Measuring Civic Knowledge among Young Adolescents in the International Civic and Citizenship Education Study

Julian Fraillon
Australian Council for Educational Research
Melbourne/Australia
julian.fraillon@acer.edu.au

Wolfram Schulz
Australian Council for Educational Research
Melbourne/Australia
wolfram.schulz@acer.edu.au

John Ainley
Australian Council for Educational Research
Melbourne/Australia
john.ainley@acer.edu.au

Paper presented to the symposium on "Evaluating the Citizenship Competences of Young Adolescents in Europe: Methods, Patterns and Trends"
at the European Conference on Educational Research (ECER)
in Cádiz, 18-21 September 2012.

Introduction

Young people's acquisition of knowledge about, and their development of and understanding of key concepts in, civics and citizenship is widely seen as an important function of most education systems. From a national perspective, civic knowledge would typically encompass aspects related to the country's political system and its institutions, rights and responsibilities of citizens, mechanisms of engagement in society as well as general concepts (such as national identity, tolerance or shared values) (Schulz, Ainley, Fraillon, Kerr & Losito, 2010a & 2010b). Depending on the curriculum for civic and citizenship education in a country, it would be expected that these aspects could be defined in relation to the respective national context.

From an international perspective, it is necessary to identify the common aspects of civics and citizenship which are shared across different countries. Examples of such common aspects include the recognition of shared principles, concepts and participatory mechanisms and motivations. It needs to be recognized that commonalities such as similar forms of government (for example, having a democratic system of government) would increase the likelihood of identifying shared cognitive aspects of civics and citizenship across different countries. Furthermore, it needs to be acknowledged that it is not possible to include in international assessment instruments questions about country-specific issues that address, for example, details about national institutions, constitutional arrangements or local civic issues.

This paper describes how knowledge and understanding of civics and citizenship was assessed as part of the *International Civic and Citizenship Education Study* (ICCS 2009). ICCS 2009 was designed to investigate the ways in which young people in a range of countries are prepared and consequently ready and able to undertake their roles as citizens. Based on a comprehensive data collection in 38 countries from Europe, Latin America and the Asian-Pacific region the study reported on students' knowledge and understanding, value beliefs and attitudes, activities and intended behaviours related to civics and citizenship (Schulz et al., 2010a & 2010b; Kerr, Sturman, Schulz & Burge, 2010; Schulz, Ainley, Lietz & Friedman, 2011).

We will discuss previous research and the theoretical background for measuring civic knowledge and understanding, then describe the way it was measured in ICCS 2009 as well as methods employed to assess construct validity, and finally report selected outcomes of the assessment.

Conceptual Background

Prior research

In political science research among adult citizens, early Gallup surveys conducted in the United States included questions on factual political knowledge in (see Berelson, Lazarsfeld & McPhee, 1954). Political knowledge is recognised as a central measure of political sophistication (Converse, 1964; Luskin 1987) and regular assessments of political knowledge have been implemented in US National Electoral Surveys (NES) (see for example Delli Carpini & Skeeter, 1993) as well as European surveys (see for example De Vreese & Bomgarden). Research using data on political knowledge from American NES surveys among adult citizens has also drawn attention to some problems in measuring political knowledge through interviewer-based adult surveys (Mondak,

1999; Barabas, 2002).

Given the aim of education systems to provide students with sufficient knowledge and understanding to be prepared for their developing roles as citizens in society, attempts to measure the outcomes of national civic and citizenship education have become more frequent over the past decades. Examples are the national assessments of civics as part of the US National Assessment of Educational Progress (NAEP), which regularly tests students' knowledge of Civics and Citizenship at Grades 4, 8, and 12 (Lutkus, Weiss, Campbell, Mazzeo, & Lazer, 1999; Niemi & Junn, 1998; Torney-Purta, 2000), and the triennial Australian National Assessment Program on Civics and Citizenship, first conducted in 2004, which assesses Grades 6 and 10 students knowledge and understanding and reports against key performance measures for this learning area (Ministerial Council on Education, Employment, Training, and Youth Affairs, 2006 & 2008; Australian Curriculum and Assessment Reporting Authority, 2011).

Within the framework of comparative studies conducted by the International Association for Evaluation of Educational Achievement(IEA), there have been several studies of measuring cognitive aspects of civic and citizenship education in cross-national research. The IEA Civic Education Study of 1971 included a test with 47 multiple-choice items for 14- year-olds in nine countries (Torney, Oppenheim, & Farnen, 1975). The IEA CIVED survey, conducted in 1999, included a test with 38 multiple-choice items for 14-year-old students in 28 countries (Torney-Purta, Lehmann, Oswald, & Schulz, 2001) and a test with 42 multiple-choice items for 17- to 18-year-olds in 16 countries (Amadeo, Torney-Purta, Lehmann, Husfeldt, & Nikolova, 2002). The most recent international IEA study of civic and citizenship education, the International Civic and Citizenship Education Study (ICCS 2009) that forms the basis of this paper, included a test consisting of 80 items (74 multiple-choice and 6 constructed response items) to assess cognitive aspects of civic learning (Schulz et al., 2010a & 2010b).

The conceptual framework for assessing civics and citizenship in ICCS 2009

The conceptual model underpinning IEA CIVED 1999 placed at its centre the individual student who is influenced by "agents of socialisation" (Torney-Purta et al., 2001). It reflects the assertion that "civic learning" is not limited to school contexts but a result of diverse processes associated with different sources and evolved from the psychological theories of ecological development and situated cognition. The conceptual framework for ICCS 2009 (Schulz, Fraillon, Ainley, Losito & Kerr, 2008) also posited that young people learn about civics and citizenship through interactions with multiple civic communities and not just through formal instruction at school.

The ICCS 2009 *Civics and Citizenship Framework* articulated a conceptual underpinning for the collection of student outcome data. It was linked to the conceptual framework of CIVED 1999, which consisted of three domains (democracy/citizenship; national identity/international relations; social cohesion/diversity) but was formulated with a broadened scope. It is organised around three dimensions. It includes a content dimension specifying the subject matter to be assessed within civics and citizenship as well as a cognitive dimension outlining the thinking processes to be assessed. In addition, it also describes an affective-behavioural dimension which encompasses the types of student perceptions and activities that were measured with a student questionnaire (see Schulz, Losito & Kerr, 2011).

The civics and citizenship framework defines and elaborates four content domains, each of which is made up a set of sub-domains incorporating elements referred to as *aspects* and *key concepts*. Content domains describe elements of civics and citizenship that include both cognitive and affective-behavioural components. The first content domain, *civic society and systems*[1], comprises the mechanisms, systems, and organisations that underpin the operation of societies. The second domain, *civic principles*[2], relates to the shared ethical foundations of civic societies. The content domain *civic participation*[3] describes with the nature of the processes and practices that define and mediate the participation of citizens in their civic communities (often referred to as active citizenship). The centrality of the individual citizen is recognised through the inclusion of the *civic identities*[4] content domain, which refers to the personal sense an individual has of being an agent of civic action who may be member of multiple communities.

In order to respond to questions on the ICCS 2009 test, students needed to apply cognitive processes to content specified in the civic and citizenship content domains. In the context of ICCS 2009, *civic knowledge* was the term used to refer to demonstrable student achievement based on the application of the cognitive processes. From this point on we will use that term when referring to student test achievement in ICCS 2009. In order to support a complete mapping of the different cognitive aspects that underpin expressions of student civic knowledge, the ICCS Assessment Framework includes two cognitive domains.

The first cognitive domain, *knowing*, encompasses types of knowledge related to concrete and abstract concepts that can be generalised across societies. They include recalling or recognising definitions, descriptions, and key properties of civic and citizenship concepts and content.

The second domain, *reasoning and analysing*, details the cognitive processes required to reach conclusions that are broader than the contents of any single piece of knowledge, including the processes involved in understanding complex sets of factors influencing civic actions and planning for and evaluating strategic solutions and outcomes. Reasoning extends from direct applications of knowledge to reach conclusions about concrete situations through to the selection and assimilation of knowledge and understanding of multiple concepts in order to reach conclusions about complex, multifaceted, unfamiliar and abstract situations.

---

[1] The domain *civic society and systems* consists of three sub-domains: *citizens* (roles, rights, responsibilities, and opportunities), *state institutions* (central to civic governance and legislation), and *civil institutions* (these mediate citizens' contact with state institutions and allow citizens to pursue many of their roles in their societies).

[2] The domain *civic principles* consists of three sub-domains: *equity* (all people having the right to fair and just treatment), *freedom* (of belief, of speech, from fear, and from want), and *social cohesion* (sense of belonging, connectedness, and common vision held by individuals and communities within a society).

[3] The domain *civic participation* consists of three sub-domains: *decision-making* (organisational governance and voting), *influencing* (debating, demonstrating, developing proposals, and selective purchasing), and *community participation* (volunteering, participating in organisations, keeping informed).

[4] The domain *civic identities* consists of two sub-domains: *civic self-image* (individuals' experience of place in each of their civic communities) and *civic connectedness* (sense of connection to different civic communities; refers also to the civic roles individuals play within each community).

The conceptual framework for ICCS 2009 established the foundation for developing an instrument for assessing students' civic knowledge. As part of the development process it was crucial to ensure that (a) cognitive aspects could be generalised and were relevant across national contexts, (b) the test instrument was reliably measuring cognitive aspects across the ability range of target grade students in participating countries, (c) the items provided a meaningful description of proficiency levels associated with civic knowledge and (d) the measures were comparable across different contexts, cultures and languages.

Study design and methods

ICCS 2009 was carried out in 38 participating countries between October 2008 and May 2009. In each country approximately 150 schools were sampled depending on characteristics of the education system using PPS (*probability proportional to size* as measured by the number of students enrolled in each school) sampling procedures. Typically, one intact class was randomly selected within each sampled school. Student samples per country ranged from 3000 to 5000 students in the target grade[5].

This paper focuses on the international student test[6] which included 80 items in seven different clusters and was administered in complete rotated design with seven randomly allocated booklets, each consisting of three 15-minutes clusters.

The development of cognitive test item material was guided by the assessment framework and was carried out in stages that involved a large number of reviews and revisions. An interactive website was established so that invited experts and representatives of national research centres could review, revise and submit draft item material. This item material was also revised and discussed at meetings to ensure a high degree of cohesion in the measurement of civic knowledge across diverse national contexts. Revisions of the ICCS item material were informed by the analysis of pilot and field trial data as well as item panelling[7], focused discussions with pilot participants and expert reviews (see Fraillon, 2011).

The final ICCS civic knowledge test comprised 80 items, of which 79 items were used in the analysis.[8] Test items were typically presented as units in which some brief contextual stimulus (an image or some text) was followed by items relating to the common context. Seventy-three items were in multiple-choice format and six items were of constructed-response format. The latter required students to provide responses of between one and four sentences in length which were scored by expert scorers employed at the national centres.[9]

---

[5] The target grade corresponded to the eighth year of schooling provided that the minimum age of students was 13.5 years, in which case the ninth grade was selected.

[6] In addition to this test ICCS included an international student questionnaire (40 minutes length), a teacher questionnaire and a school questionnaire (completed by the Principal) as well as regionally specific instruments.

[7] Panelling is an exhaustive collaborative review of test items by expert test developers.

[8] One item showed insufficient measurement properties to warrant inclusion in the final set of items for analysis.

[9] The ICCS test of civic knowledge also included a link to the 1999 CIVED survey through the inclusion of 17 secure items from the CIVED item pool. The inclusion of these allowed us to

The test items were allocated to seven clusters that were assembled into a fully balanced rotated test design with seven test booklets, each with a testing time of 45 minutes. As for the field trial, one cluster comprised the set of secure CIVED items.

The newly developed ICCS test items and CIVED trend items were all mapped to the ICCS assessment framework. Table 1 shows the mapping of the final set of test items to the content and cognitive domains of the ICCS civics and citizenship framework.

<Insert Table 1 here>

Because the ICCS framework was broader than that for CIVED, most of the CIVED items were mapped to the content domain *civic society and systems* and the cognitive domain *knowing*. The newly developed ICCS items were written to complement the framework coverage of the CIVED items and consequently were mainly referring to the content domains of *civic principles*, *civic participation* and *civic identities* and the cognitive domain *reasoning and analyzing*. The relatively low proportions of items addressing the content domains *civic participation* and *civic identities* content was due to the fact that less of this kind of often complex conceptual content could reasonably be asked of students in this age group.

Item response modelling with the Rasch model (Rasch, 1960) was applied to assess the psychometric characteristics of items at different stage throughout the item development phase as well as to derive the cognitive scale (see Appendix A). To obtain more accurate population estimates, ICCS 2009 employed plausible value methodology which relies on all available information from student tests and questionnaires (von Davier, Gonzalez, & Mislevy, 2009). The final reporting scale was set to a metric that had a mean of 500 (the *ICCS average score*) and a standard deviation of 100 for the equally weighted national samples. The ACER Conquest, Version 2.0 software (Wu, Adams, Wilson, & Haldane, 2007) was used to scale the ICCS 2009 test data.

Translation and Verification of the ICCS Test Instrument

A critical set of processes in supporting the use of data collected using the ICCS 2009 instruments to make comparisons across countries relate to the adaptation, translation and verification of the instruments from their original English language source into the language of testing in each country. These processes were applied to all ICCS instruments and are described in detail in the ICCS 2009 Technical Report (Malak et. al., 2011). In addition to translations and review carried out by each national center, the ICCS test items and stimuli went through a rigorous three-part international verification process:

(i) negotiation of suggested adaptations such as place and people's names and the use of country-specific terms relating to civic institutions and other organisations.

(ii) independent verification of instrument translations; and

(iii) verification (by the ISC) of the layout of the test items and stimuli.

An independent review of the translation verification record was also conducted by international quality-control monitors, Together these processes aimed to minimise the

---

measure changes in performance for countries that participated in both ICCS 2009 and CIVED.

influence of translation on the meaning of the test item and stimulus contents to each student.

Psychometric characteristics of the ICCS 2009 test instrument

To establish a measure of civic knowledge, it was important to use test items that covered the different levels of achievement found in the target population. The population consisted of students in their eighth year of schooling across 38 different countries. Figure 1 shows the distribution of cognitive abilities among ICCS students (based on a representative sub-sample used for the final calibration[10]) and the location of items[11]. In addition, the plot shows the location of average student abilities in participating countries on the civic knowledge scale.

The range of item difficulties matched the abilities found in the student population. However, the average item difficulties were somewhat lower than the average student abilities. Overall, the test items were better at targeting students in the lower than in the higher civic knowledge ranges. The extent to which the test targeted student abilities varied across countries according to the distribution of student achievement within each country. Keeping in mind the variance of student achievement within countries, the test targeting for each country is summarised by the locations of average student abilities in each participating country.

<Insert Figure 1 here >

The overall reliability of the international test was 0.84[12]. Table 2 shows the median test reliabilities (Cronbach's alpha) across booklets for national samples. The within-country test reliabilities ranged from 0.70 to 0.88 and the median country reliability was 0.83. Only in six countries was the median reliability below 0.80.

<Insert Table 2 here>

Goodness of fit for individual items can be determined by calculating a mean square statistic (Wright & Masters, 1982) which provides an indication of the extent to which each item fits the item response model. Given the absence of clear rules for acceptable item fit and issues related with sole reliance on item fit statistics (see, for example, Rost & von Davier, 1994) the fit of ICCS 2009 test items was assessed using a broader range of item statistics including weighted mean square statistics, item-total correlations and item characteristic curves (see Schulz & Fraillon, 2011a & 2011b).

ICCS 2009 used open-ended test items in a cross-national assessment of civic knowledge. Student responses were scored according to scoring guides that were further refined based on experiences in the international field trial of test items. In addition to

---

[10] The calibration of student item parameters involved randomly selecting subsamples of 500 students from each national sample. This process ensured that each country that had met sample participation requirements was equally represented in the sample. The random selection was based on the final student weights, and the final calibration sample included data from 18,000 students.

[11] The response probability was p = 0.5. For a multiple-choice item this means that a student with a score equal to the item difficulty estimate was expected to have a 50 percent probability of giving the correct answer.

[12] An overall reliability estimate is an IRT-based measure provided by the ACER ConQuest software (see Wu et al., 2007).

assessing the general fit of these items, it was also necessary to monitor the reliability of scoring across participating countries. Within countries, for each of the seven booklets, subsamples of about 100 student records were scored twice by different scorers. This double- scoring procedure enabled the ICCS researchers to assess the inter-scorer reliabilities. On average, scorer agreement for the six items was between 85 and 89 percent.

As in other IEA studies, national data from open-response items only items for which there was a minimum of 70 percent agreement between scorers were retained for scaling and inclusion in the international database. This adjudication rule was applied separately for each open-response item scored in each country.

Assessing cross-national measurement equivalence

One important assumption for making comparisons in cross-national research is measurement invariance which typically is based on common parameters for scaling applied to a diverse range of populations. Measurement invariance is achieved if individuals obtaining the same score based on the same measurement instrument have the same standing on the underlying construct that is measured. In comparative international studies, it is usual to start from common source instruments that are subsequently adapted and translated for data collections in participating countries.

There is considerable evidence that language differences can have substantial effects on measurement equivalence. Whilst ICCS 2009 like most international studies (Yu, Malak, Schulz & Friedman, 2011; Chrostowski & Malak, 2004; Grisay, 2002) implemented rigorous translation verifications to achieve a maximum of "linguistic equivalence", research shows that even slight deviations in wording may lead to differences in item responses. Furthermore and of particular relevance in an assessment of civic-related knowledge, non-equivalence is also likely to be affected by the cultural differences among national contexts (van de Veijver & Tanzer, 1997; Byrne, 2003).

In order to investigate parameter invariance of test items (item-by-country interaction), national calibrations of item difficulties (δ, see appendix A) were compared with international item parameters. To make these sets of parameters comparable, the sum of item parameters was set to be 0. Confidence intervals for each national item parameter were computed based on their standard errors with an adjustment for design effects and for multiple comparisons.

Figure 2 shows the example of an "item-by-country interaction graph" for one ICCS 2009 test item. The figure shows considerable variation in the estimated item parameters across countries. The item was not used for the scaling of final test scores. Similar graphs were produced for each test item and taken into account when adjudicating test items at the international and national levels. Furthermore, information about occurrence of cross-national differential item functioning (DIF) was used to identify items for post-verification checks after completion of the main data collection.

<Insert Figure 2 here>

Generally, the ICCS 2009 test items showed only limited item-by-country interactions. However, in cases where national item difficulties deviated substantially from the international item difficulty the items were omitted from scaling for those national samples.

As in other international studies, it is interesting to review the extent to which variations across countries follow cultural or linguistic patterns. To this end, we conducted a hierarchical cluster analysis of item-by-country interactions. The analysis was limited to those multiple-choice items for which no translation errors in any countries had been detected.[13] Figure 3 shows the dendogram resulting from this hierarchical cluster analysis. The "tree" diagram represents the degree of similarity in item-by-country interactions between participating countries. The further to the left the lines from individual country clusters merge, the more similar were the patterns of deviations from the international calibration.

<Insert Figure 3 here>

The graph shows evidence of a primary clustering of item-by-country interactions by groups of countries with the same language family and similar geographical locations. Groupings can be seen for English-speaking, German-speaking and Spanish-speaking countries each had quite similar patterns of deviations from the international parameters. The analysis also shows a secondary clustering around geographical region that distinguishes Latin American countries and Spain from other European countries. The heterogeneity across the ICCS Asian regional countries was reflected in the lack of clear clustering by region or language shown in Figure 3.

It should also be noted that there were some exceptions from clustering by common languages and/or cultures. For example, the pattern of item-by-country interaction exhibited in Lithuania (which used three different languages for testing) appeared to more similar to Malta and the group of English-speaking countries. Furthermore, whilst there was considerable similarity between Norway and Sweden, the pattern item-by-country interactions in Denmark (where students also speak a Scandinavian language) seemed to be closer to Finland (where a majority of students speaks Finnish, a non-Scandinavian language).

The establishment of proficiency levels of civic knowledge

A described proficiency scale provides an indication of the nature of the construct, instantiates the conceptual development between lower and higher achievement and underpins this with an empirical basis for reporting achievement. The development of a described proficiency scale for ICCS 2009 was informed by a review of the content of items and empirical data about their relative difficulties. Item descriptors were ordered by item difficulty to produce an item map. Following an analysis of the item map and the distribution of student achievement, proficiency levels were established with a width of 84 scale points and level boundaries at 395, 479, and 563 scale points. Student scores below 395 scale points were identified as indicating civic knowledge below the level targeted by the assessment instrument.

The proficiency level descriptions are syntheses of the item descriptors within each level. They describe a hierarchy of civic knowledge in terms of increasing sophistication of content knowledge and cognitive process. Because the scale was derived empirically, increasing levels on the scale represent increasingly complex

[13] Open-ended items were not included in this analysis given that in two countries (Dominican Republic and Indonesia) systematic deviations in item difficulty were detected and therefore not included in the final scaling.

content and cognitive processes as demonstrated through performance. The scale does not extend from simple content at the bottom to reasoning and analyzing at the top. The cognitive processes of knowing and of reasoning and analyzing are evident across all levels of the scale.

Each level of the scale references the degree to which students appreciate the interconnectedness of civic systems, as well as the sense students have of the impact of civic participation on their communities. The scale broadly reflects development encompassing the concrete, familiar, and mechanistic elements of civics and citizenship through to the wider policy and institutional processes that determine the shape of our civic communities.

The scale is hierarchical in the sense that civic knowledge becomes more sophisticated as student achievement progresses up the scale. However, it is also developmental because of the assumption that any given student is probably able to demonstrate achievement of the scale content below his or her measured level of achievement.

Table 3 shows the described scale of civic knowledge developed for ICCS 2009 and provides descriptions of the scale's contents and the nature of the progression between the proficiency levels.

<Insert Table 3 here>

Level 1 of the scale is characterized by students' engagement with the fundamental principles and broad concepts that underpin civics and citizenship. Students operating at this level demonstrated familiarity with the "big ideas" of civics and citizenship; they were generally able to determine accurately what was fair or unfair in familiar contexts and to demonstrate some knowledge of the most basic operations of civic and civil institutions. Students working at Level 1 also typically demonstrated awareness of citizens' capacity to influence their own local context. The key factors that differentiate Level 1 achievement from that of higher levels relate to the degree of specificity of students' knowledge and the amount of mechanistic rather than relational thinking that students express in regard to the operations of civic and civil institutions. Internationally, 16 per cent of students were below this level and 26 per cent of students were at this level.

Students working at Level 2 typically demonstrated some specific knowledge and understanding of the most pervasive civic and citizenship institutions, systems, and concepts. These students demonstrated understanding of the interconnectedness of civic and civil institutions, and the processes and systems through which they operate (rather than only being able to identify their most obvious characteristics). Students working at Level 2 were also able to demonstrate understanding of the connection between principles or key ideas and how these operate in policy or practice in everyday, familiar contexts. They could relate some formal civic processes to their everyday experience and demonstrated awareness that the potential sphere of influence (and, by inference, responsibility) of active citizens lies beyond their own local context. One key factor differentiating Level 2 from Level 3 is the degree to which students use knowledge and understanding to evaluate and justify policies and practices. Thirty-one per cent of students were at this level.

Students working at Level 3 demonstrate a holistic rather than a segmented knowledge and understanding of civic and citizenship concepts. They made evaluative judgments about the merits of policies and behaviours from given perspectives, justified positions or propositions, and hypothesized outcomes based on their understanding of civic and

citizenship systems and practices. Students working at Level 3 demonstrated understanding of active citizenship practice as a means to an end rather than as an "automatic response" expected in a given context. These students were thus able to evaluate active citizenship behaviours in light of their desired outcomes. Twenty-eight per cent of students, across all countries recorded scores classified in this top level.

Variation of civic knowledge within and across countries

Table 4 shows the distribution of student achievement on the civic knowledge test for all countries. The national average scores ranged from 380 to 576 scale points. This range was equivalent to almost two between-student standard deviations which indicates considerable differences between participating countries.

<Insert Table 4 here>

The distributions of scores within each country can be seen in Table 4, where the length of the bars depicts the distribution of student scores within each country. The spread of scores within countries appeared to be unrelated to the average scale score for that county. In most countries, the distance between the lowest 5 percent and the highest 95 per cent of civic knowledge scores was around 300 scale points.

Fourteen countries had national averages that were significantly below the ICCS average, and 18 countries had national averages that were significantly higher than the international average. The average scale scores of four countries were not statistically significantly different from the ICCS 2009 average of 500 scale points. The difference between the bottom quartile and the top quartile (i.e., the area covering the middle half of the averages for countries) was 60 scale points—more than half a between-student standard deviation.

The four countries with the highest average scores—Finland, Denmark, the Republic of Korea, and Chinese Taipei—form a small group near the top of the scale. These countries cover a range of 17 scale points, which is followed by a gap of 22 scale points to the next country, Sweden. At the lower end of the scale, the average performance of students in the Dominican Republic is 44 scale points below that of Paraguay.

Table 4 also includes the Human Development Index (HDI) value for each country[14]. A strong association can be seen across the countries listed in Table 4 between HDI and average civic knowledge scale scores ($r = 0.75$). In addition, there is variation in the average age of students in the target grade (Grade 8) across countries. The average student age ranged from 13.7 to 15.5 years, although only a few countries were at the extreme ends of this range. The relationship between student age and civic knowledge scale scores varies within countries and across countries. Patterns in association between average student age across countries and average civic knowledge scale scores are superficially less clear than the patterns for HDI, partly because average student age across countries relates to local conditions, such as the age at which children begin school, and to student retention and progression rates, factors that may, in turn, be associated with HDI.

[14] The HDI, provided by the United Nations Development Programme (UNDP), is "a composite index measuring average achievement in three basic dimensions of human development including a healthy life, access to knowledge and a decent standard of living" (UNDP, 2009).

We conducted a regression analysis at the country level to assist interpretation of the relationship between average student age, HDI, and average civic knowledge scale scores across countries and to account for the potential interaction between HDI and student age as predictors of civic knowledge scale scores. The outcome variable in the analysis was the average ICCS 2009 civic knowledge scale score for each country; the average student age and the HDI for each country were predictors. Both HDI and student age were significantly positively associated with average civic knowledge scale scores. Across countries, one year of average student age was associated with an increase of 35 civic knowledge scale points whilst an increase of 0.1 in HDI corresponded to an increase of 54 civic knowledge scale points. Similar country-level results were found by Schulz, Fraillon, Ainley and van de gaer (2011) when using multi-level analysis of civic knowledge with ICCS 2009 data.

Conclusion

ICCS was concerned with measuring the extent to which students in 38 countries knew about and understood elements and concepts of citizenship as well as traditional civics. The key to this exercise was developing a conceptual framework that could establish the basis for an assessment instrument. The assessment instrument was based on four content domains (civic society and systems, civic principles, civic participation and civic identities). It assessed students' basic knowledge in these domains and the extent to which they could reason and analyse about issues based on that content. The focus of the test creation (including subsequent quality control in adapting and translating the test instrument for use within each participating country) was on establishing an instrument that addressed content relating to the underlying trait of civic knowledge and understanding that was both applicable (and accessible), and equivalent across the participating countries. Reviews of item dimensionality showed that the instrument measured a one-dimensional construct with high reliability. Reviews of measurement equivalence revealed patterns of item-by-country interaction that corresponded to cultural and linguistic differences. However, generally there was only limited variation in measurement characteristics across countries.

The measure of civic knowledge represented progress from limited knowledge and understanding of principles and concepts of civics and citizenship through to a deeper understanding of the wider policy climate and institutional processes shaping civic societies. Whilst students at the first proficiency level were characterized by a mechanistic knowledge of operations and institutions as well as engagement with fundamental principles and broader concepts, at the next higher level they demonstrated knowledge and understanding of the main civic and citizenship institutions, systems, and concepts as well as an appreciation of the interconnectedness of institutions and processes. At the highest level, students showed the capacity to apply knowledge and understanding to evaluate or justify policies, practices, and behaviours.

The results from ICCS 2009 revealed considerable variation among and within countries in the extent of civic knowledge. On a scale where the international mean was set to 500, and the standard deviation to 100, scale points, the average civic knowledge scores ranged from 380 to 576. This range was equivalent to almost two international student-level standard deviations. The difference between the bottom quartile and the top quartile of countries (which represents the middle half of the distribution of national averages) was about 60 scale points. There was even greater variation in civic knowledge within participating countries. This is illustrated by the fact that the distance

between the lowest five percent and the highest 95 percent of civic knowledge scores was typically about 300 scale points.

The development of the ICCS assessment framework, test instrument and achievement scale have demonstrated that it is possible to define, describe and to establish a common measure of a broad core of internationally comparable aspects of knowledge and understanding of civics and citizenship knowledge among 13-to-14-year olds. It is planned that this successful work will form the conceptual and measurement platform for ongoing and future research on the nature of learning outcomes of civic and citizenship education and their associations with contexts in which young people are prepared for their roles as citizens.

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).

Australian Curriculum and Assessment Reporting Authority (2011). *National Assessment Program Civics and Citizenship Year 6 and Year 10 Report 2010*. Sydney: Australian Curriculum and Assessment Reporting Authority.

Barabas, J. (2002). Another Look at the Measurement of Political Knowledge. Political Analysis, 10:2, 209-222.

Berelson, B. R., Lazarsfeld, P. F., & McPhee, W. N. (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.

Byrne, B. M. (2003). Testing for equivalent self-concept measurement across culture. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International Advances in Self-research: Speaking to the future* (pp. 291–314). Greenwich: Information Age Publishing.

Chrostowski, S. J. & Malak, B. (2004). Translation and Cultural Adaptation of the TIMSS 2003 Instruments. In M. O. Martin, I. V. S. Mullis & S. J. Chrostowski (Eds.) TIMSS 2003. *Technical Report* (pp. 93-108). Amsterdam: IEA.

Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. In: D. E. Apter (Ed.). *Ideology and Discontent* (pp. 206-261). New York: Free Press.

De Vreese, C. H. & Boomgaarden, H. (2006). News, Political Knowledge and Participation: The Differential Effects of News Media Exposure on Political Knowledge and Participation. *Acta Politica*, 41, 317-341.

Delli Carpini, M. X. & Keeter, S. (1993). Measuring Political Knowledge: Putting First Things First, *American Journal of Political Science*, 37:4, 1179-1206.

Fraillon, J. (2011). ICCS test development. In W. Schulz, J. Ainley, J., & J. Fraillon (Eds.). *ICCS 2009 Technical Report* (pp. 21-31). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Grisay, A. (2002). Translation and Cultural Appropriateness of the Test and Survey Material. In R. J. Adams & M. Wu (Eds.). *PISA 2000. Technical Report* (pp. 57-70). Paris: OECD Publications.

Kerr, D., Sturman, L., Schulz, W., & Bethan, B. (2010). *ICCS 2009 European Report. Civic knowledge, attitudes and engagement among lower secondary school students in twenty-four European countries*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Luskin, R. C. (1987). Measuring Political Sophistication. *American Journal of Political Science*, 31:4, 856-899.

Lutkus, A. D., Weiss, A. R., Campbell, J. R., Mazzeo, J. and Lazer, S. (1999). *The NAEP Civics Report Card for the Nation.* Washington, D.C.: National Center for Education Statistics.

Malak, B, Yu, A, Schulz, W and Friedman, T (2011). Translation and national adaptations of ICCS 2009 instruments. In W. Schulz, J. Ainley, J., & J. Fraillon (Eds.). ICCS 2009 Technical Report (pp. 51-58). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Masters, G. N. & Wright, B. D. (1997). The Partial Credit Model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 101-122), New York: Springer.

Ministerial Council on Education, Employment, Training, and Youth Affairs (2006). *National Assessment Program. Civics and Citizenship Years 6 and 10 Report 2004.* Carlton South: Ministerial Council on Education, Employment, Training and Youth Affairs.

Ministerial Council on Education, Employment, Training, and Youth Affairs (2008). *National Assessment Program. Civics and Citizenship Years 6 and 10 Report 2007.* Carlton South: Ministerial Council on Education, Employment, Training and Youth Affairs.

Mondak, J. J. (1999). Reconsidering the Measurement of Political Knowledge. *Political Analysis*, 8:1, 57-82.

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

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen & Lydiche.

Rost, J., & von Davier, M. (1994). A conditional item-fit index for Rasch models. *Applied Psychological Measurement*, 18(2), 171–182.

Schulz, W. & Fraillon, J. (2011a). Scaling procedures for ICCS test items. In W. Schulz, J. Ainley, J., & J. Fraillon (Eds.). *ICCS 2009 Technical Report* (pp. 129-156). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W. & Fraillon, J. (2011b). The Analysis of Measurement Equivalence in International Studies using the Rasch Model. *Educational Research and Evaluation*, 17(6), 447-464.

Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010a). *Initial Findings from the IEA International Civic and Citizenship Study*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010b). *ICCS 2009 International Report: Civic knowledge, attitudes, and engagement among lower secondary school students in thirty-eight countries*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Ainley, J., Friedman, T. & Lietz, P. (2011). *ICCS 2009 Latin American Report: Civic knowledge and attitudes among lower secondary students in six Latin American countries*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Fraillon, J., Ainley, J. & van de gaer, E. (2011). *Multi-level analysis of factors explaining differences in civic knowledge.* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in New Orleans, 8-12 April.

Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008) *International Civic and Citizenship Education Study. Assessment Framework.* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Schulz, W., Losito, B., & Kerr, D. (2011). ICCS questionnaire development. In W. Schulz, J. Ainley, J., & J. Fraillon (Eds.). ICCS 2009 Technical Report. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

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

Torney-Purta, J. (2000) An international perspective on the NAEP Civics Report Card, *The Social Studies*, 94, 148–50.

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 (IEA).

van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 47, 263-279.

von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, Volume 2, 9–36.

Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. (2007). ACER ConQuest 2.0: General item response modelling software [computer program manual]. Camberwell, VIC: ACER Press.

Yu, A., Malak, B., Schulz, W., & Friedman, T. (2011). Translation and national adaptations of ICCS 2009 instruments. In W. Schulz, J. Ainley, J., & J. Fraillon (Eds.). *ICCS 2009 Technical Report* (pp. 51-58). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Appendix A

When using the one-parameter (Rasch) model (Rasch, 1960) for dichotomous items, a test respondent's probability of selecting Category 1 instead of 0 is modelled as

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

where $P_i(\theta)$ is the probability for respondent *n* to score 1 on item *i*, θ_n is the estimated ability of respondent *n*, and δ_i is the estimated location of item *i* on this dimension. For each item, item responses are modelled as a function of the latent trait θ_n.

In the case of items with more than two categories *k+1* (as, for example, with Likert-type items), this model is generalised as the partial credit model (Masters & Wright, 1997), which takes the form of

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

Here, $P_{xi}(\theta)$ denotes the probability of respondent *n* scoring *x* on item *i* and θ_n denotes their ability. The item parameter δ_i gives the location of the item on the latent continuum and τ_{ij} denotes an additional parameter for each step j between adjacent categories.

Table 1: Main study item mapping to assessment framework

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

Figure 1: Mapping of student abilities, item difficulties and country locations

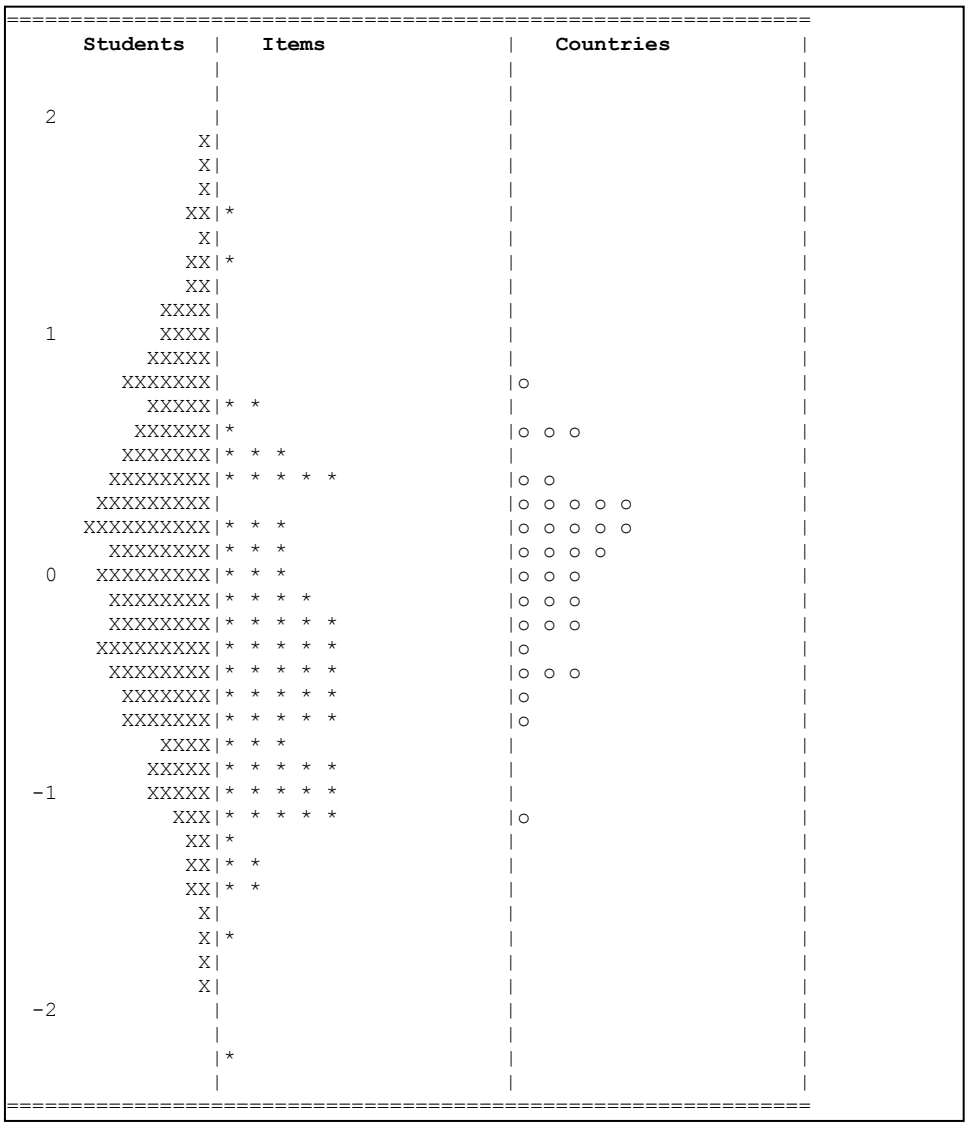


Table 2: Median test reliabilities across booklets for national samples

Country	Median test reliability across booklets	Median number of items per booklet
Austria	0.86	32
Belgium (Flemish)	0.82	32
Bulgaria	0.88	32
Chile	0.83	32
Chinese Taipei	0.83	31
Colombia	0.81	32
Cyprus	0.84	32
Czech Republic	0.81	32
Denmark	0.84	32
Dominican Republic	0.70	29
England	0.87	32
Estonia	0.84	32
Finland	0.81	32
Greece	0.87	32
Guatemala	0.78	32
Hong Kong SAR	0.84	30
Indonesia	0.72	27
Ireland	0.87	32
Italy	0.82	32
Korea, Republic of	0.77	27
Latvia	0.78	29
Liechtenstein	0.85	32
Lithuania	0.80	32
Luxembourg	0.86	31
Malta	0.85	30
Mexico	0.82	32
Netherlands	0.85	32
New Zealand	0.88	32
Norway	0.84	32
Paraguay	0.84	32
Poland	0.85	32
Russian Federation	0.82	32
Slovak Republic	0.82	32
Slovenia	0.83	32
Spain	0.84	32
Sweden	0.85	32
Switzerland	0.83	32
Thailand	0.78	31
ICCS median	0.83	32

Table 3: ICCS 2009 proficiency levels with descriptions

Level 3: 563 score points and above

Students working at Level 3 make connections between the processes of social and political organisation and influence, and the legal and institutional mechanisms used to control them. They generate accurate hypotheses on the benefits, motivations, and likely outcomes of institutional policies and citizens' actions. They integrate, justify, and evaluate given positions, policies or laws based on the principles that underpin them. Students demonstrate familiarity with broad international economic forces and the strategic nature of active participation.

Students working at Level 3, for example:

- Identify likely strategic aims of a program of ethical consumption
- Suggest mechanisms by which open public debate and communication can benefit society
- Suggest related benefits of widespread cognitive intercultural understanding in society
- Justify the separation of powers between the judiciary and the parliament
- Relate the principle of fair and equal governance to laws regarding disclosure of financial donations to political parties
- Evaluate a policy with respect to equality and inclusiveness
- Identify the main feature of free market economies and multinational company ownership.

Level 2: 479 to 562 score points

Students working at Level 2 demonstrate familiarity with the broad concept of representative democracy as a political system. They recognise ways in which institutions and laws can be used to protect and promote a society's values and principles. They recognise the potential role of citizens as voters in a representative democracy, and they generalize principles and values from specific examples of policies and laws (including human rights). Students demonstrate understanding of the influence that active citizenship can have beyond the local community. They generalize the role of the individual active citizen to broader civic societies and the world.

Students working at Level 2, for example:

- Relate the independence of a statutory authority to maintenance of public trust in decisions made by the authority
- Generalizes the economic risk to developing countries of globalization from a local context
- Identify that informed citizens are better able to make decisions when voting in elections
- Relate the responsibility to vote with the representativeness of a democracy
- Describe the main role of a legislature/parliament
- Define the main role of a constitution
- Relate the responsibility for environmental protection to individual people.

Level 1: 395 to 478 score points

Students working at Level 1 demonstrate familiarity with equality, social cohesion, and freedom as principles of democracy. They relate these broad principles to everyday examples of situations in which protection of or challenge to the principles are demonstrated. Students also demonstrate familiarity with fundamental concepts of the individual as an active citizen: they recognise the necessity for individuals to obey the law; they relate individual courses of action to likely outcomes; and they relate personal characteristics to the capacity of an individual to effect civic change.

Students working at Level 1, for example:

- Relate freedom of the press to the accuracy of information provided to the public by the media
- Justify voluntary voting in the context of freedom of political expression
- Identify that democratic leaders should be aware of the needs of the people over whom they have authority
- Justifies voluntary voting in the context of freedom of political expression
- Recognize that the UN Universal Declaration of Human Rights is intended to apply to all people
- Generalize about the value of the internet as a communicative tool in civic participation
- Recognise the civic motivation behind an act of ethical consumerism.

Figure 2: Example of item-by-country interaction graph

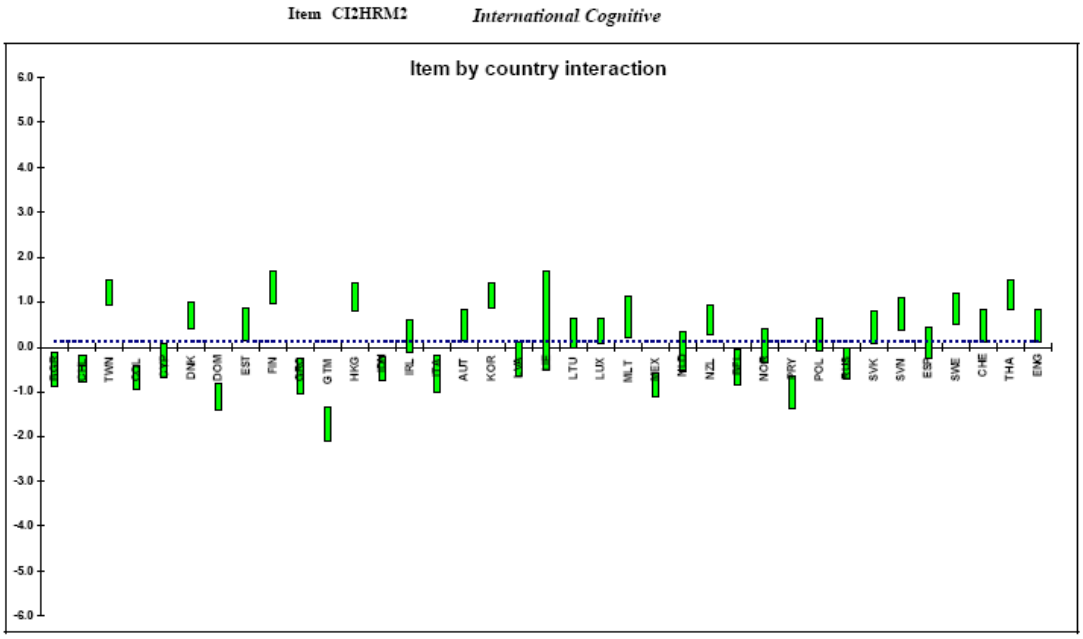


Figure 3: Dendogram of hierarchical cluster analysis of item-by-country interactions

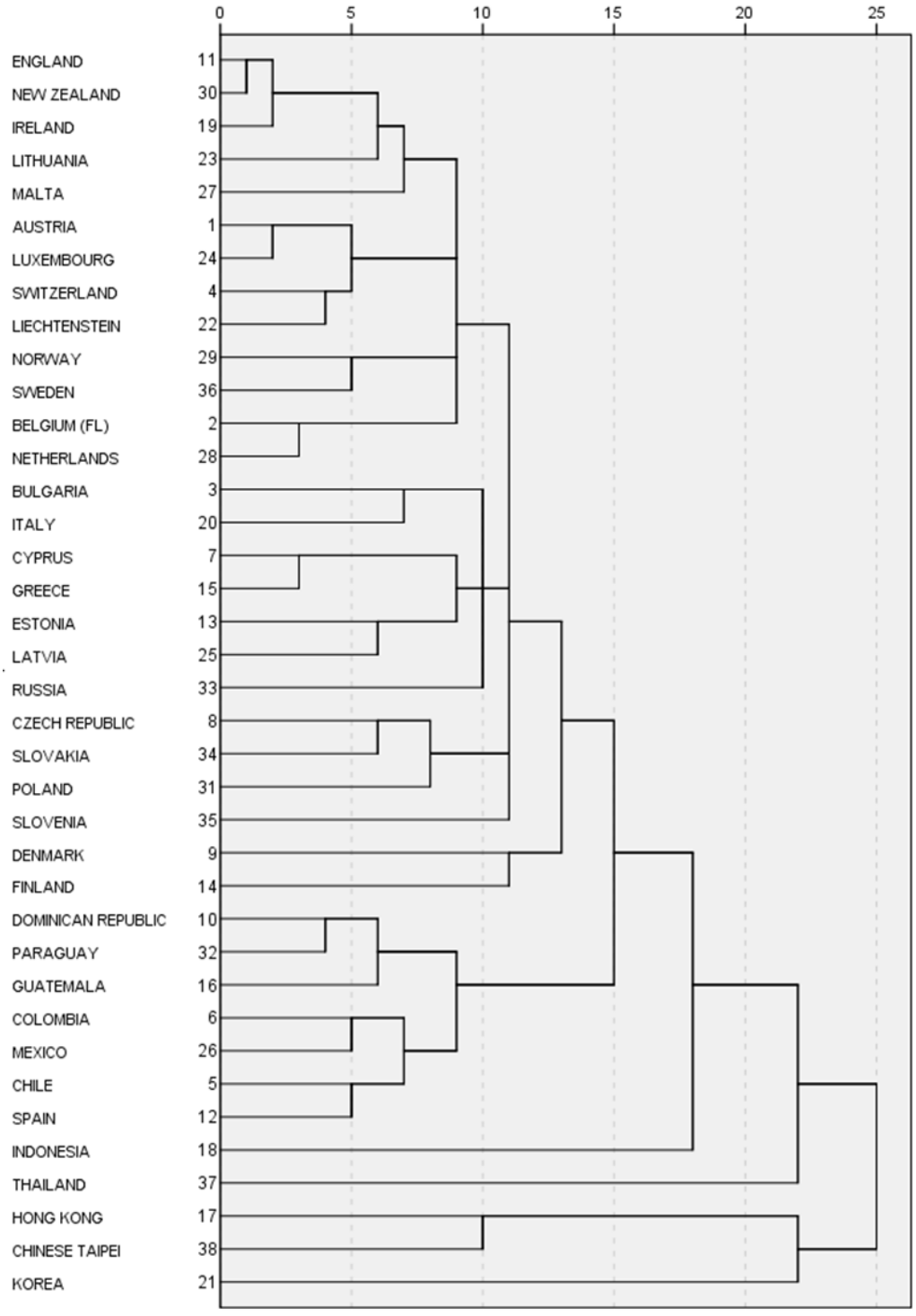
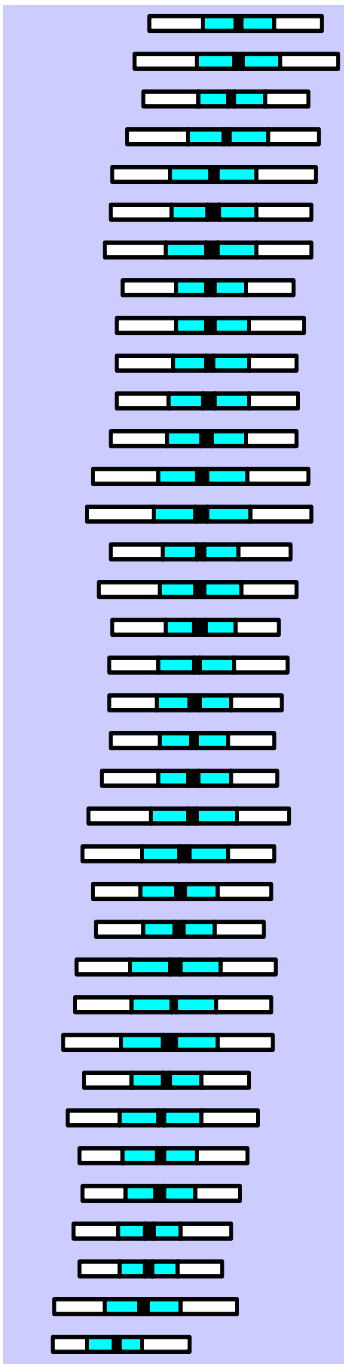


Table 4: National averages of civic knowledge

Country	Years of schooling	Average Age	Civic Knowledge 200 300 400 500 600 700 800	Average Scale Score		HDI
Finland	8	14.7		576 (2.4)	▲	0.96
Denmark †	8	14.9		576 (3.6)	▲	0.96
Korea, Republic of[1]	8	14.7		565 (1.9)	▲	0.94
Chinese Taipei	8	14.2		559 (2.4)	▲	0.94
Sweden	8	14.8		537 (3.1)	▲	0.96
Poland	8	14.9		536 (4.7)	▲	0.88
Ireland	8	14.3		534 (4.6)	▲	0.97
Switzerland †	8	14.7		531 (3.8)	▲	0.96
Liechtenstein	8	14.8		531 (3.3)	▲	0.95
Italy	8	13.8		531 (3.3)	▲	0.95
Slovak Republic[2]	8	14.4		529 (4.5)	▲	0.88
Estonia	8	15.0		525 (4.5)	▲	0.88
England ‡	9	14.0		519 (4.4)	▲	0.95
New Zealand †	9	14.0		517 (5.0)	▲	0.95
Slovenia	8	13.7		516 (2.7)	▲	0.93
Norway †	8	13.7		515 (3.4)	▲	0.97
Belgium (Flemish) †	8	13.9		514 (4.7)	▲	0.95
Czech Republic †	8	14.4		510 (2.4)	▲	0.90
Russian Federation	8	14.7		506 (3.8)		0.82
Lithuania	8	14.7		505 (2.8)		0.87
Spain	8	14.1		505 (4.1)		0.96
Austria	8	14.4		503 (4.0)		0.96
Malta	9	13.9		490 (4.5)	▼	0.90
Chile	8	14.2		483 (3.5)	▼	0.88
Latvia	8	14.8		482 (4.0)	▼	0.87
Greece	8	13.7		476 (4.4)	▼	0.94
Luxembourg	8	14.6		473 (2.2)	▼	0.96
Bulgaria	8	14.7		466 (5.0)	▼	0.84
Colombia	8	14.4		462 (2.9)	▼	0.81
Cyprus	8	13.9		453 (2.4)	▼	0.91
Mexico	8	14.1		452 (2.8)	▼	0.85
Thailand †	8	14.4		452 (3.7)	▼	0.78
Guatemala[1]	8	15.5		435 (3.8)	▼	0.70
Indonesia	8	14.3		433 (3.4)	▼	0.73
Paraguay[1]	9	14.9		424 (3.4)	▼	0.76
Dominican Republic	8	14.8		380 (2.4)	▼	0.78