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Basic schools	90	0	82	7	1
Gymnasia	60	0	54	5	1
Total	**150**	**0**	**136**	**12**	**2**

Denmark **Target Population**

In Denmark, the national desired target population was Grade 8 and Grade 9 as a national option.

Coverage and Exclusions

School-level exclusions consisted of special schools for functionally and mentally disabled with atypical educational systems.

Sample Design

- Explicit stratification was not used.
- Implicit stratification was organized by school type (public and private).

Table D.8 Allocation of school sample in Denmark

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Denmark	250	0	178	0	72
Total	**250**	**0**	**178**	**0**	**72**

England

Target Population

In England, the national desired target population was Grade 10.

Coverage and Exclusions

Exclusions within the national defined target population were special schools for the functionally and mentally disabled with atypical educational systems and small schools.

Sample Design

- No explicit stratification was used.
- Implicit stratification was organized by school performance and school type, for a total of 30 implicit strata.

Table D.9 Allocation of school sample in England

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
England	150	0	80	48	22
Total	**150**	**0**	**80**	**48**	**22**

Estonia

Target Population

In Estonia, the national desired target population was Grade 8.

Coverage and Exclusions

Exclusions within the national defined target population were special schools for the functionally and mentally disabled with atypical educational systems and small schools.

Sample Design

- Explicit stratification was organized by regions and small schools, for a total of seven explicit strata.

- Implicit stratification was organized by urbanization and language of instruction, for a total of six implicit strata.

Table D.10 Allocation of school sample in Estonia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Central Estonia	21	0	18	0	3
North-East Estonia	27	0	23	0	4
Southern Estonia	29	0	24	1	4
Western Estonia	19	0	16	0	3
Northern Estonia	48	0	40	1	7
Small schools	8	0	7	0	1
Large schools	18	0	15	0	3
Total	**170**	**0**	**143**	**2**	**25**

Finland

Target Population

In Finland, the national desired target population was Grade 8.

Coverage and Exclusions

Exclusions within the national defined target population consisted of the autonomous province of Ahvenanmaa, Rudolf Steiner schools, special schools for the functionally and mentally disabled with atypical educational systems, small schools, and language schools.

Sample Design

- Explicit stratification was organized by major regions (Uusimaa, Southern Finland, Eastern Finland, Mid Finland and Northern Finland), for a total of five explicit strata.

- Implicit stratification was organized by municipalities (urban, semi urban, and rural), for a total of three implicit strata.

Table D.11 Allocation of school sample in Finland

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Uusimaa	30	0	26	2	2
Southern Finland	30	0	28	2	0
Eastern Finland	30	0	28	0	2
Mid Finland	30	0	29	1	0
Northern Finland	30	0	30	0	0
Total	**150**	**0**	**141**	**5**	**4**

Target Population
In Germany, the national desired target population was Grade 8.

Coverage and Exclusions
Exclusions within the national defined target population were schools that follow a different curriculum and special schools for the functionally and mentally disabled with atypical educational systems. Three federal states ("Bundeslaender"), Bremen, Hessen, and Niedersachsen) refused to participate in the study, and one federal state (Baden-Wuerttemberg) did not permit testing in high schools ("Gymnasium").

Sample Design
- Explicit stratification was organized by federal states (13 federal states) and school type (Gymnasium, Gesamtschule, Realschule, Hauptschule, and other schools including Regelschule, Mittelschule, Regionalschule, and Sekundarschule), for a total of 48 explicit strata.

Table D.12 Allocation of school sample in Germany

Explicit Stratum*	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
BB Gymnasium	2	0	1	1	0
BB Gesamtschule	6	0	4	2	0
BB Realschule	1	0	1	0	0
BE Gymnasium	2	0	1	1	0
BE Gesamtschule	3	0	1	1	1
BE Realschule	2	0	1	1	0
BE Hauptschule	4	0	2	2	0
BW Realschule	6	0	3	1	2
BW Hauptschule	7	0	6	1	0
BY Gymnasium	6	0	3	3	0
BY Gesamtschule	1	0	1	0	0
BY Realschule	7	0	4	3	0
BY Hauptschule	9	0	4	4	1
HH Gymnasium	1	0	1	0	0
HH Gesamtschule	2	0	1	0	1
HH Hauptschule	4	0	1	1	2
MV Gymnasium	2	0	2	0	0
MV Gesamtschule	1	0	0	1	0
MV Realschule	2	0	2	0	0
MV Hauptschule	4	0	3	1	0
MV Other schools**	2	0	0	2	0
NW Gymnasium	11	0	7	4	0
NW Gesamtschule	9	0	3	4	2
NW Realschule	9	0	6	2	1
NW Hauptschule	9	0	7	2	0

Notes: *BB=Brandenburg, BE=Berlin, BW=Baden-Wuerttemberg, BY=Bayern, HH= Hamburg, MV=Mecklenburg-Vorpommern, NW=Nordrhein-Westfalen, RP=Rheinland-Pfalz.

"Other schools" are schools with different study programs like Regelschule, Mittelschule, Regionalschule, or Sekundarschule.

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
RP Gymnasium	2	0	2	0	0
RP Gesamtschule	1	0	0	1	0
RP Realschule	2	0	2	0	0
RP Hauptschule	3	0	2	0	1
SH Gymnasium	1	0	1	0	0
SH Gesamtschule	2	0	0	1	1
SH Realschule	2	0	2	0	0
SH Hauptschule	2	0	2	0	0
SL Gymnasium	1	0	1	0	0
SL Gesamtschule	1	0	1	0	0
SL Realschule	1	0	1	0	0
SL Hauptschule	4	0	2	2	0
SL Other schools**	2	0	0	2	0
SN Gymnasium	4	0	1	3	0
SN Other schools**	18	0	10	8	0
ST Gymnasium	2	0	2	0	0
ST Gesamtschule	1	0	0	1	0
ST Realschule	3	0	2	1	0
ST Hauptschule	4	0	3	1	0
ST Other schools**	1	0	1	0	0
TH Gymnasium	2	0	1	1	0
TH Gesamtschule	1	0	0	1	0
TH Other schools**	10	0	6	3	1
Total	**182**	**0**	**107**	**62**	**13**

Notes: *SH=Schleswig-Holstein, SL=Saarland, SN=Sachsen, ST=Sachsen-Anhalt, TH=Thueringen.

** "Other schools" are schools with different study programs like Regelschule, Mittelschule, Regionalschule, or Sekundarschule.

Greece

Target Population

In Greece, the national desired target population was Grade 9.

Coverage and Exclusions

Exclusions within the national defined target population were schools delivering special education for the functionally and mentally disabled with atypical educational systems, very small schools, and non-native-language-speaking schools.

Sample Design

- Explicit stratification was organized by regions, for a total of 12 explicit strata.
- Implicit stratification was organized by school type (public and private) and by urbanization.

Table D.13 Allocation of school sample in Greece

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
East Macedon-Thrace	8	0	7	0	1
Central Macedonia	24	0	22	1	1
West Macedonia	5	0	5	0	0
Epirus	6	0	6	0	0
Thessaly	11	0	11	0	0
Ionian Islands	3	0	2	0	1
Western Greece	13	0	13	0	0
Sterea	7	0	7	0	0
Attiki	54	0	43	5	6
Peloponnese	6	0	5	1	0
Aegean Islands	7	0	5	0	2
Crete	9	0	9	0	0
Total	**153**	**0**	**135**	**7**	**11**

Hong Kong (SAR)

Target Population

In Hong Kong (SAR), the national desired target population was Grade 9.

Coverage and Exclusions

Exclusions within the national defined target population were schools delivering special education for the functionally and mentally disabled with atypical educational systems and non-native-language-speaking schools.

Sample Design

- No explicit stratification was used.

- Implicit stratification was organized by financing mode, for a total of two implicit strata.

Table D.14 Allocation of school sample in Hong Kong (SAR)

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Hong Kong (SAR)	150	0	137	13	0
Total	**150**	**0**	**137**	**13**	**0**

Hungary

Target Population

In Hungary, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of special schools for the disabled and very small schools (MOS<10.)

Sample Design

- No explicit stratification was used.
- Implicit stratification was organized by urbanization and school size.

Table D.15 Allocation of school sample in Hungary

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Hungary	150	0	146	0	4
Total	**150**	**0**	**146**	**0**	**4**

Italy   Target Population
In Italy, the national desired target population was Grade 8 (population 2a) and Grade 9 (population 2b).

Coverage and Exclusions
School-level exclusions consisted of non-government middle schools, special schools for the functionally and mentally disabled with atypical educational systems, and non-native-language- speaking schools.

Sample Design
- No explicit stratification was used (for both populations).
- For Grade 8, implicit stratification was organized by region and by type of municipality, for a total of 40 implicit strata. Implicit stratification for Grade 9 was organized by region and by school category (classical education, artistic education, technical education, and vocational education), for a total of 80 implicit strata.

Table D.16 Allocation of school sample in Italy

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Italy	172	0	161	11	0
Total	**172**	**0**	**161**	**11**	**0**

Latvia

Target Population

In Latvia, the national desired target population was Grade 8.

Coverage and Exclusions

Coverage in Latvia included students whose language of instruction is Latvian or Russian. School-level exclusions consisted of very small schools and special schools for the functionally and mentally disabled with atypical educational systems.

Sample Design

- Explicit stratification was organized by school size and language of instruction (Latvian/Russian), for a total of five explicit strata.

- Implicit stratification was organized by urbanization and districts, for a total of 10 implicit strata.

- Schools in the Latvian- and Russian-speaking "very large schools" and the "small rural Latvian-speaking schools" strata were selected with equal probability.

Table D.17 Allocation of school sample in Latvia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Very large Latvian-speaking schools	11	0	11	0	0
Large Latvian-speaking schools	52	0	49	1	2
Small rural Latvian-speaking schools	12	0	12	0	0
Very large Russian-speaking schools	9	0	7	0	2
Other Russian-speaking schools	66	0	49	1	16
Total	**150**	**0**	**128**	**2**	**20**

Lithuania

Target Population

In Lithuania, the national desired target population was Grade 9.

Coverage and Exclusions

Coverage in Lithuania included students whose language of instruction was Lithuanian or , Russian, or Polish. School-level exclusions consisted of special schools and very small schools (MOS<7).

Sample Design

- Explicit stratification was organized by school size and language of instruction (Latvian/Russian/Polish).

- Implicit stratification was organized by school type (basic and secondary), for a total of four implicit strata.

Table D.18 Allocation of school sample in Lithuania

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Lithuanian-speaking large schools	133	0	127	3	3
Lithuanian-speaking small schools	17	0	16	0	1
Polish-speaking large schools	4	0	3	0	1
Polish-speaking small schools	2	0	2	0	0
Russian-speaking large schools	16	0	12	4	0
Russian-speaking small schools	2	0	2	0	0
Total	**174**	**0**	**162**	**7**	**5**

Norway

Target Population

In Norway, the national desired target population was Grade 9.

Coverage and Exclusions

School-level exclusions consisted of schools delivering special education, schools for the functionally and mentally disabled, schools following different curricula, and non-native-language-speaking schools.

Sample Design

• No explicit stratification was used.

• Implicit stratification was organized by geographical regions, for a total of six implicit strata.

Table D.19 Allocation of school sample in Norway

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Norway	200	0	150	4	46
Total	**200**	**0**	**150**	**4**	**46**

Poland

Target Population

In Poland, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of schools delivering special education for the functionally and mentally disabled with atypical educational systems.

Sample Design

• No explicit stratification was used.

• No implicit stratification was used.

Table D.20 Allocation of school sample in Poland

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Poland	200	0	163	16	21
Total	**200**	**0**	**163**	**16**	**21**

Target Population

In Portugal, the national desired target population was Grade 9.

Coverage and Exclusions

School-level exclusions consisted of special schools for functionally and mentally disabled and small schools.

Sample Design

- Explicit stratification was organized by region, for a total of seven explicit strata.

- Implicit stratification was organized by school size.

Table D.21 Allocation of school sample in Portugal

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Norte	55	0	54	0	1
Centro	27	0	26	1	0
Lisboa e Vale Do Tejo	47	0	46	1	0
Alentejo	8	0	8	0	0
Algarve	5	0	5	0	0
Madeira	4	0	4	0	0
Azores	4	0	4	0	0
Total	**150**	**0**	**147**	**2**	**1**

Romania

Target Population

In Romania, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of special schools for the physically and mentally disabled and small schools (MOS<8).

Sample Design

- Explicit stratification was organized by school size, for a total of two explicit strata.
- Implicit stratification was organized by urbanization (rural and urban), for a total of three implicit strata.

Table D.22 Allocation of school sample in Romania

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Large schools	125	0	121	0	4
Small rural schools	25	0	25	0	0
Total	**150**	**0**	**146**	**0**	**4**

Russian Federation

Target Population
In the Russian Federation, the national desired target population was Grade 8.

Coverage and Exclusions
School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems and non-native-language-speaking schools.

Sample Design
- Preliminary sampling was of 45 regions from a list of 89 regions. Nineteen regions were large enough to be sampled with certainty and are marked with (°) in Table D.23.

- No explicit stratification was used. The explicit strata shown in Table D.23 are the 45 sampled regions.

- Implicit stratification was organized by school size (small and large schools) and by urbanization (village, settlement, small town, middle town, large town, and metropolis), for large schools only

- Four schools were sampled per region. More schools were sampled in some certainty regions.

Table D.23 Allocation of school sample in the Russian Federation

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Bashkortostan °	4	0	4	0	0
Kabardino-Balkaria	4	0	4	0	0
Kalmykia	4	0	4	0	0
Marii Al	4	0	4	0	0
Tataria	4	0	4	0	0
Udmuttia	4	0	4	0	0
Krasnodar Kr.°	6	0	6	0	0
Altay Kr.°	4	0	4	0	0
Krasnoyarsk Kr.°	4	0	4	0	0
Primor Kr.	4	0	4	0	0
Stavropol Kr.°	4	0	4	0	0
Habarovsk Kr.	4	0	4	0	0
Belgorod Obl.	4	0	4	0	0
Vladimir Obl.	4	0	4	0	0
Volgograd Obl.°	4	0	3	1	0
Vologda Obl.	4	0	4	0	0
Ust Orda Ok. & Irkutsk Obl.°	4	0	4	0	0
Kemerovo Obl.°	4	0	3	0	1
Kirov Obl.	4	0	3	0	1
Leningrad Obl.	4	0	4	0	0
Moscow Obl.°	6	0	6	0	0
Murmansk Obl.	4	0	4	0	0
N. Novgorod Obl.°	4	0	4	0	0

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Novgorod Obl.	4	0	4	0	0
Omsk Obl.°	4	0	4	0	0
Novosibirsk Obl.	4	0	4	0	0
Orenburg Obl.	4	0	4	0	0
Orel Obl.	4	0	4	0	0
Komi Perm Ok. & Perm Obl.°	4	0	3	1	0
Rostov Obl.°	4	0	4	0	0
Rasan Obl.	4	0	4	0	0
Samara Obl.°	4	0	4	0	0
Saratov Obl.°	4	0	4	0	0
Sahalin Obl.	4	0	3	0	1
Sverdlovsk Obl.°	6	0	6	0	0
Smolensk Obl.	4	0	4	0	0
Tambov Obl.	4	0	4	0	0
Tver Obl.	4	0	4	0	0
Tomsk Obl.	4	0	3	1	0
Ulianovsk Obl.	4	0	4	0	0
Chelyabinsk Obl.°	4	0	3	0	1
Chita Obl.	4	0	3	0	1
Moscow°	8	0	8	0	0
Sankt Petersburg°	4	0	4	0	0
Khanty Mansi Ok.	4	0	4	0	0
Total	**190**	**0**	**182**	**3**	**5**

Slovak Republic

Target Population
In Slovakia, the national desired target population was Grade 8.

Coverage and Exclusions
School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems, non-native-language-speaking schools, and very small schools (MOS<13).

Sample Design
- Explicit stratification was organized by school size (very large gymnasiums and large gymnasiums) and by school level (basic school and gymnasium), for a total of three explicit strata.
- Implicit stratification was organized by regions and school type (private and other), for a total of 11 implicit strata.
- "Very large gymnasia" were selected with equal probabilities.
- The large-school sample size in the two gymnasia strata produced estimates by school level.

Table D.24 Allocation of school sample in the Slovak Republic

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Very large gymnasia	2	0	2	0	0
Large gymnasia	28	0	21	5	2
Basic schools	120	0	94	23	3
Total	**150**	**0**	**117**	**28**	**5**

Slovenia

Target Population
In Slovenia, the national desired target population was Grade 8.

Coverage and Exclusions
School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems and non-native-language-speaking schools.

Sample Design
- Explicit stratification was organized by school size (very large schools and large schools), for a total of two explicit strata.
- Implicit stratification was organized by urbanization (five levels), for a total of six implicit strata.

Table D.25 Allocation of school sample in Slovenia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Large schools	148	0	139	8	1
Very large schools	2	0	2	0	0
Total	**150**	**0**	**141**	**8**	**1**

Sweden

Target Population

In Sweden, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems, small schools, and schools with students who had been in Swedish schools less than one year.

Sample Design

- Explicit stratification was organized by major regions, immigrant status of school, and private schools, for a total of seven explicit strata.

- Implicit stratification was organized by school size.

Table D.26 Allocation of school sample in Sweden

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Big cities	25	0	22	0	3
Suburbs	25	0	24	1	0
Other cities	25	0	24	0	1
Rural and other areas	25	0	22	0	3
Schools with many immigrants	25	0	23	0	2
Private schools	21	0	16	2	3
Large private schools	4	0	4	0	0
Total	**150**	**0**	**135**	**3**	**12**

Switzerland   Target Population

In Switzerland, the national desired target population was Grade 8.

Coverage and Exclusions

School-level exclusions consisted of very small schools.

Sample Design

- There was no explicit stratification.
- Implicit stratification was organized by languages and cantons.

Table D.27 Allocation of school sample in Switzerland

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Switzerland	181	0	128	29	24
Total	**181**	**0**	**128**	**29**	**24**

United States of America

Target Population

In the United States, the national desired target population was Grade 9.

Coverage and Exclusions

School-level exclusions consisted of small schools in the territories.

Sample Design

- Preliminary sampling was made of 52 primary sampling units (PSUs) from a list of 1,027 PSUs; 10 PSUs were large enough to be sampled with certainty.

- These PSUs were developed with the following general properties:
    - for metropolitan areas, the PSUs consisted of cities and their surrounding suburban areas
    - for non-metropolitan areas, the PSUs consisted of single counties or groups of counties
    - the PSUs were always contiguous geographic areas.

  Thus, special explicit stratification was applied to the USA design by school type and PSU size.

- Implicit stratification was organized by religious denomination and PSU within the private schools and by PSU and minority status within the public schools.

Table D.28 Allocation of school sample in the United States of America

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Private—Certainty PSUs	4	0	3	0	1
Private—Small PSUs	7	0	5	1	1
Public—Certainty PSUs	42	0	21	12	9
Public—Large PSUs	18	0	13	1	4
Public—Small PSUs (Metro)	51	0	28	12	11
Public—Small PSUs (Non-Metro)	28	0	26	2	0
Total	**150**	**0**	**96**	**28**	**26**

D.2 CIVED OLDER POPULATION (UPPER SECONDARY STUDENTS)

Introduction

The following material describes the target population definition and necessary coverage and exclusions. This material shows, as well, the sample design for each country within the older population participating in the IEA Civic Education Study (CivEd).

Chile

Target Population
In Chile, the national desired target population was Grade 12 for all regions within the country.

Coverage and Exclusions
School-level exclusions consisted of special schools for the physically and mentally disabled with atypical educational systems and schools where students were unable to speak or to read the Spanish language.

Sample Design
- Explicit stratification by administrative relationship with the state (municipal, subsidized, and private) and by school track (scientific-humanistic, technical-professional, and paid) gave a total of five explicit strata.
- Implicit stratification was organized by region (north, south, and metropolitan area), for a total of three implicit strata.

Table D.29 Allocation of school sample in Chile

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Municipal scientific-humanistic	30	0	30	0	0
Municipal technical-professional	30	0	30	0	0
Subsidized scientific-humanistic	30	0	30	0	0
Subsidized technical-professional	30	0	30	0	0
Private paid schools	60	0	57	3	0
Total	**180**	**0**	**177**	**3**	**0**

Colombia

Target Population

In Colombia, the national desired target population was Grade 11. Note that the information provided on sampling was insufficient for adjudication and weighting.

Coverage and Exclusions

School-level exclusions consisted of very small schools (MCS/2).

Sample Design

• Explicit stratification was used by regions (Atlántica, Bogotá, Occidente, Centro Oriente, Orinoquía-Amazonía) and type of schools (private or public), for a total of five explicit strata.

• Implicit stratification was organized by the population of the municipality into four different levels and by the average score of schools in social studies—Colombia's national examination (Grade 11)—in four levels as well, for a total of 16 implicit strata.

Cyprus

Target Population

In Cyprus, the national desired target population was Grade 11.

Coverage and Exclusions

All existing schools were included and sampled.

Sample Design

- Explicit stratification was used by urbanization (rural and urban schools) and regions (Nicosia, Limassol, Larfam, and Paphos).
- Implicit stratification was not used.

Table D.30 Allocation of school sample in Cyprus

Explicit Stratum		Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
				Sampled	Replacement	
Cyprus	Nicosia (urban)	11	0	11	0	0
Cyprus	Nicosia (rural)	2	0	2	0	0
Cyprus	Limassol (urban)	10	0	10	0	0
Cyprus	Limassol (rural)	3	0	3	0	0
Cyprus	Larfarm (urban)	5	0	5	0	0
Cyprus	Larfarm (rural)	4	0	4	0	0
Cyprus	Paphos (urban)	4	0	4	0	0
Cyprus	Paphos (rural)	3	0	2	0	1
Total		**42**	**0**	**41**	**0**	**1**

Czech Republic

Target Population
Within the Czech Republic, the national desired target population applied was Grade 12.

Coverage and Exclusions
School-level exclusions consisted of very small schools, special schools for the functionally and mentally disabled with atypical educational systems, and Polish language schools.

Sample Design
- Explicit stratification was organized by school track (academic, technical, and vocational), for a total of three explicit strata.

- No implicit stratification was used.

Table D.31 Allocation of school sample in Czech Republic

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Gymnasia	30	0	29	0	1
Technical schools	60	0	58	1	1
Vocational schools	60	0	56	3	1
Total	**150**	**0**	**143**	**4**	**3**

Denmark

Target Population (pending)

In Denmark, the national desired target population was Grade 12.

Coverage and Exclusions

All schools were included.

Sample Design

- No explicit stratification was used.
- Implicit stratification was organized by school type.

Table D.32 Allocation of school sample in Denmark

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Denmark	200	0	141	0	59
Total	**200**	**0**	**141**	**0**	**59**

Estonia

Target Population
In Estonia, the national desired target population was Grade 12.

Coverage and Exclusions
School-level exclusions consisted of special schools for the functionally and mentally disabled with atypical educational systems.

Sample Design
• Explicit stratification was organized by five different regions (Central, North-East, Southern, Western, and Northern Estonia), for a total of five explicit strata.

• No implicit stratification was used.

Table D.33 Allocation of school sample in Estonia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Central Estonia	28	0	21	1	6
North-East Estonia	41	0	32	0	9
Southern Estonia	40	0	29	0	11
Western Estonia	25	0	15	2	8
Northern Estonia	78	0	42	0	36
Total	**212**	**0**	**139**	**3**	**70**

Hong Kong (SAR)

Target Population
In Hong Kong (SAR), the national desired target population was Grade 12. Note that the information provided on sampling was insufficient for adjudication and weighting.

Coverage and Exclusions
School-level exclusions consisted of non-native-language-speaking schools.

Sample Design
- No explicit stratification was used.
- Implicit stratification was organized by mode of financing (government-aided, government, and private schools), for a total of three implicit strata.

Israel

Target Population

In Israel, the national desired target population was Grade 11.

Coverage and Exclusions

School-level exclusions consisted of students from ultra-orthodox Jewish schools and Arab private church schools, special education schools, and schools that do not follow the state curriculum.

Sample Design

- Explicit stratification was used by type of school (Hebrew and Arab), for a total of two explicit strata.

- Implicit stratification was organized by school type (regular state and state religious schools), region, socioeconomic status within region, and number of eligible students in each listed school within region (descending and ascending), for a total of 11 implicit strata.

Table D.34 Allocation of school sample in Israel

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Hebrew schools	160	0	138	19	3
Arab schools	77	0	76	0	1
Total	**237**	**0**	**214**	**19**	**4**

Latvia

Target Population

In Latvia, the national desired target population was Grade 12.

Coverage and Exclusions

Coverage in Latvia included students whose language of instruction is Latvian or Russian. School-level exclusions consisted of special schools for the physically and mentally disabled and very small schools.

Sample Design

- Explicit stratification was used by school size (very small schools, moderately small schools, and large schools), for a total of three explicit strata.

- Implicit stratification was organized by school type and urbanization, for a total of four implicit strata.

Table D.35 Allocation of school sample in Latvia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Very small schools	3	0	3	0	0
Moderately small schools	18	0	16	0	2
Large schools	124	0	98	8	18
Total	**145**	**0**	**117**	**8**	**20**

Note: Sample special note necessary.

Norway

Target Population

In Norway, the national desired target population was Grade 12.

Coverage and Exclusions

Exclusions within the national defined target population were schools for individual education, remote schools, and extremely small schools.

Sample Design

- Explicit stratification was not used.

- Implicit stratification was organized by counties, for a total of 19 implicit strata.

Table D.36 Allocation of school sample in Norway

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Large schools	5	0	4	0	1
Other schools	145	0	120	0	25
Total	**150**	**0**	**124**	**0**	**26**

Poland

Target Population

In Poland, the national desired target population was Grade 11.

Coverage and Exclusions

School-level exclusions consisted of schools for students with various forms of physical disability and included job preparation schools, special basic vocational schools, special technical secondary schools, special vocational secondary schools, and special general secondary schools.

Sample Design

- No explicit stratification was used.
- Implicit stratification was organized by provinces, for a total of 16 implicit strata.

Table D.37 Allocation of school sample in Poland

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Poland	150	0	135	15	0
Total	**150**	**0**	**135**	**15**	**0**

Portugal

Target Population
In Portugal, the national desired target population applied was Grade 11.

Coverage and Exclusions
All schools were included.

Sample Design
- Explicit stratification was organized by region (Norte, Centro, Lisboa e Vale Do Tejo, Alentejo, Algarve, Madeira, and Azores), for a total of 11 explicit strata.
- Implicit stratification was organized by school size and school type (regular or professional), for a total of three implicit strata.

Table D.38 Allocation of school sample in Portugal

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Norte	51	0	49	0	2
Centro	27	0	27	0	0
Lisboa e Vale Do Tejo	52	0	48	0	4
Alentejo	7	0	7	0	0
Algarve	5	0	4	0	1
Madeira	4	0	3	0	1
Azores	4	0	4	0	0
Alentejo professional	1	0	1	0	0
Algarve professional	2	0	2	0	0
Madeira professional	2	0	2	0	0
Azores professional	2	0	2	0	0
Total	**157**	**0**	**149**	**0**	**8**

Russian Federation

Target Population

In the Russian Federation, the national desired target population applied was Grade 11.

Coverage and Exclusions

School-level exclusions consisted of special schools for the physically and mentally disabled and schools with non-Russian teaching language.

Sample Design

- Explicit stratification was organized by regions.
- Implicit stratification was organized by school type (rural or urban).

Table D.39 Allocation of school sample in the Russian Federation

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Bashkortostan	4	0	4	0	0
Kabardino-Balkaria	4	0	4	0	0
Kalmykia	3	0	3	0	0
Marii Al	4	0	4	0	0
Tataria	3	0	3	0	0
Udmuttia	4	0	4	0	0
Krasnodar Kr.	4	0	4	0	0
Altay Kr.	4	0	4	0	0
Krasnoyarsk Kr.	4	0	4	0	0
Primor Kr.	3	0	3	0	0
Stavropol Kr.	4	0	4	0	0
Habarovsk Kr.	6	0	6	0	0
Belgorod Obl.	4	0	4	0	0
Vladimir Obl.	4	0	4	0	0
Volgograd Obl.	4	0	4	0	0
Vologda Obl.	8	0	7	0	1
Ust Orda Ol. & Irkutsk Obl.	5	0	3	0	2
Kemerovo Obl.	4	0	4	0	0
Kirov Obl.	4	0	4	0	0
Leningrad Obl.	3	0	3	0	0
Moscow Obl.	4	0	4	0	0
Murmansk Obl.	4	0	4	0	0
N. Novgorod Obl.	3	0	3	0	0
Novgorod Obl.	3	0	3	0	0
Omsk Obl.	3	0	3	0	0
Novosibirsk Obl.	4	0	4	0	0
Orenburg Obl.	3	0	3	0	0
Orel Obl.	4	0	4	0	0
Komi Perm Ok. & Perm Obl.	4	0	4	0	0
Rostov Obl.	4	0	4	0	0
Rasan Obl.	4	0	4	0	0
Saratov Obl.	6	0	6	0	0
Sahalin Obl.	4	0	4	0	0

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Sverdlovsk Obl.	4	0	4	0	0
Smolensk Obl.	3	0	3	0	0
Tambov Obl.	3	0	3	0	0
Tver Obl.	3	0	3	0	0
Tomsk Obl.	2	0	2	0	0
Ulianovsk Obl.	3	0	3	0	0
Chelyabinsk Obl.	3	0	3	0	0
Chita Obl.	4	0	4	0	0
Moscow	4	0	4	0	0
Sankt Petersburg	3	0	3	0	0
Khanty Mansi Ok.	4	0	3	0	1
Total	**169**	**0**	**165**	**0**	**4**

Note: Samara Obl. was not included in the sample.

Slovenia

Target Population

In Slovenia, the national desired target population was Grade 12.

Coverage and Exclusions

Except for three special schools, all schools were included.

Sample Design
- No explicit stratification was used.
- Implicit stratification was organized by geographical area, for a total of three implicit strata.

Table D.40 Allocation of school sample in Slovenia

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Slovenia	151	0	145	0	6
Total	**151**	**0**	**145**	**0**	**6**

Sweden

Target Population

In Sweden, the national desired target population applied was Grade 12.

Coverage and Exclusions

School-level exclusions consisted of very small schools.

Sample Design

- Explicit stratification was organized as for the 14-year-old students by type of community in order to concentrate the sample of schools within strata to the more relevant parts of Sweden, for a total of three explicit strata.

- Implicit stratification was sorted by type of school, for a total of four implicit strata.

Table D.41 Allocation of school sample in Sweden

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Big cities	35	0	30	1	4
Suburbs	35	0	31	0	4
Medium to rural	35	0	32	0	3
Total	**105**	**0**	**93**	**1**	**11**

Switzerland Target Population
In Switzerland, the national desired target population applied was Grade 11.

Coverage and Exclusions
School-level exclusions consisted of special schools for the physically and mentally disabled.

Sample Design
- Explicit stratification was organized by school type, for a total of six strata.
- Implicit stratification was sorted by school track, for a total of one implicit strata.

Table D.42 Allocation of school sample in Switzerland

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools		Non-Participating Schools
			Sampled	Replacement	
Gymnasium	10	0	10	0	0
Berufsbildung	47	0	43	0	4
Diplommittelschule	4	0	4	0	0
Unterichtsberufe	4	0	4	0	0
Berufsmaturität	4	0	4	0	0
Anlehre	4	0	4	0	0
Total	**73**	**0**	**69**	**0**	**4**

APPENDIX E: DESIGN EFFECTS AND EFFECTIVE SAMPLE SIZE TABLES

Table E.1: Total civic knowledge (content knowledge and interpretative skills)

Country	Total Number of Students Assessed	Design Effect	Effective Sample Size
Australia	3,301	5.0	660
Belgium (French)	2,075	5.5	379
Bulgaria	2,857	11.7	243
Chile	5,677	9.5	600
Colombia	4,889	17.8	275
Cyprus	3,102	2.0	1,568
Czech Republic	3,599	7.0	513
Denmark	3,192	2.1	1,498
England	3,011	3.2	929
Estonia	3,418	3.8	900
Finland	2,776	3.1	909
Germany	3,685	2.6	1,414
Greece	3,448	4.3	793
Hong Kong (SAR)	4,993	11.7	426
Hungary	3,165	3.9	807
Italy	3,808	6.1	628
Latvia	2,570	6.8	377
Lithuania	3,494	6.4	547
Norway	3,310	1.8	1,824
Poland	3,372	19.2	176
Portugal	3,234	6.3	511
Romania	2,993	8.5	352
Russian Federation	2,129	8.4	254
Slovak Republic	3,460	6.2	557
Slovenia	3,064	2.1	1,481
Sweden	3,061	4.8	631
Switzerland	3,096	6.6	468
United States	2,786	7.7	360

Table E.2:Total civic knowledge (content knowledge and interpretative skills)

Country	Total Number of Students Assessed	Design Effect	Effective Sample Size
Chile	5,750	6.4	900
Cyprus	1,694	9.9	171
Czech Republic	3,349	4.3	786
Denmark	2,745	1.6	1,706
Estonia	3,175	6.5	488
Israel	5,750	10.5	548
Latvia	2,756	14.3	192
Norway	2,076	5.8	359
Poland	4,041	10.6	382
Portugal	2,734	3.6	750
Russian Federation	1,787	6.0	299
Slovenia	3,728	16.4	227
Sweden	2,677	3.6	739
Switzerland	1,270	11.9	106

APPENDIX F: THE CIVED INSTRUMENTS

F.1 CIVIC KNOWLEDGE TEST ITEMS

Release Items for 14-year-old Students

2. Which of the following is an accurate statement about laws?

   A. ☒ Laws forbid or require certain actions [behaviours].

   B. ☐ Laws are made by the police.

   C. ☐ Laws are valid only if all citizens have voted to accept them.

   D. ☐ Laws prevent criticism of the government.

3. Which of the following is a political right? The right ...

   A. ☐ of pupils to learn about politics in school

   B. ☒ of citizens to vote and stand for [run for] election

   C. ☐ of adults to have a job

   D. ☐ of politicians to have a salary

5. A woman who has a young child is interviewed for a job at a travel agency. Which of the following is an example of discrimination [injustice]? She does not get the job because ...

   A. ☐ she has no previous experience.

   B. ☒ she is a mother.

   C. ☐ she speaks only one language.

   D. ☐ she demands a high salary.

7. In a democratic country [society] having many organisations for people to join is important because this provides ...

   A. ☐ a group to defend members who are arrested.

   B. ☐ many sources of taxes for the government.

   C. ☒ opportunities to express different points of view.

   D. ☐ a way for the government to tell people about new laws.

11. In democratic countries what is the function of having more than one political party?

   A. ☒ To represent different opinions [interests] in the national legislature [e.g. Parliament, Congress]

   B. ☐ To limit political corruption

   C. ☐ To prevent political demonstrations

   D. ☐ To encourage economic competition

12. In a democratic political system, which of the following ought to govern the country?

   A. ☐ Moral or religious leaders

   B. ☐ A small group of well-educated people

   C. ☒ Popularly elected representatives

   D. ☐ Experts on government and political affairs

16. What is the major purpose of the United Nations?
    A. ☐ Safeguarding trade between countries
    B. ☒ Maintaining peace and security among countries
    C. ☐ Deciding where countries' boundaries should be
    D. ☐ Keeping criminals from escaping to other countries

17. Which of the following is most likely to cause a government to be called non-democratic?
    A. ☒ People are prevented from criticising [not allowed to criticise] the government.
    B. ☐ The political parties criticise each other often.
    C. ☐ People must pay very high taxes.
    D. ☐ Every citizen has the right to a job.

18. Which of the following is most likely to happen if a large publisher buys many of the [smaller] newspapers in a country?
    A. ☐ Government censorship of the news is more likely.
    B. ☒ There will be less diversity of opinions presented.
    C. ☐ The price of the country's newspapers will be lowered.
    D. ☐ The amount of advertising in the newspapers will be reduced.

The next three questions are based on the following imaginary political leaflet [political advertisement].

> **We citizens have had enough!**
> A vote for the Silver Party means a vote for higher taxes.
> It means an end to economic growth and a waste of our nation's resources.
> Vote instead for economic growth and free enterprise.
> Vote for more money left in everyone's wallet!
> Let's not waste another 4 years! VOTE FOR THE GOLD PARTY.

23. This is an election leaflet [political advertisement] which has probably been issued by ...
    A. ☐ the Silver Party.
    B. ☒ a party or group in opposition to [running against] the Silver Party.
    C. ☐ a group which tries to be sure elections are fair.
    D. ☐ the Silver Party and the Gold Party together.

24. The authors of the leaflet think that higher taxes are ...

A. ☐ a good thing.

B. ☐ necessary in a [free] market economy.

C. ☐ necessary for economic growth.

D. ☒ a bad thing.

25. The party or group that has issued this leaflet is likely also to be in favour of ...

A. ☒ reducing state [government] control of the economy.

B. ☐ lowering of the voting age.

C. ☐ capital punishment.

D. ☐ more frequent elections.

26. Two people work at the same job but one is paid less than the other. The principle of equality would be violated if the person is paid less because of ...

A. ☐ fewer educational qualifications.

B. ☐ less work experience.

C. ☐ working for fewer hours.

D. ☒ gender [sex].

The next question differs from those earlier in the test. The following question contains three statements of fact and one statement of opinion. Read each question, and then choose the opinion.

31. Three of these statements are facts and one is an opinion. Which of the following is an OPINION?

A. ☒ Actions by individual countries are the best way to solve environmental problems.

B. ☐ Many countries contribute to the pollution of the environment.

C. ☐ Some countries offer to co-operate in order to diminish acid rain.

D. ☐ Water pollution often comes from several different sources.
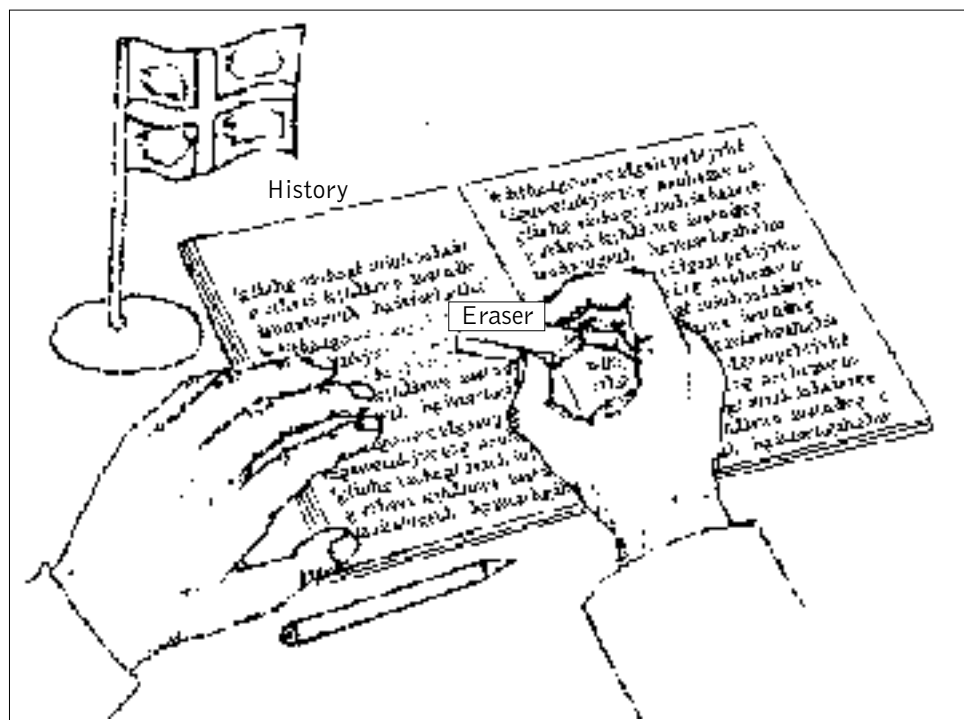
36. What is the message or main point of this cartoon? History textbooks ...

   A. ☒ are sometimes changed to avoid mentioning problematic events from the past.

   B. ☐ for children must be shorter than books written for adults.

   C. ☐ are full of information that is not interesting.

   D. ☐ should be written using a computer and not a pencil.

The next question differs from those earlier in the test. The following question contains three statements of opinion and one statement of fact. Read each question, and then choose the fact.

38. Three of these statements are opinions and one is a fact. Which of the following is a FACT [the factual statement]?

   A. ☐ People with very low incomes should not pay any taxes.

   B. ☒ In many countries rich people pay higher taxes than poor people.

   C. ☐ It is fair that some citizens pay higher taxes than others.

   D. ☐ Donations to charity are the best way to reduce differences between rich and poor.

Additional Release Items for Upper Secondary Students

The next question is based on the following part of an article from an imaginary newspaper.

> **OPEC TO REDUCE OIL PRODUCTION**
>
> The Organisation of Petroleum Exporting Countries met in Vienna last week. Leaders of the oil producing countries agreed to reduce the amount of oil each of them pump. At a press conference today, the organisation's spokesman said that this will help to keep the world's economy in balance.

34. Why would the OPEC countries decide to reduce their oil production?

   A. ☐ To help economic growth in the world

   B. ☒ To keep oil prices from falling

   C. ☐ To save oil for future generations

   D. ☐ To promote the development of new energy sources

35. If there were a high protective tariff [tax] in Brazil on cars made in Japan, who would most directly benefit?

   A. ☐ car-makers in Japan

   B. ☐ people in Brazil who buy cars made in Japan

   C. ☒ car-makers in Brazil

   D. ☐ the government in Japan

F.2 STUDENT BACKGROUND QUESTIONNAIRE

1. On what date were you born?

    Write in the month, day and year.

    _____ month _____ day _____ year

2. Are you a girl or a boy?

    Tick one box only.

    girl   ☐   1
    boy   ☐   2

3. Which best describes you?

    Tick one box only.

    COUNTRY LIST

    [A] .... ☐   1
    [B] .... ☐   2
    [C] .... ☐   3
    [D] .... ☐   4
    [E] .... ☐   5

4. Were you born in [country of test]?

    No .... ☐   1
    Yes .... ☐   2

5. If you were not born in [country of test], how old were you when you came to [country of test]?

    Write in your age at the time.

    I was .... years old when I came to [name of the country].

6. How often do you speak [language of test] at home?

    Tick one box only.

    Never   ☐   1
    Sometimes   ☐   2
    Always or almost always   ☐   3

7. Does any of these people live at home with you most or all of the time?

	no 1	yes 2
Mother or stepmother or female guardian	☐	☐
Father or stepfather or male guardian	☐	☐

8. Altogether, how many people live in your home?

   Write in the total number of people.

   .... *(Don't forget to include yourself.)*

9. Do you get a daily newspaper at home?

   No .... ☐ 1

   Yes .... ☐ 2

10. About how many books are there in your home?

    Do not count newspapers, magazines or books for school; tick one box only.

    None ☐ 1

    1 - 10 ☐ 2

    11 - 50 ☐ 3

    51 - 100 ☐ 4

    101 - 200 ☐ 5

    More than 200 ☐ 6

11. How many years of further education do you expect to complete after this year?

    Please include vocational education and/or higher education. Tick one box only.

    0 years ☐ 1

    1 or 2 years ☐ 2

    3 or 4 years ☐ 3

    5 or 6 years ☐ 4

    7 or 8 years ☐ 5

    9 or 10 years ☐ 6

    More than 10 years ☐ 7

12. How far in school did your mother and father go?

    Tick only one box in each column.

	Mother	Father
Did not finish elementary school	☐ 1	☐ 1
Finished elementary school	☐ 2	☐ 2
Finished some high school	☐ 3	☐ 3
Finished high school	☐ 4	☐ 4
Some vocational/technical education after high school	☐ 5	☐ 5
Some community college, college, or university courses	☐ 6	☐ 6
Completed a bachelor's degree at a college or university	☐ 7	☐ 7
I don't know	☐ 0	☐ 0

13. Have you participated in the following organisations?

Tick the appropriate box in each row.

	no 1	yes 2
a) A student council/student government [class or school parliament]	☐	☐
b) A youth organisation affiliated with a political party or union	☐	☐
c) A group which prepares a school newspaper	☐	☐
d) An environmental organisation	☐	☐
e) A U. N. or UNESCO Club	☐	☐
f) A student exchange or school partnership program	☐	☐
g) A human rights organisation	☐	☐
h) A group conducting [voluntary] activities to help the community	☐	☐
i) A charity collecting money for a social cause	☐	☐
j) Boy or Girl Scouts [Guides]	☐	☐
k) A cultural association [organisation] based on ethnicity	☐	☐
l) A computer club	☐	☐
m) An art, music or drama organisation	☐	☐
n) A sports organisation or team	☐	☐
o) An organisation sponsored by a religious group	☐	☐

14. Think about all the organisations listed above. How often do you attend meetings or activities for any or all of these organisations?

Almost every day (4 or more days a week)	☐ 1
Several days (1 to 3 days a week)	☐ 2
A few times each month	☐ 3
Never or almost never	☐ 4

In the next few questions think about the days on which you attend school.

15. How often do you spend time [directly] after school talking [hanging out] with your friends?

Almost every day (4 or more days a week)	☐ 1
Several days (1 to 3 days a week)	☐ 2
A few times each month	☐ 3
Never or almost never	☐ 4

16. How often do you spend time during the evening [after dinner or after
    —] outside your home with your friends?

    Almost every day (4 or more days a week)   ☐ 1

    Several days (1 to 3 days a week)   ☐ 2

    A few times each month   ☐ 3

    Never or almost never   ☐ 4

17. How much time do you spend watching television or videos on school
    days?

    no time   ☐ 1

    less than 1 hour   ☐ 2

    1 to 2 hours   ☐ 3

    3 to 5 hours   ☐ 4

    more than 5 hours   ☐ 5

F.3 LIKERT-TYPE ITEMS ON STUDENT CONCEPTS, ATTITUDES, AND ACTION

	Scaled items
	Items not in scales

Section A: Democracy

You are going to read a list of things that might happen in a country that is a democracy. Each one of them could either be good and have positive results for democracy or it could be bad and have negative results for democracy.

There are **no right answers and no wrong answers** to these questions, because we just want to know what you think about democracy and the things that might influence it.

Please tick the box in the column which best fits your opinion. If you think that the statement does not apply, put a tick in the circle in the last column.

What is good and what is bad for democracy?

		very bad for democracy 1	somewhat bad for democracy 2	somewhat good for democracy 3	very good for democracy 4	don't know/ doesn't apply 0
A1	When everyone has the right to express their opinions freely that is	☐	☐	☐	☐	◯
A2	When differences in income and wealth between the rich and the poor are small, that is	☐	☐	☐	☐	◯
A3	When political leaders in power give jobs in the government [public sector] to members of their family, that is	☐	☐	☐	☐	◯
A4	When newspapers are free of all government [state, political] control, that is	☐	☐	☐	☐	◯
A5	When private businesses have no restrictions from government, that is	☐	☐	☐	☐	◯
A6	When one company owns all the newspapers, that is	☐	☐	☐	☐	◯
A7	When people demand their political and social rights, that is	☐	☐	☐	☐	◯

Democracy (continued)

	very bad for democracy	somewhat bad for democracy	somewhat good for democracy	very good for democracy	don't know/ doesn't apply
	1	2	3	4	0
A8 When immigrants are expected to give up the language and customs of their former countries, that is	☐	☐	☐	☐	○
A9 When political parties have rules that support women to become political leaders, that is	☐	☐	☐	☐	○
A10 When people who are critical of the government are forbidden from speaking at public meetings, that is	☐	☐	☐	☐	○
A11 When citizens have the right to elect political leaders freely, that is	☐	☐	☐	☐	○
A12 When courts and judges are influenced by politicians, that is	☐	☐	☐	☐	○
A13 When many different organisations [associations] are available [exist] for people who wish to belong to them, that is	☐	☐	☐	☐	○
A14 When there is a separation [segregation] between the church [institutional church] and the state [government], that is	☐	☐	☐	☐	○
A15 When young people have an obligation [are obliged] to participate in activities to benefit [help] the community [society], that is	☐	☐	☐	☐	○
A16 When a minimum income [living standard] is assured for everyone, that is	☐	☐	☐	☐	○
A17 When political parties have different opinions [positions] on important issues, that is	☐	☐	☐	☐	○
A18 When people participate in political parties in order to influence government, that is	☐	☐	☐	☐	○

Democracy (continued)

	very bad for democracy	somewhat bad for democracy	somewhat good for democracy	very good for democracy	don't know/ doesn't apply
	1	2	3	4	0
A19 When laws that women claim are unfair to them are changed, that is	☐	☐	☐	☐	○
A20 When all the television stations present the same opinion about politics, that is	☐	☐	☐	☐	○
A21 When people refuse to obey a law which violates human rights, that is	☐	☐	☐	☐	○
A22 When newspapers are forbidden to publish stories that might offend ethnic groups [immigrant groups, racial groups, national groups], that is	☐	☐	☐	☐	○
A23 When wealthy business people have more influence on government than others, that is	☐	☐	☐	☐	○
A24 When government leaders are trusted without question, that is	☐	☐	☐	☐	○
A25 When people peacefully protest against a law they believe to be unjust, that is	☐	☐	☐	☐	○

Section B: Good Citizens

In this section there are some statements that could be used to explain what a good adult citizen is or what a good adult citizen does. There are no right and wrong answers to these questions.

For each of these statements, tick one box to show how important you believe each is for explaining what a good adult citizen is or does.

An adult who is a good citizen ...

	not important	somewhat unimportant	somewhat important	very important	don't know/ doesn't apply
	1	2	3	4	0
B1 obeys the law	☐	☐	☐	☐	○
B2 votes in every election[a]	☐	☐	☐	☐	○
B3 joins a political party[a]	☐	☐	☐	☐	○
B4 works hard	☐	☐	☐	☐	○

Good citizens (continued)

		not important 1	somewhat unimportant 2	somewhat important 3	very important 4	don't know/ doesn't apply 0
B5	would participate in a peaceful protest against a law believed to be unjust[b]	☐	☐	☐	☐	○
B6	knows about the country's history[a]	☐	☐	☐	☐	○
B7	would be willing to serve in the military to defend the country	☐	☐	☐	☐	○
B8	follows political issues in the newspaper, on the radio or on TV[a]	☐	☐	☐	☐	○
B9	participates in activities to benefit people in the community [society] [b]	☐	☐	☐	☐	○
B10	shows respect for government representatives [leaders, officials] [a]	☐	☐	☐	☐	○
B11	takes part in activities promoting human rights[b]	☐	☐	☐	☐	○
B12	engages in political discussions[a]	☐	☐	☐	☐	○
B13	takes part in activities to protect the environment[b]	☐	☐	☐	☐	○
B14	is patriotic and loyal [devoted] to the country	☐	☐	☐	☐	○
B15	would be willing to ignore [disregard] a law that violated human rights	☐	☐	☐	☐	○

[a] Scaled Item: **Importance of Conventional Citizenship.**
[b] Scaled Item: **Importance of Social-movement-related Citizenship.**

Section C: Government

Below you will find some statements about the responsibilities of the government [state].

What responsibilities should the government have?

Read each of these statements and tick the appropriate box to decide whether what is described should or should not be the government's [state's] responsibility.

	definitely should not be the government's responsibility	probably should not be the government's responsibility	probably should be the government's responsibility	definitely should be the government's responsibility	don't know
	1	2	3	4	0
C1 To guarantee a job for everyone who wants one[d]	☐	☐	☐	☐	◯
C2 To keep prices under control[d]	☐	☐	☐	☐	◯
C3 To provide basic health care for everyone[c]	☐	☐	☐	☐	◯
C4 To provide an adequate [decent] standard of living for old people[c]	☐	☐	☐	☐	◯
C5 To provide industries with the support they need to grow[d]	☐	☐	☐	☐	◯
C6 To provide an adequate [decent] standard of living for the unemployed[d]	☐	☐	☐	☐	◯
C7 To reduce differences in income and wealth among people[d]	☐	☐	☐	☐	◯
C8 To provide free basic education for all[c]	☐	☐	☐	☐	◯
C9 To ensure [be sure there are] equal political opportunities for men and women[c]	☐	☐	☐	☐	◯
C10 To control pollution of the environment[c]	☐	☐	☐	☐	◯
C11 To guarantee peace and order [stability] within the country[c]	☐	☐	☐	☐	◯
C12 To promote honesty and moral behaviour among people in the country[c]	☐	☐	☐	☐	◯

[c] Scaled Item: **Concept of Society-related Government Responsibilities.**

[d] Scaled Item: **Concept of Economy-related Government Responsibilities.**

SECTION D: TRUST IN INSTITUTIONS

In this section we will name several institutions in this country [name of country]:

How much of the time can you trust each of the following institutions?

Consider each of these institutions and select the box in the column which shows how you feel you can trust them.

		never	only some of the time	most of the time	always	don't know
		1	2	3	4	0
D1	The national [federal] government [in _____(the national seat of government)]e	☐	☐	☐	☐	◯
D2	The local council or government of your town or citye	☐	☐	☐	☐	◯
D3	Courtse	☐	☐	☐	☐	◯
D4	The policee	☐	☐	☐	☐	◯
D5	News on television	☐	☐	☐	☐	◯
D6	News on the radio	☐	☐	☐	☐	◯
D7	News in the press [newspapers]	☐	☐	☐	☐	◯
D8	Political partiese	☐	☐	☐	☐	◯
D9	United Nations	☐	☐	☐	☐	◯
D10	Schools [educational institutions]	☐	☐	☐	☐	◯
D11	National Parliament [Congress]e	☐	☐	☐	☐	◯
D12	The people who live in this country [name of country]	☐	☐	☐	☐	◯

e Scaled Item: **Trust in Government-related Institutions.**

In the next sections you will find statements on different topics.

You may agree with some of the statements and disagree with others. Sometimes you will feel that you disagree or agree strongly, and sometimes you will feel less strongly.

There are no right and wrong answers to these questions, we just want to know your opinion.

Section E: Our Country

In this section you will find some statements about this country [name of country].

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		strongly disagree	disagree	agree	strongly agree	don't know
		1	2	3	4	0
E1	To help protect jobs in this country [name of country] we should buy products made in this country [name of country]	☐	☐	☐	☐	◯
E2	We should keep [prevent] other countries from trying to influence political decisions in this country [name of country]	☐	☐	☐	☐	◯
E3	The flag of this country [name of country] is important to me[f]	☐	☐	☐	☐	◯
E4	We should always be alert and stop threats from other countries to this country [name of country]'s political independence	☐	☐	☐	☐	◯
E5	This country [name of country] deserves respect from other countries for what we have accomplished	☐	☐	☐	☐	◯
E6	There is little to be proud of in this country [name of country]'s history	☐	☐	☐	☐	◯
E7	I have great love for this country [name of country][f]	☐	☐	☐	☐	◯
E8	People should support their country even if they think their country is doing something wrong	☐	☐	☐	☐	◯

Our country (continued)

		strongly disagree	disagree	agree	strongly agree	don't know
		1	2	3	4	0
E9	This country [name of country] should be proud of what it has achieved[f]	☐	☐	☐	☐	◯
E10	The national anthem of this country [name of country] is important to me	☐	☐	☐	☐	◯
E11	I would prefer to live permanently in another country[f]	☐	☐	☐	☐	◯
E12	We should stop outsiders from influencing this country [name of country]'s traditions and culture	☐	☐	☐	☐	◯

[f] Scaled Item: **Positive Attitudes toward One's Nation.**

Section F: Opportunities 1

In this section there are some statements about the chances which members of certain groups REALLY DO HAVE in this country [name of country].

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		strongly disagree	disagree	agree	strongly agree	don't know
		1	2	3	4	0
F1	Children who are members of certain ethnic groups [immigrant groups, national groups, racial groups] have fewer chances than other children to get a [good] secondary [high school] education in this country	☐	☐	☐	☐	○
F2	Girls have fewer chances than boys to get a [good] secondary [high school] education in this country	☐	☐	☐	☐	○
F3	Children from poor families have fewer chances than others to get a [good] secondary [high school] education in this country	☐	☐	☐	☐	○
F4	Children who live in rural [farming] areas have fewer chances than others to get a [good] secondary [high school] education in this country	☐	☐	☐	☐	○
F5	Adults who are members of certain ethnic groups [immigrant groups, national groups, racial groups] have fewer chances than others to get good jobs in this country	☐	☐	☐	☐	○
F6	Women have fewer chances than men to get good jobs in this country	☐	☐	☐	☐	○

Section G: Opportunities 2

In this section there are some statements about the opportunities which members of certain groups SHOULD HAVE in this country [name of the country].

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		strongly disagree	disagree	agree	strongly agree	don't know
		1	2	3	4	0
G1	Women should run for public office [a seat in the legislature] and take part in the government just as men do[h]	☐	☐	☐	☐	◯
G2	All ethnic [racial or national] groups should have equal chances to get a good education in this country	☐	☐	☐	☐	◯
G3	Members of anti-democratic groups [groups that are against democracy] should be prohibited from hosting a television show talking about these [their] ideas	☐	☐	☐	☐	◯
G4	Women should have the same rights as men in every way[h]	☐	☐	☐	☐	◯
G5	All ethnic [racial or national] groups should have equal chances to get good jobs in this country	☐	☐	☐	☐	◯
G6	Women should stay out of politics[h]	☐	☐	☐	☐	◯
G7	Members of anti-democratic groups [groups that are against democracy] should be prohibited from organising peaceful [non-violent] demonstrations or rallies	☐	☐	☐	☐	◯
G8	Schools should teach students to respect members of all ethnic [racial or national] groups	☐	☐	☐	☐	◯
G9	When jobs are scarce, men [should] have more right to a job than women[h]	☐	☐	☐	☐	◯
G10	Members of anti-democratic groups [groups that are against democracy] should be prohibited from running in an election for political office	☐	☐	☐	☐	◯

Opportunities 2 (continued)

	strongly disagree	disagree	agree	strongly agree	don't know
	1	2	3	4	0
G11 Men and women should get equal pay when they are in the same jobs [occupations][h]	☐	☐	☐	☐	◯
G12 Members of all ethnic [racial or national] groups should be encouraged to run in elections for political office	☐	☐	☐	☐	◯
G13 Men are better qualified to be political leaders than women[h]	☐	☐	☐	☐	◯
G14 Members of anti-democratic groups [groups that are against democracy] should be prohibited from making public speeches about these [their] ideas	☐	☐	☐	☐	◯

[h] Scaled Item: **Positive Attitudes toward Women's Political and Economic Rights.**

Section H: Immigrants

Listed below you will read several statements about immigrants and immigration in this country [name of country].

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		strongly disagree	disagree	agree	strongly agree	don't know
		1	2	3	4	0
H1	Immigrants should have the opportunity [option] to keep [continue speaking] their own language[g]	☐	☐	☐	☐	◯
H2	Immigrants' children should have the same opportunities for education that other children in the country have[g]	☐	☐	☐	☐	◯
H3	Immigrants who live in a country for several years should have the opportunity to vote in elections[g]	☐	☐	☐	☐	◯
H4	Immigrants should have the opportunity [option] to keep [continue] their own customs and lifestyle[g]	☐	☐	☐	☐	◯
H5	Immigrants should have all the same rights that everyone else in a country has[g]	☐	☐	☐	☐	◯
H6	Immigrants should be forbidden to engage in political activity	☐	☐	☐	☐	◯
H7	Having many immigrants makes it difficult for a country to be united and patriotic	☐	☐	☐	☐	◯
H8	All countries should accept refugees who are trying to escape from wars or political persecution in other countries	☐	☐	☐	☐	◯

[g] Scaled Item: **Trust in Government-related Institutions.**

Section I: The Political System

In this section there are some statements about the political system and your personal view on politics in general.

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		strongly disagree 1	disagree 2	agree 3	strongly agree 4	don't know 0
I1	The government [people in government] cares [care] a lot about what all of us think about new laws	☐	☐	☐	☐	○
I2	I know more about politics than most people my age	☐	☐	☐	☐	○
I3	The government [people in government] is [are] doing its best to find out what people [ordinary people] want	☐	☐	☐	☐	○
I4	The powerful leaders in government [Government] care very little about the opinions of people [ordinary people]	☐	☐	☐	☐	○
I5	When political issues or problems are being discussed, I usually have something to say	☐	☐	☐	☐	○
I6	In this country a few individuals have a lot of political power while the rest of the people have very little power	☐	☐	☐	☐	○
I7	The politicians quickly forget the needs of the voters who elected them.	☐	☐	☐	☐	○
I8	I am able to understand most political issues easily	☐	☐	☐	☐	○
I9	When people get together [organise] to demand change, the leaders in government listen	☐	☐	☐	☐	○
I10	I am interested in politics	☐	☐	☐	☐	○

Section J: School

Listed below you will find some statements on students' partipation in school life.

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		strongly disagree 1	disagree 2	agree 3	strongly agree 4	don't know 0
J1	Electing student representatives to suggest changes in how the school is run [how to solve school problems] makes schools better[j]	☐	☐	☐	☐	○
J2	Lots of positive changes happen in this school when students work together[j]	☐	☐	☐	☐	○
J3	Organising groups of students to state their opinions could help solve problems in this school[j]	☐	☐	☐	☐	○
J4	If members of my class felt they were unfairly treated, I would be willing to go with them to speak to the teacher	☐	☐	☐	☐	○
J5	Students acting together [in groups] can have more influence on what happens in this school than students acting alone [by themselves][j]	☐	☐	☐	☐	○
J6	I am interested in participating in discussions about school problems	☐	☐	☐	☐	○
J7	When school problems are being discussed I usually have something to say	☐	☐	☐	☐	○

[j] Scaled Item: **Confidence in Participation at School.**

Section K: School Curriculum

In this section we would like to know what you have learned in school.

Please read each statement listed below and select the box in the column which corresponds to the way you feel about the statement.

	strongly disagree 1	disagree 2	agree 3	strongly agree 4	don't know 0
K1 In school I have learned to understand people who have different ideas	☐	☐	☐	☐	○
K2 In school I have learned to co-operate [work together] in groups with other students	☐	☐	☐	☐	○
K3 In school I have learned to contribute to solving problems in the community [society]	☐	☐	☐	☐	○
K4 In school I have learned to be a patriotic and loyal [committed] citizen of my country	☐	☐	☐	☐	○
K5 In school I have learned how to act to protect the environment	☐	☐	☐	☐	○
K6 In school I have learned to be concerned about what happens in other countries	☐	☐	☐	☐	○
K7 In school I have learned about the importance of voting in national and local elections	☐	☐	☐	☐	○

Section L: Political Action 1

In this set of questions you will find some activities related to politics.

For each of these activities, tick the box to show how often you do it.

How often do you have discussions of what is happening in your national [your country's] politics [government]?

		never 1	rarely 2	sometimes 3	often 4	don't know 0
L1	With people of your own age [peers]	☐	☐	☐	☐	○
L2	With parents or other adult family members	☐	☐	☐	☐	○
L3	With teachers	☐	☐	☐	☐	○

How often do you have discussions of what is happening in international politics?

		never 1	rarely 2	sometimes 3	often 4	don't know 0
L4	With people of your own age [peers]	☐	☐	☐	☐	○
L5	With parents or other adult family members	☐	☐	☐	☐	○
L6	With teachers	☐	☐	☐	☐	○

How often do you ...

		never 1	rarely 2	sometimes 3	often 4	don't know 0
L7	read articles (stories) in the newspaper about what is happening in this country?	☐	☐	☐	☐	○
L8	read articles (stories) in the newspaper about what is happening in other countries?	☐	☐	☐	☐	○
L9	listen to news broadcasts on television?	☐	☐	☐	☐	○
L10	listen to news broadcasts on the radio?	☐	☐	☐	☐	○

Section M: Political Action 2

Listed below are several types of action that adults could take: When you are an adult, what do you expect that you will do?

Tick one box in each column for each action to show how likely you would be to do it.

		I will certainly not do this 1	I will probably not do this 2	I will probably do this 3	I will certainly do this 4	don't know 0
M1	Vote in national elections	☐	☐	☐	☐	○
M2	Get information about candidates before voting in an election	☐	☐	☐	☐	○
M3	Join a political party[i]	☐	☐	☐	☐	○
M4	Write letters to a newspaper about social or political concerns[i]	☐	☐	☐	☐	○
M5	Be a candidate for a local or city office[i]	☐	☐	☐	☐	○

[i] Scaled Item: **Political Activities.**

Listed below are several types of action that you as a young person could take during the next few years: What do you expect that you will do?

Again tick one box in each column for each action to show how likely you would be to do it. If you don't know, put a tick in the circle in the last column.

		I will certainly not do this 1	I will probably not do this 2	I will probably do this 3	I will certainly do this 4	don't know 0
M6	Volunteer time to help [benefit] [poor or elderly] people in the community	☐	☐	☐	☐	○
M7	Collect money for a social cause	☐	☐	☐	☐	○
M8	Collect signatures for a petition	☐	☐	☐	☐	○
M9	Participate in a non-violent [peaceful] protest march or rally	☐	☐	☐	☐	○
M10	Spray-paint protest slogans on walls	☐	☐	☐	☐	○
M11	Block traffic as a form of protest	☐	☐	☐	☐	○
M12	Occupy public buildings as a form of protest	☐	☐	☐	☐	○

Section N: Classrooms

The next part of the questionnaire includes some statements about things that happen in your school. When answering these questions think especially about classes in history, civic education or social studies [other civic-related subjects].

Please read each statement and select the box in the column which corresponds to the way you feel about the statement.

		never 1	rarely 2	sometimes 3	often 4	don't know 0
N1	Students feel free to disagree openly with their teachers about political and social issues during class[k]	☐	☐	☐	☐	○
N2	Students are encouraged to make up their own minds about issues[k]	☐	☐	☐	☐	○
N3	Teachers respect our opinions and encourage us to express them during class[k]	☐	☐	☐	☐	○
N4	Teachers place great importance [stress, emphasis] on learning facts or dates when presenting history or political events	☐	☐	☐	☐	○
N5	Students feel free to express opinions in class even when their opinions are different from most of the other students[k]	☐	☐	☐	☐	○
N6	Teachers require students to memorise dates or definitions	☐	☐	☐	☐	○
N7	Teachers encourage us to discuss political or social issues about which people have different opinions[k]	☐	☐	☐	☐	○
N8	Teachers present several sides of [positions on] an issue when explaining it in class[k]	☐	☐	☐	☐	○
N9	Students bring up current political events for discussion in class	☐	☐	☐	☐	○
N10	Memorising dates and facts is the best way to get a good grade [mark] from teachers in these classes	☐	☐	☐	☐	○

N11 Teachers lecture and the students take notes ☐ ☐ ☐ ☐ ◯

N12 Students work on material from the textbook ☐ ☐ ☐ ☐ ◯

ᵏ Scaled Item: **Open Classroom Climate for Discussion.**

Section O: Effectiveness of Political Action (Older Population only)

There are many opinions on how one can effectively influence decisions in society. Here is a list of a few ways that are used. How effective do you think they are.

For each of these activities, tick one box to show how effective you think it is.

	not at all effective 1	little effective 2	somewhat effective 3	very effective 4	don't know 0
O1 Working in political parties	☐	☐	☐	☐	◯
O2 Working in local action groups	☐	☐	☐	☐	◯
O3 Working in trade unions	☐	☐	☐	☐	◯
O4 Voting in elections	☐	☐	☐	☐	◯
O5 Personally contacting influential people	☐	☐	☐	☐	◯
O6 Participating in public demonstrations	☐	☐	☐	☐	◯
O7 Working to get attention by the press, radio and TV	☐	☐	☐	☐	◯
O8 Participating in illegal protest activities	☐	☐	☐	☐	◯

Section P: Use of Military Force (Older Population only)

In your opinion, how justified is the use of military force by one country against another for each of the following reasons.

For each of these reasons, tick one box to show how justified you think the use of military force is.

		definitely unjustified 1	unjustified 2	justified 3	definitely justified 4	don't know 0
P1	To enforce compliance with UN resolutions	☐	☐	☐	☐	○
P2	To defend itself when attacked	☐	☐	☐	☐	○
P3	To act against terrorist organisations	☐	☐	☐	☐	○
P4	To stop violations of human rights	☐	☐	☐	☐	○
P5	To get back [retrieve] occupied territories	☐	☐	☐	☐	○
P6	To prevent a possible future attack	☐	☐	☐	☐	○
P7	To get back [retrieve] holy [sacred] places	☐	☐	☐	☐	○
P8	To destroy nuclear and chemical weapons factories	☐	☐	☐	☐	○

F.4 TEACHER QUESTIONNAIRE (STANDARD POPULATION ONLY)

Part 1: Work Experience and Education

1. What [civic-related subject(s)] do you teach this school year?

   _____

   _____

   _____

2. Do you teach in the tested class?

   No   ☐ 1

   Yes   ☐ 2

   [If yes, what subject   ]

3. Are you the home room/class teacher of the tested class?

   No   ☐ 1

   Yes   ☐ 2

4. For how many years, including the present year, have you been teaching altogether?

   _____ years.

5. For how many years, including the present year, have you been teaching [civic education or a civic education-related subject]?

   _____ years.

6. What is the highest level of formal education you have completed?

   [A] .... ☐ 1

   [B] .... ☐ 2

   [C] .... ☐ 3

   [D] .... ☐ 4

   [E] .... ☐ 5

7. Do you hold a degree from an academic/teacher education institution in a discipline related to civic education?

   No   ☐ 1

   Yes   ☐ 2

   If yes, please name...

   discipline/s: *level of the degree/s:*

   1. _____   1. _____

   2. _____   2. _____

   3. _____   3. _____

   4. _____   4. _____

   5. _____   5. _____

8. Have you participated in in-service professional development activities or training in a discipline related to civic education?

No ☐ 1

Yes ☐ 2

If yes, what was/were the name/s of the course/s:

1. _____

2. _____

3. _____

4. _____

5. _____

9. How old are you?

Under 25 ☐ 1

25–29 ☐ 2

30–39 ☐ 3

40–49 ☐ 4

50–59 ☐ 5

60 or more ☐ 6

10. Are you female or male?

Female ☐ 1

Male ☐ 2

Part 2: Views on Civic Education

With the statements in this section we would like to inquire about your views on civic education in the curriculum of schools.

Section A: How should civic education be taught?

Please rate the statements below on the following scale:

		strongly disagree	disagree	agree	strongly agree
Civic education ...		1	2	3	4
A1	should be taught as a specific subject	☐	☐	☐	☐
A2	should be taught integrated into subjects related to human and social sciences, like history, geography, languages, religion, ethics, law				
A3	should be integrated into all subjects taught at school				
A4	should be an extra-curricular activity				

Section B: What is worth learning in civic education?

Please rate the statements below on the following scale:

		strongly disagree	disagree	agree	strongly agree
		1	2	3	4
B1	There is broad consensus in our society as to what is worth learning in civic education	☐	☐	☐	☐
B2	Teachers should negotiate with students what is to be studied in civic education	☐	☐	☐	☐
B3	Teachers should teach according to curriculum standards/requirements in the area of civic education	☐	☐	☐	☐
B4	What is important in civic education cannot be taught in school	☐	☐	☐	☐
B5	Because of conflicts and different opinions in society there cannot be agreement on what should be taught in civic education	☐	☐	☐	☐
B6	Changes have been so rapid in recent years that teachers often do not know what to teach in civic education	☐	☐	☐	☐

Section C: How much does civic education matter?

Please rate the statements below on the following scale:

		strongly disagree 1	disagree 2	agree 3	strongly agree 4
C1	Teaching civic education makes a difference for students' political and civic development	☐	☐	☐	☐
C2	Teaching civic education at school matters a great deal for our country	☐	☐	☐	☐
C3	Schools are irrelevant for the development of students' attitudes and opinions about matters of citizenship	☐	☐	☐	☐
C4	Education authorities pay little attention to civic education	☐	☐	☐	☐

Section D: What is emphasised in civic education at your school?

Tick only one box for each column!

When I look at civic education in my school, I believe most emphasis...

	is placed on	*should be* placed on
knowledge about society	☐ 1	☐ 1
student independent (critical) thinking	☐ 2	☐ 2
student participation in community and political activities	☐ 3	☐ 3
development of values	☐ 4	☐ 4

Section E: What do students learn in your school?

Please rate the statements below on the following scale:

		strongly disagree 1	disagree 2	agree 3	strongly agree 4
E1	In our school students learn to understand people who have different [ideas/points of view]	☐	☐	☐	☐
E2	In our school students learn to co-operate [work together] in groups with other students	☐	☐	☐	☐
E3	In our school students learn to contribute to solve problems in the community [society]	☐	☐	☐	☐
E4	In our school students learn to be patriotic and loyal [committed] citizens of their country	☐	☐	☐	☐

What do students learn in your school (continued)

		strongly disagree 1	disagree 2	agree 3	strongly agree 4
E5	In our school students learn how to act to protect the environment	☐	☐	☐	☐
E6	In our school students learn to be concerned about what happens in other countries	☐	☐	☐	☐
E7	In our school students learn about the importance of voting in national and local elections	☐	☐	☐	☐

Section F: What should students learn to become good citizens?

Please rate the items below on the following scale:

To become a good adult citizen students should learn to recognise the importance of ...

		strongly disagree 1	disagree 2	agree 3	strongly agree 4
F1	obeying the law	☐	☐	☐	☐
F2	voting on every election	☐	☐	☐	☐
F3	joining a political party	☐	☐	☐	☐
F4	working hard	☐	☐	☐	☐
F5	participating in a peaceful protest against a law believed to be injust	☐	☐	☐	☐
F6	knowing about the country's history	☐	☐	☐	☐
F7	being willing to serve in the military to defend the country	☐	☐	☐	☐
F8	reading about [following] political issues in the newspaper, on the radio or on TV	☐	☐	☐	☐
F9	participating in activities to help people in the community [society]	☐	☐	☐	☐
F10	showing respect for government representatives [leaders, officials]	☐	☐	☐	☐
F11	taking part in activities promoting human rights	☐	☐	☐	☐
F12	engaging in political discussion	☐	☐	☐	☐
F13	being patriotic and loyal [devoted] to the country	☐	☐	☐	☐
F14	ignoring [disregarding] a law that violated human rights	☐	☐	☐	☐

Part 3: The Teaching of [Civic Education-related] Subjects, Activities, and Lessons

Section G: How do you plan for civic education? When you prepare for civic education-related activities, from what sources do you draw?

Please rate the importance of each source on the following scale:

		not important 1	less important 2	important 3	very important 4
G1	Official curricula or curricular guidelines or frameworks	☐	☐	☐	☐
G2	Official requirements (standards) in the area of civic education	☐	☐	☐	☐
G3	Your own ideas of what is important to know in civic education	☐	☐	☐	☐
G4	Original sources (such as constitutions, human rights declarations)	☐	☐	☐	☐
G5	[Approved] Textbooks	☐	☐	☐	☐
G6	Materials published by commercial companies, public institutes, or private foundations	☐	☐	☐	☐
G7	Self-produced materials	☐	☐	☐	☐
G8	Media (newspapers, magazines, television)	☐	☐	☐	☐

Section H: What topics do you teach?

The following table contains a list of 20 topics.

Please rate them by checking the appropriate boxes in the table according to...

List of Topics	(a) How important do you think this topic is for civic education?				(b) How confident (well equipped) do you feel to deal with this topic?				(c) How much opportunity do students up to & including grade [xx] have to learn this topic?			
	not important 1	of little importance 2	important 3	very important 4	not at all 1	Little confident 2	confident 3	very confident 4	not at all 1	little 2	considerable 3	very much 4
H1 National Constitution and State/ political institutions	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H2 Citizens' rights and obligations	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H3 Different conceptions of democracy	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H4 [Different/Comparative] political systems	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H5 Election and electoral systems	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H6 The judicial system	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H7 Human and civil rights	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H8 Important events in the nation's history	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H9 International organisations	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H10 International problems and relations	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H11 Migrations of people	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H12 Economic issues	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H13 Social Welfare	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H14 Trade/Labour Unions	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H15 Equal opportunities for women and men	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H16 Cultural differences and minorities	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H17 Environmental issues	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H18 Civic virtues	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H19 Dangers of propaganda and manipulation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
H20 Media	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

PART 4: INSTRUCTION

The following list presents activities that can be used in [civic related education].

Section I: How often are the following activities used in your classes?

Please indicate how frequently the following activities are used in your classes:

		never 1	sometimes 2	often 3	very often 4
I1	The teacher chooses the issues to be discussed in class	☐	☐	☐	☐
I2	Students work on projects that involve gathering information outside of school	☐	☐	☐	☐
I3	Students study textbooks	☐	☐	☐	☐
I4	Students work on drill sheets or work sheets	☐	☐	☐	☐
I5	Students work in groups on different topics and prepare presentations	☐	☐	☐	☐
I6	Students participate in role play and simulations	☐	☐	☐	☐
I7	The teacher asks questions and the students answer	☐	☐	☐	☐
I8	The teacher lectures [presents the subject] and the students take notes	☐	☐	☐	☐
I9	The teacher includes discussion on controversial issues in class	☐	☐	☐	☐
I10	Students participate in events or activities in the community (society)	☐	☐	☐	☐

Section J: In your view, what needs to be improved about civic education in your school?

*Select the **three** most important items listed below by checking the three appropriate boxes.*

We need...

J1   more materials and textbooks ☐

J2   better materials and textbooks ☐

J3   additional training in teaching methods ☐

J4   additional training in subject matter knowledge ☐

J5   more co-operation between teachers in different subject areas ☐

J6   more instructional time allocated to civic education ☐

J7   more co-operation with external experts ☐

J8   more opportunities for special projects ☐

J9   more resources for extra-curricular activities ☐

J10 more autonomy for school decisions ☐

Section K: How do you assess students?

*Please check appropriate box. Tick **two** boxes only!*

When I assess students in civic-related education, I primarily rely on...

K1 written compositions or essays ☐

K2 multiple-choice tests ☐

K3 oral assessments ☐

K4 oral participation ☐

K5 other forms of assessment: ☐
      please specify: _____

K6 no specific assessment ☐

F.5   SCHOOL QUESTIONNAIRE

1.   How many years will you have been a school principal by the end of this school year?

In your total career _____ years.

In your present school _____ years.

2.   In your school are there home-room [class] teachers who cover civic-related subjects?

No   ☐   1

Yes   ☐   2

3.   In your school are there teachers who specialize in a civic-related subject?

No   ☐   1

Yes   ☐   2

4.   Is this school participating in any special program(s) or project(s) related to Civic Education

No   ☐   1

Yes   ☐   2

If the answer is yes, please list the program(s) or project(s) this school is participating in:

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

5. Are the following organizations available for students to join in the school or community?

		No 1	Yes 2
a)	A student council/student government [class or school parliament]	☐	☐
b)	A youth organisation affiliated with a political party or union	☐	☐
c)	A group which prepares a school newspaper	☐	☐
d)	An environmental organisation	☐	☐
e)	A U. N. or UNESCO Club	☐	☐
f)	A student exchange or school partnership program	☐	☐
g)	A human rights organisation	☐	☐
h)	A group conducting [voluntary] activities to help the community	☐	☐
i)	A charity collecting money for a social cause	☐	☐
j)	Boy or Girl Scouts [Guides]	☐	☐
k)	A cultural association [organisation] based on ethnicity	☐	☐
l)	A computer club	☐	☐
m)	An art, music or drama organisation	☐	☐
n)	A sports organisation or team	☐	☐
o)	An organisation sponsored by a religious group	☐	☐

6. How should civic education be taught?

*Please give us **your personal opinion** and rate the statements below on the following scale:*
Civic education ...

		strongly disagree 1	disagree 2	agree 3	strongly agree 4
a)	should be taught in/as a specific subject	☐	☐	☐	☐
b)	should be taught integrated into subjects related to human and social sciences, like history, geography, languages, religion, ethics, law	☐	☐	☐	☐
c)	should be integrated into all subjects taught at school	☐	☐	☐	☐
d)	should be an extra-curricular activity	☐	☐	☐	☐

7. Which of the following statements holds true for students of this school?

*Please give us **your personal opinion** and rate the statements below on the following scale:*

		strongly disagree 1	disagree 2	agree 3	strongly agree 4
a)	Students in this school learn to understand people who have different ideas	☐	☐	☐	☐
b)	Students in this school learn to co-operate [work together] in groups with other students	☐	☐	☐	☐
c)	Students in this school learn to contribute to solving problems in the community [society]	☐	☐	☐	☐
d)	Students in this school learn to be a patriotic and loyal [committed] citizen of our country	☐	☐	☐	☐
e)	Students in this school learn how to act to protect the environment	☐	☐	☐	☐
f)	Students in this school learn to be concerned about what happens in other countries	☐	☐	☐	☐
g)	Students in this school learn about the importance of voting in national and local elections	☐	☐	☐	☐

8. In your school, how do parents become involved?

They...	never 1	sometimes 2	often 3
a) notify the school about learning problems of their children	☐	☐	☐
b) make sure that their child completes his/her homework	☐	☐	☐
c) raise and/or contribute funds other than tuition fees	☐	☐	☐

9. What is your best guess about the socioeconomic [background] of students in your schools:

   a) below average   _____%

   b) close to average   _____%

   c) above average   _____%

10. What is the total enrollment of full-time students in your school?

   Please, write in a number on each line.

   a) In target grade   _____ boys   _____ girls

   b) In school   _____ boys   _____ girls

   c) Average size of class in target grade   _____

   d) Average size of class in school   _____

11. Is your school ...

   a state school   ☐ 1

   a private school   ☐ 2

12. Please, indicate how frequently each of the following occurs at your school:

	never 1	sometimes 2	often 3
a) Vandalism	☐	☐	☐
b) Drugs	☐	☐	☐
c) Truancy	☐	☐	☐
d) Racism	☐	☐	☐
e) Religious intolerance	☐	☐	☐
f) Alcohol	☐	☐	☐
g) Bullying	☐	☐	☐
h) Violence	☐	☐	☐

13. How are the students in your school admitted?

They are assigned   ☐ 1

They have a choice   ☐ 2

Other   ☐ 3

please specify: _____

14. The following questions refer to different aspects of instructional time for Grade _____ [target grade]

   Please, write in a number on each line.

   How many instructional weeks are there in a school year?

   Weeks _____

   How many class periods are there in a school week?

   Hours _____

   How many instructional minutes are there in an average class period?

   Minutes _____

15. Which of the following grade levels are found in your school? [Country-specific Grade Description]

	no 1	yes 2
a.) Pre-kindergarten	☐	☐
b.) Kindergarten	☐	☐
c.) Grade 1	☐	☐
d.) Grade 2	☐	☐
e.) Grade 3	☐	☐
f.) Grade 4	☐	☐
g.) Grade 5	☐	☐
h.) Grade 6	☐	☐
i.) Grade 7	☐	☐
j.) Grade 8	☐	☐
k.) Grade 9	☐	☐
l.) Grade 10	☐	☐
m.) Grade 11	☐	☐
n.) Grade 12	☐	☐
o.) Grade 13	☐	☐

16. How many full-time (or full-time equivalent) teaching teachers are there in your school?

   Exclude non-teaching principal and administrators. If there are no male or no female teachers, please enter '0'.

   _____ male teachers.

   _____ female teachers.

17. How many hours (class period) per week are students required to take in...

a.) *History [subject 1]*

	Grade ___ [previous grade]	Grade ___ [target grade]
less than one hour	☐ 1	☐ 1
one to two hours	☐ 2	☐ 2
three to four hours	☐ 3	☐ 3
five to six hours	☐ 4	☐ 4

b.) *Social Sciences/Civic Education [subject 2]*

	Grade ___ [previous grade]	Grade ___ [target grade]
less than one hour	☐ 1	☐ 1
one to two hours	☐ 2	☐ 2
three to four hours	☐ 3	☐ 3
five to six hours	☐ 4	☐ 4

c.) *Law and Economics [International option] [subject 3]*

	Grade ___ [previous grade]	Grade ___ [target grade]
less than one hour	☐ 1	☐ 1
one to two hours	☐ 2	☐ 2
three to four hours	☐ 3	☐ 3
five to six hours	☐ 4	☐ 4