

## **Cross-National Equivalence of Students' Perceptions of Good Citizenship in ICCS**

**Dorota Weziak-Bialowolska,**

European Commission – Joint Research Centre, Ispra, Italy

**Maria Magdalena Isac,**

European Commission – Joint Research Centre, Ispra, Italy

### **Abstract**

The study reported in this article assesses whether two concepts measured in the International Civic and Citizenship Education Study (ICCS) 2009, namely (1) student's perceptions of the importance of conventional citizenship and (2) student's perceptions of the importance of social-movement-related citizenship, are comparable across all countries participating in the study. In particular by means of confirmatory multi-group factor analysis we tested for invariance of measurements across countries which can ensure that meaningful comparisons can be conducted. The results show a complex picture indicating that the concepts are not necessarily comparable across all countries involved in ICCS and their use for secondary data analysis can be more or less feasible across all countries depending on the research questions and the methodology applied. Scientific and practical implications are discussed.

### **Key words:**

Citizenship values, measurement invariance, concept equivalence, factor analysis, ICCS

## **1. Introduction**

The increasing availability and complexity of cross-national datasets from international large-scale assessment studies in education provide researchers with the possibility of investigating a numerous number of research questions in a comparative perspective. However, such studies can be valuable only if the researchers are well aware of some of the limitations that the studies might have and are able to apply the appropriate methodology. By stating this, we mean, among others, the assessment of measurement invariance of questionnaire data enabling comparisons of constructs across countries.

An example of such a large-scale assessment study in education, which has not received enough attention with regard to cross-country construct equivalence, is the International Civic and Citizenship Education Study (ICCS) 2009 carried out by the International Association for the Evaluation of Educational Achievement (IEA). The aim of the ICCS is to investigate “the ways in which young people are prepared to undertake their roles as citizens in democracy” (Schulz et al., 2010, p.13) in a range of countries. Because it benefited from the experience gathered in two previous international IEA civic education studies conducted in 1971 (Torney, Oppenheim, & Farnen, 1975) and 1999 Civic Education Study (CIVED) (Torney-Purta, Lehmann, Oswald, & Schulz, 2001) and was guided by broad theoretical models (see Torney-Purta et al., 2001; Schulz et al., 2010), the ICCS conceptual framework (Schulz, Fraillon, Ainley, Losito, & Kerr, 2008) grew to assess a numerous number of concepts. The main assumption was that young people learn about civics and citizenship through their interactions with multiple communities. Therefore the study gathered information not only regarding student civic competences (outputs) but also contextual factors that map the characteristics of individual students, their learning experiences in the family, the school and the wider community. As a result, the ICCS study is currently the most comprehensive cross-national source of information regarding civic and citizenship education and student civic outcomes. The number of elements measured enables researchers to formulate and test several hypotheses with the aim of looking particularly at civic competences and civic learning but not only. Secondary analysis can address several other topics such as classroom climate, school resources, school governance etc.

Since the very beginning stages in the study development, the ICCS team paid serious attention to guaranteeing methodological quality (e.g. employing a rigorous sampling strategy, consulting international expert groups for instrument development, conducting pilot and field trial studies) (Schulz, 2009; Schulz et al. 2010). Moreover, after the ICCS data was collected, processed and released, the IEA technical report (Schulz, Ainley, & Fraillon, 2011) provided extensive information to guide future decisions. Nevertheless, although cross-country validity of some constructs was assessed during the field trial stage of ICCS (see Schulz et al. 2011, p. 157) to our knowledge, the ICCS documentation available at the moment does not provide sufficient information regarding the cross-country comparability of all concepts and measures. The implication is that most scales in ICCS are still to be validated in order to compare constructs with some confidence across countries. It is, therefore, up to the researcher to perform tests of measurement invariance of theoretical constructs and establish whether meaningful comparisons can be carried out. Moreover, this is necessary not only in order to conduct a series of country comparisons with regard to construct scale but also to estimate different types of statistical models with latent constructs across countries.

Therefore, the aim of this paper is to assess the cross-country equivalence of two concepts measured in ICCS across all countries participating in the study. It is carried out within a factor analytical framework. We want to stress, however, that these are just examples and cross-country equivalence should be examined regarding all other concepts/instruments used in international large-scale-assessments when one desires to compare several countries.

In our study, we have chosen to investigate two concepts referring to citizenship values or student perceptions of “good citizenship” behavior namely: (1) student’s perceptions of the importance of conventional citizenship and (2) student’s perceptions of the importance of social-movement-related citizenship. We have selected these concepts because they receive much attention in international comparative studies and are of particular interest for contemporary educational policies and practices (Sherrod, Torney-Purta, & Flanagan, 2010; Hoskins, Barber, Van Nijlen, & Villalba, 2011; Hoskins, Villalba, & Saisana, 2012; Schulz et al. 2010; Torney-Purta et al. 2001; Torney-Purta, 2002; Torney-Purta, & Richardson, 2004; Hoskins, Janmaat, & Villalba, 2012; Isac, Maslowski, Creemers & Van der Werf, 2013; Torney-Purta, & Richardson, 2004). To the best of our knowledge, so far, there have not been any studies on cross-country equivalence of the citizenship values concepts, we have chosen. Furthermore, as noted by the

experts involved in the design and data processing of the ICCS study (see Schulz, 2008), the assessment of measurement invariance of questionnaire data has not received enough attention with the implication that secondary data analysis might result in comparisons of constructs that are not necessarily appropriate given the properties of the data.

In what follows we describe the conceptualization and operationalization of the concepts chosen for this exercise. They are followed by an account of the available data, the methodology of the measurement invariance assessment used and the analysis strategy. Further on, we report our results and formulate the main conclusions as well as discussing scientific and practical implications.

## **2. Citizenship values**

Citizenship values, often regarded as learning outcomes of citizenship education programs, are considered important determinants of social and political participation in adulthood and therefore receive extensive attention in research on the development of citizenship (see Sherrod, Torney-Purta, & Flanagan, 2010). They are also frequently employed in cross-country comparisons of young people's perceptions of the importance of conventional and social-movement citizenship in the field of youth civic engagement (see Barber & Torney-Purta, 2012; Sherrod et al., 2010). The focus is on monitoring such outcomes for research and educational policy purposes (e.g. Hoskins, Barber, Van Nijlen, & Villalba, 2011; Hoskins, Villalba, & Saisana, 2012; Schulz et al. 2010; Torney-Purta et al. 2001; Torney-Purta, 2002; Torney-Purta, & Richardson, 2004) and/or on examining the individual, contextual and school-related determinants of citizenship values development often with a multilevel analysis methodology (e.g. Hoskins, Janmaat, & Villalba, 2012; Isac, Maslowski, Creemers & Van der Werf, 2013; Torney-Purta, & Richardson, 2004). In both cases, the underlying assumption is that the concepts have similar meanings in the analyzed countries, however it is not empirically verified.

The citizenship values measured in ICCS are conceptualized as student perceptions of different types of “good citizenship” behavior (see Schulz et al. 2008; Schulz et al. 2010; Torney-Purta et al. 2001) based broadly on Almond and Verba's (1963) description of political culture. Dalton (2006), quoted in Schulz et al. (2010), describes two facets of good citizenship depending on whether the types of endorsed behaviors are driven by a sense of “citizen duty” or by elements of

liberal or communitarian norms of citizenship. Citizenship values are accordingly conceptualized as two interrelated sets of values capturing beliefs regarding “good citizenship” behavior manifested in both conventional and social-movement forms of participation. Conventional participation relates to citizen behavior associated with compliance with social norms or democratic duties such as voting in national elections, joining a political party or following political issues in the media. On the other hand, social movement participation is concerned with more active forms of participation in society such as participating in peaceful protests, in activities benefiting the community or promoting human rights (see Schulz et al., 2010). These sets of values are not necessarily contradictory but can reflect contrasting emphasis on the role of a democratic citizen with younger people more likely to endorse the later value category (see Dalton, 2006; Torney-Purta et al. 2001; Schulz et al. 2010).

### **3. Data**

All the data selected for the analysis comes from the IEA-ICCS study 2009 which was conducted in 38 European, Asian, South and Central American and Oceanic countries. In this study two sets of items measuring *student citizenship values* were used: students’ perceptions of conventional (CITCON) (6 statements) and social-movement-related (CITSOC) citizenship (4 statements) (see Table 1). The data come from the student survey (14-year-olds). The sample size varies between 357 in Lichtenstein and 1,964 in the Netherlands to 6,576 in Mexico with a total of 140,650 across all 38 countries. Preliminary analysis based on the descriptive statistics led us to the decision of excluding Lichtenstein due to small sample size.

Table 1 illustrates the two sets of items measuring citizenship: students’ perceptions of good conventional (CITCON) (6 statements) and social-movement-related (CITSOC) citizenship (4 statements). The items were included in the ICCS student background questionnaire (see Schulz et al. 2011) and therefore rated by 14-years-olds. Each of them was measured on a 4 point Likert scale ranging from “very important” to “not important at all”. The corresponding items were used to derive two scales in ICCS (see Schulz et al. 2011, p. 174). The measurement qualities of the scales were estimated by means of confirmatory factor analysis on the pooled dataset (Schulz et al. 2011, p. 176).

Table 1. Student citizenship values – items codes and wording

<i>How important are the following behaviors for being a good adult citizen?</i>	<b>Conventional Citizenship (CITCON)</b>		
	<b>IS2P21A</b> Voting in every national election	<i>1. Very important 2. Quite important 3. Not very important 4. Not important at all</i>	
	<b>IS2P21B</b> Joining a political party		
	<b>IS2P21C</b> Learning about the country's history		
	<b>IS2P21D</b> Following political issues in the newspaper, on the radio on TV or on the internet		
	<b>IS2P21E</b> Showing respect for government representatives		
	<b>IS2P21F</b> Engaging in political discussions		
	<b>Social-movement-related citizenship (CITSOC)</b>		
	<b>IS2P21G</b> Participating in peaceful protests against laws believed to be unjust		
	<b>IS2P21H</b> Participating in activities to benefit people in the <local community>		
<b>IS2P21I</b> Taking part in activities promoting human rights			
<b>IS2P21J</b> Taking part in activities to protect the environment			

#### 4. Testing for measurement invariance

In order to compare country scores on a scale, it is necessary to establish the cross-country comparability of the scales. By cross-country construct comparability, which is also defined as measurement invariance (MI), we understand that scale scores from different countries measure the same construct with the same measurement unit (Wu et al., 2007).

Davidov (2008) reports that there are three major techniques of measurement invariance testing, namely differential item functioning approach – which can be found for instance in Jansen (2011), Item Response Theory Models –in De Jong et al. (2007) and factor analysis framework – in Davidov (2008), Davidov et al. (2008), Gregorich (2006) and Wu et al. (2007), among others. However, the most frequently used one is multi-group confirmatory factor analysis, which is also applied in our analysis.

Meredith (1993) distinguishes four level of measurement invariance: (1) configural invariance, (2) weak invariance, (3) strong invariance, and (4) strict invariance. To these four types of MI, Gregorich (2006) adds another one that should precede them – dimensional invariance, i.e. the verification of the number of factors in each group in an exploratory factor analysis framework. Byrne (2008) proposes another classification. She classifies weak and strong invariances as measurement equivalence and strict invariance as structural equivalence.

In the factor analysis framework, configural invariance means that the same factor model is specified across compared groups (Horn and Mcardle, 1992). The weak invariance, called also

pattern invariance (Meredith and Teresi 2006) or metric factorial invariance (Davidov 2008; Meredith & Teresi 2006; de Jong et al. 2007) requires invariant factor loadings across groups. The strong measurement invariance, called also strong factorial invariance (Meredith & Teresi 2006) or scalar invariance (Davidov, 2008; de Jong et al., 2007), requires that factor intercepts are identical across groups. The strict measurement invariance requires that, in addition to equal factor loadings and factor intercepts, the manifest variable residuals are equal across groups. Taking into regard the typology of the MI, the process of its establishing is clearly hierarchical. It starts with establishing the well-fitting baseline model for each group separately and then proceeds with testing subsequent types of invariance (Byrne, 2008). Establishing: (i) configural invariance ensures that common factors are associated with the same items across groups, (ii) weak measurement invariance ensures that the common factors have the same meaning across groups and the same measurement unit, (iii) strong measurement invariance enables meaningful comparisons of the group means as the factors have both the same measurement unit and the same reference point, (iv) strict measurement invariance means that comparisons of group means and observed variances are defensible (Gregorich, 2006).

In other words, when the researcher aims at comparing across countries the relations among factor or factors and other observable variables, the weak measurement invariance suffices. However, when she wants to perform valid cross-country comparisons of scale scores of construct of interest (comparisons of factor means), the strong measurement invariance should be first established. However, it must be noted that there is no common agreement on which type of measurement invariance is sufficient to compare factor means. Meredith (1993), Wu, Li, and Zumbo (2007), Lubke & Dolon, (2003) strongly claim that it is the strict measurement invariance, whereas according to Davidov (2008), Davidov et al. (2008) and Byrne & van de Vijver (2010) meaningful information can be obtained assuming only strong measurement invariance. Byrne & Van de Vijver (2010) claim also that there is a widespread consensus about the fact that testing for strict equivalence not only is of least importance, but also somewhat unreasonable and, referring to the paper of Selig et al. (2008), not recommended. All types of invariance can be verified either in full or partial version (Baumgartner and Steenkamp, 1992; Byrne, 2008; Byrne et al., 1989; de Jong et al., 2007; Gregorich, 2006; Millsap and Kwok, 2004). In the full version of invariance, equality constraints concern all manifest variables, whereas in the partial version some of them can be relaxed. It means that only the subset of items

meeting the metric, strong, or strict factorial invariance criteria are used to estimate group differences, which under the conditions of partial invariance, provide substantive and defensible information (Gregorich, 2006). The specification of partial measurement invariance is commonly used particularly in the area of cross-cultural research (Byrne and Van de Vijver, 2010).

When the measurement invariance is not satisfied, subgroups of countries have to be found that are measurement invariant (Welkenhuysen-Gijbels et al., 2003). This approach is particularly valid in the cross-cultural equivalence studies, where (i) the construct of interest may be structurally and psychometrically inappropriate or (ii) the clusters of countries may exhibit both within-cluster homogeneity and between-cluster heterogeneity (Byrne and Van de Vijver, 2010).

## **5. Modeling Strategy**

In our analysis to check the measurement invariance properties we use a multi-group factor analytical framework and we check for four levels of the measurement invariance – configural, weak (metric), strong (scalar) and strict – going from the least restrictive to the most restrictive model. It is noteworthy that our imperative is to establish the configural invariance first (see Byrne and van de Vijver 2010). Only if it holds, do we proceed to checking higher levels of equivalence.

All analyses are performed under the full measurement equivalence framework only. We run the analysis for the whole group of countries participating in the ICCS, but also for subgroups of countries. The countries are grouped based on the geographical location (see table Table 8 in Appendix). We distinguish European, South and Central American, and Asian (including New Zealand) countries. Having repeated the analyses for Asian countries excluding New Zealand, any substantive change is recorded. When the model including all countries fits well, the analyses for sub-groups of countries are avoided.

Since the conceptualization of the citizenships values concepts is well established and there is a common agreement that conventional citizenship and social-movement related citizenships constitute two correlated one-dimensional concepts, the step on establishing the configural model was abandoned. However, since in our research we adopted the bottom-up approach, we started with checking the dimensionality of the citizenship concepts in each country separately. It means that we tested if one-factor model of CITCON and CITSOC respectively are well fitted in each

of the country. Then we proceeded to verification of the measurement invariance properties for each citizenship scale separately. It meant that at each level of equivalence first one-factor model was checked: the model of CITCON and the model of CITSOC. Then, following Schulz and Friedman (2011, p. 176), we proceeded to verification of these properties specifying the two-factor model.

In factor analytical models the goodness-of-fit statistics are indicative of a well-fitting model. Among the broad range of fit indices, we choose to report root-mean-square error of approximation (RMSEA), 90% confidence interval for RMSEA, Tucker-Lewis index (TLI) and comparative-fit index (CFI), as implemented in Mplus (Muthén and Muthén, 2012). Concerning the RMSEA, values below 0.08 inform us about satisfactory low level of noise in the model (Browne and Cudeck, 1993) and values below 0.05 about very low level of noise (Hu and Bentler 1999). We want also the upper boundary of 90% confidence interval to be below 0.08 (Hu and Bentler, 1999). As for CFI and TLI, the model is satisfactory if they are over 0.95. Values over 0.90 are also considered acceptable (Hu and Bentler, 1999; Marsh et al., 2012). However, supported by (Hu and Bentler, 1999; Kline, 2011; Marsh et al., 2012, 2004) among others, we treat these cutoff values as only rough guidelines.

All CFA analyses were run using Mplus 6.1 and descriptive statistics were run in IBM SPSS Statistics Version 20. Due to the categorical character of items, in the estimation procedure, the robust weighted least squares estimator is used. This is a default estimator in the analysis using categorical indicators in Mplus (Muthén and Muthén 2012). Sampling weights are also taken into account in the analyses.

## **6. Results**

### **6.1. Conventional citizenship**

The results for Conventional citizenship concept showed that the fit of the one-factor model for 37 countries analyzed separately may be questioned. Our analyses indicated that the one-factor model was not well fitted in all countries participating in the ICCS. There were only 16 countries (highlighted in bold in Table 2) that fitted the model perfectly. Additional 10 countries (underlined ones) may be assessed as fitting the model, if a less strict approach to fit assessment is adopted, namely  $RMSEA < 0.10$ . The remaining 11 countries did not fit the model. These were

Chinese Taipei, Cyprus, Denmark, the Dominican Republic, Estonia, Greece, Guatemala, Hong Kong, the Republic of Korea, Malta, and Thailand.

Table 2. Conventional citizenship (CITCON) by country – fit statistics

Country	RMSEA	CFI	TLI	Country	RMSEA	CFI	TLI
<u>Austria</u>	<u>0.084</u>	<u>0.931</u>	<u>0.939</u>	<b>Lithuania</b>	<b>0.077</b>	<b>0.933</b>	<b>0.917</b>
<b>Bulgaria</b>	<b>0.066</b>	<b>0.965</b>	<b>0.961</b>	<b>Luxembourg</b>	<b>0.075</b>	<b>0.952</b>	<b>0.952</b>
<u>Chile</u>	<u>0.090</u>	<u>0.932</u>	<u>0.924</u>	Malta	0.104	0.896	0.896
Chinese Taipei	0.115	0.936	0.944	<b>Mexico</b>	<b>0.066</b>	<b>0.950</b>	<b>0.944</b>
<b>Colombia</b>	<b>0.060</b>	<b>0.944</b>	<b>0.923</b>	<u>Netherlands</u>	<u>0.081</u>	<u>0.920</u>	<u>0.920</u>
Cyprus	0.133	0.871	0.871	<b>New Zealand</b>	<b>0.068</b>	<b>0.964</b>	<b>0.964</b>
<b>Czech Republic</b>	<b>0.078</b>	<b>0.958</b>	<b>0.958</b>	<u>Norway</u>	<u>0.091</u>	<u>0.951</u>	<u>0.957</u>
Denmark	0.104	0.887	0.873	<b>Paraguay</b>	<b>0.062</b>	<b>0.927</b>	<b>0.911</b>
Dominican Republic	0.072	0.864	0.797	<u>Poland</u>	<u>0.084</u>	<u>0.944</u>	<u>0.950</u>
Estonia	0.113	0.905	0.881	<u>Russian Federation</u>	<u>0.091</u>	<u>0.951</u>	<u>0.945</u>
<b>Finland</b>	<b>0.071</b>	<b>0.972</b>	<b>0.975</b>	<b>Slovak Republic</b>	<b>0.066</b>	<b>0.953</b>	<b>0.953</b>
Greece	0.103	0.882	0.852	<u>Slovenia</u>	<u>0.092</u>	<u>0.941</u>	<u>0.948</u>
Guatemala	0.073	0.885	0.872	<u>Spain</u>	<u>0.088</u>	<u>0.934</u>	<u>0.934</u>
Hong Kong, SAR	0.117	0.930	0.939	<u>Sweden</u>	<u>0.087</u>	<u>0.964</u>	<u>0.968</u>
<b>Indonesia</b>	<b>0.083</b>	<b>0.900</b>	<b>0.862</b>	<b>Switzerland</b>	<b>0.063</b>	<b>0.956</b>	<b>0.956</b>
<b>Ireland</b>	<b>0.071</b>	<b>0.957</b>	<b>0.962</b>	Thailand	0.089	0.891	0.891
<b>Italy</b>	<b>0.073</b>	<b>0.931</b>	<b>0.923</b>	<b>England</b>	<b>0.078</b>	<b>0.961</b>	<b>0.966</b>
Korea, Republic of	0.112	0.923	0.931	<b>Belgium (Flemish)</b>	<b>0.079</b>	<b>0.947</b>	<b>0.940</b>
<b>Latvia</b>	<b>0.071</b>	<b>0.934</b>	<b>0.918</b>				

As regards the measurement invariance checking, the results for Conventional citizenship concept show that also the fit of the configural one-factor model for 37 countries may be questioned. However, taking into regard the high number of countries in ICCS and reported by Byrne and Van de Vijver (2010) frequently occurring problems with establishing measurement equivalence in large-scale cross cultural studies like ICCS, we decided to adopt in this special case less restrictive decision rules to the model fit assessment: namely  $RMSEA < 0.10$  and  $TLI$  and  $CFI > 0.90$ . Only after this decision was taken, can we assume that configural invariance holds for all countries.

The highest level of measurement invariance achieved for CITCON scale for all countries is weak measurement invariance. It holds, however, also for all subgroups of countries but Asian countries. Furthermore Southern and Central American countries constitute the best group of

countries in terms of scale equivalence. They are characterized by weak and strong measurement invariance.

These results imply that if one wants to compare average performance on the CITCON scale among countries, she can do that for Southern and Central American countries only. However if the researcher's interests are in comparing the strength of the relation between the CITCON scale and other variables, like for instance student achievement in a test of civic knowledge or other indicators of citizenship competencies, she is entitled to do so. Some caution should nevertheless be taken when comparing Asian countries only.

Table 3. Conventional citizenship (CITCON) – fit statistics

Country	RMSEA	90% CI	CFI	TLI
<b>One factor model with configural invariance – MG-CFA</b>				
all 37 countries	0.083	0.081 – 0.084	0.952	0.919
<b>One factor model with full weak measurement invariance – MG-CFA</b>				
all 37 countries	0.079	0.077 – 0.080	0.938	0.927
Asia + New Zealand	0.092	0.890 – 0.095	0.933	0.919
Europe	0.078	0.076 – 0.079	0.943	0.934
South and Central America	0.060	0.057 – 0.063	0.937	0.924
<b>One factor model with full strong measurement invariance – MG-CFA</b>				
all 37 countries	0.135	0.135 – 0.136	0.647	0.784
Asia + New Zealand	0.184	0.182 – 0.186	0.514	0.673
Europe	0.116	0.115 – 0.117	0.744	0.843
South and Central America	0.083	0.081 – 0.085	0.781	0.853

## 6.2. Social-movement citizenship

The results for Social-movement citizenship concept showed better fit of the one-factor model for 37 countries analyzed separately compared to the Conventional citizenship one-factor model. Although our analyses indicated that the one-factor model was not well fitted in all countries participating in the ICCS, in most of them it was. For 18 countries (bolded ones in Table 4) fit of the model was perfect. For additional 9 countries (underlined ones) the fit of the model was acceptable (with an application of a less strict approach to fit assessment, namely  $RMSEA < 0.1$ ). Only for 8 countries the model did not fit. These are the Czech Republic, Denmark, Hong Kong, Italy, the Republic of Korea, Lithuania, Malta and Slovenia.

Table 4. Social-movement citizenship (CITSOC) by country – fit statistics

Country	RMSEA	CFI	TLI	Country	RMSEA	CFI	TLI
<b>Austria</b>	<b>0.063</b>	<b>0.993</b>	<b>0.985</b>	Lithuania	0.158	0.949	0.898
<u>Bulgaria</u>	<u>0.094</u>	<u>0.987</u>	<u>0.974</u>	<b>Luxembourg</b>	<b>0.053</b>	<b>0.994</b>	<b>0.991</b>
<b>Chile</b>	<b>0.036</b>	<b>0.998</b>	<b>0.995</b>	Malta	0.131	0.967	0.935
<b>Chinese Taipei</b>	<b>0.055</b>	<b>0.998</b>	<b>0.997</b>	<b>Mexico</b>	<b>0.054</b>	<b>0.992</b>	<b>0.983</b>
<u>Colombia</u>	<u>0.081</u>	<u>0.970</u>	<u>0.940</u>	<b>Netherlands</b>	<b>0.012</b>	<b>1.000</b>	<b>0.999</b>
<b>Cyprus</b>	<b>0.079</b>	<b>0.990</b>	<b>0.985</b>	<u>New Zealand</u>	<u>0.086</u>	<u>0.989</u>	<u>0.984</u>
Czech Republic	0.123	0.981	0.972	<u>Norway</u>	<u>0.087</u>	<u>0.989</u>	<u>0.984</u>
Denmark	0.100	0.988	0.982	<b>Paraguay</b>	<b>0.065</b>	<b>0.976</b>	<b>0.939</b>
<b>Dominican Republic</b>	<b>0.045</b>	<b>0.987</b>	<b>0.966</b>	<b>Poland</b>	<b>0.074</b>	<b>0.993</b>	<b>0.987</b>
<b>Estonia</b>	<b>0.017</b>	<b>1.000</b>	<b>0.999</b>	<u>Russian Federation</u>	<u>0.099</u>	<u>0.985</u>	<u>0.970</u>
<b>Finland</b>	<b>0.045</b>	<b>0.998</b>	<b>0.996</b>	<u>Slovak Republic</u>	<u>0.099</u>	<u>0.985</u>	<u>0.969</u>
<u>Greece</u>	<u>0.083</u>	<u>0.988</u>	<u>0.976</u>	Slovenia	0.140	0.963	0.945
<b>Guatemala</b>	<b>0.033</b>	<b>0.996</b>	<b>0.991</b>	<u>Spain</u>	<u>0.094</u>	<u>0.985</u>	<u>0.977</u>
Hong Kong, SAR	0.132	0.980	0.969	<b>Sweden</b>	<b>0.070</b>	<b>0.996</b>	<b>0.994</b>
<b>Indonesia</b>	<b>0.067</b>	<b>0.985</b>	<b>0.963</b>	<b>Switzerland</b>	<b>0.072</b>	<b>0.992</b>	<b>0.984</b>
<u>Ireland</u>	<u>0.096</u>	<u>0.989</u>	<u>0.983</u>	<b>Thailand</b>	<b>0.034</b>	<b>0.997</b>	<b>0.993</b>
Italy	0.133	0.971	0.942	<b>England</b>	<b>0.067</b>	<b>0.995</b>	<b>0.993</b>
Korea, Republic of	0.145	0.985	0.978	<b>Belgium (Flemish)</b>	<b>0.039</b>	<b>0.998</b>	<b>0.998</b>
<u>Latvia</u>	<u>0.097</u>	<u>0.971</u>	<u>0.942</u>				

As regards the configural one-factor model of social-movement citizenship, again, following strictly the rules of thumb for fit statistics would lead to its rejection for 37 countries. In this case we also decided to adopt less restrictive decision rules to the model fit assessment: namely  $RMSEA < 0.10$  and  $TLI$  and  $CFI > 0.90$ . Only after this decision was taken, can we assume that configural invariance holds for all countries. Taking into regard all 37 countries, the social-movement citizenship scale is characterized by configural and weak invariance. Referring to strong invariance, it does not hold for all countries.

These results imply that if one wants to compare average performance on the CITSOC scale among countries cannot do that. However if the researcher's interests are in comparing the strength of relation among CITSOC scale and other variables, she is entitled to do so.

Table 5. Social-movement-related citizenship (CITSOC) – fit statistics

Countries	RMSEA	90% CI	CFI	TLI
<b>One factor model with configural invariance – MG-CFA</b>				
All 37 countries	0.085	0.081 – 0.088	0.992	0.975
<b>One factor model with full weak measurement invariance – MG-CFA</b>				

All 37 countries	0.063	0.061 – 0.066	0.991	0.986
Asia + New Zealand	0.076	0.071 – 0.080	0.991	0.986
Europe	0.065	0.062 – 0.068	0.991	0.986
South and Central America	0.040	0.035 – 0.045	0.992	0.987
<b>One factor model with full strong measurement invariance – MG-CFA</b>				
All 37 countries	0.096	0.095 – 0.097	0.938	0.968
Asia + New Zealand	0.144	0.141 – 0.147	0.911	0.948
Europe	0.085	0.083 – 0.087	0.955	0.976
South and Central America	0.042	0.039 – 0.046	0.974	0.985
<b>One factor model with full strict measurement invariance – MG-CFA</b>				
South and Central America	0.051	0.048 – 0.054	0.918	0.926

### 6.3. Conventional and Social-movement citizenship

At the final stage we checked the measurement invariance properties assuming that the specification of citizenship scales relates to a two-factor model. The outcome of the analyses done for each country separately showed that again not for all countries a two factor model fitted data sufficiently. For 16 countries (highlighted in bold in Table 4) the fit of the model was highly . For additional 7 countries (underlined ones) fit of the model was acceptable (with an application of less strict approach to fit assessment, namely  $RMSEA < 0.10$ ). However for 14 countries the model did not fit. These were Bulgaria, Chinese Taipei, Colombia, Cyprus, the Dominican Republic, Estonia, Greece, Guatemala, Hong Kong, Indonesia, Malta, Norway, Paraguay, Slovenia and Thailand.

Most of these countries popped up as misfitting while checking the fit of conventional citizenship model. The only exception was Slovenia, which did not fit the social-movement related citizenship model. Hong Kong and Malta were the only two countries that did not fit both one-factor citizenship models and a two-factor model. Furthermore, we noticed that for Bulgaria, Colombia, Indonesia, Norway and Paraguay, although both Conventional citizenship and Social-movement related citizenship models taken separately fitted the data, two-factor model was not fitted. On the other hand, for Denmark and the Republic of Korea, although the one-factor models model misfitted the data, a two factor model fitted the data perfectly. It means that in a two-factor model for these countries strong compensation for misfit occurred.

Table 6. Citizenship: conventional and social-movement related by country – fit statistics

Country	RMSEA	CFI	TLI	Country	RMSEA	CFI	TLI
---------	-------	-----	-----	---------	-------	-----	-----

<u>Austria</u>	<u>0.076</u>	<u>0.891</u>	<u>0.945</u>	<b>Lithuania</b>	<b>0.073</b>	<b>0.905</b>	<b>0.936</b>
Bulgaria	0.099	0.863	0.907	<b>Luxembourg</b>	<b>0.061</b>	<b>0.942</b>	<b>0.969</b>
<u>Chile</u>	<u>0.080</u>	<u>0.894</u>	<u>0.934</u>	Malta	0.103	0.813	0.899
Chinese Taipei	0.113	0.895	0.954	<b>Mexico</b>	<b>0.070</b>	<b>0.904</b>	<b>0.942</b>
Colombia	0.068	0.872	0.898	<b>Netherlands</b>	<b>0.069</b>	<b>0.913</b>	<b>0.949</b>
Cyprus	0.101	0.859	0.937	<b>New Zealand</b>	<b>0.075</b>	<b>0.927</b>	<b>0.964</b>
<u>Czech Republic</u>	<u>0.081</u>	<u>0.931</u>	<u>0.959</u>	Norway	0.090	0.891	0.960
<u>Denmark</u>	<u>0.086</u>	<u>0.902</u>	<u>0.943</u>	Paraguay	0.083	0.798	0.812
Dominican Republic	0.072	0.836	0.836	<b>Poland</b>	<b>0.076</b>	<b>0.932</b>	<b>0.961</b>
<u>Estonia</u>	<u>0.085</u>	<u>0.903</u>	<u>0.932</u>	<b>Russian Federation</b>	<b>0.070</b>	<b>0.936</b>	<b>0.965</b>
<b>Finland</b>	<b>0.077</b>	<b>0.941</b>	<b>0.969</b>	<b>Slovak Republic</b>	<b>0.060</b>	<b>0.943</b>	<b>0.965</b>
Greece	0.114	0.787	0.836	Slovenia	0.101	0.863	0.926
Guatemala	0.081	0.786	0.855	<b>Spain</b>	<b>0.080</b>	<b>0.908</b>	<b>0.949</b>
Hong Kong, SAR	0.099	0.889	0.959	<b>Sweden</b>	<b>0.080</b>	<b>0.943</b>	<b>0.975</b>
Indonesia	0.073	0.884	0.896	<b>Switzerland</b>	<b>0.060</b>	<b>0.926</b>	<b>0.961</b>
<b>Ireland</b>	<b>0.071</b>	<b>0.938</b>	<b>0.969</b>	Thailand	0.093	0.765	0.894
<b>Italy</b>	<b>0.074</b>	<b>0.911</b>	<b>0.939</b>	<b>England</b>	<b>0.075</b>	<b>0.942</b>	<b>0.973</b>
<u>Korea, Republic of</u>	<u>0.095</u>	<u>0.925</u>	<u>0.968</u>	<u>Belgium (Flemish)</u>	<u>0.091</u>	<u>0.923</u>	<u>0.947</u>
<b>Latvia</b>	<b>0.060</b>	<b>0.922</b>	<b>0.942</b>				

Analyses performed for all countries together showed that only the configural model and the model with weak measurement invariance fitted the data well. Furthermore, the weak MI was also confirmed for the European countries. Nevertheless, the strong measurement invariance, in which the researcher aiming at countries comparisons is most interested in, is confirmed neither for all countries nor for the groups of them.

Table 7. Citizenship: conventional and social-movement related (CITCON&CITSOC) – fit statistics

<b>countries</b>	<b>RMSEA</b>	<b>90% CI</b>	<b>CFI</b>	<b>TLI</b>
<b>Two factor model with configural invariance – MG-CFA</b>				
All 37 countries	0.076	0.076 – 0.077	0.939	0.919
<b>Two factor model with full weak measurement invariance – MG-CFA</b>				
All 37 countries	0.071	0.071 – 0.072	0.937	0.929
Asia + New Zealand	0.082	0.080 – 0.083	0.939	0.930
Europe	0.071	0.070 – 0.072	0.941	0.933
South and Central America	0.064	0.063 – 0.066	0.899	0.893
<b>Two factor model with full strong measurement invariance – MG-CFA</b>				
All 37 countries	0.102	0.102 – 0.103	0.873	0.917
Asia + New Zealand	0.139	0.138 – 0.141	0.746	0.794
Europe	0.090	0.090 – 0.091	0.857	0.891

South and Central America	0.068	0.067 – 0.070	0.850	0.879
---------------------------	-------	---------------	-------	-------

These results imply that even taking into account the correlation between the two components of the citizenship values concept, the cross-country comparisons of CITCON and CITSOC means are not justified, especially under the factor analytical framework. Nevertheless, the comparisons of the strength of relation among CITCON and CITSOC scales and other variables are justified provided that the correlation among these citizenship values scales is taken into account. However it must be noted that even so some caution should be taken when comparing Asian countries only.

## 7. Discussion

The increasing availability and complexity of cross-national datasets from international large scale assessment studies in education such as ICCS contribute significantly to scientific advancements. They provide researchers with the possibility of investigating a numerous number of research questions in a comparative perspective and arrive to results relevant to education policy and practice. In this article we have tried to show that such studies can be extremely valuable only if the researchers are well aware of some of the limitations that the studies might have and are able to apply the appropriate methodology. As noted by the experts involved in the design and data processing of the ICCS study (see Schulz, 2009), the assessment of measurement invariance of questionnaire data has not received enough attention with the implication that secondary data analysis might result in comparisons of constructs that are not necessarily appropriate given the properties of the data. We have therefore illustrated this issue by investigating whether valid comparisons of two concepts frequently used for comparative secondary data analysis are indeed possible. As aforementioned, these are just examples and cross-country equivalence should be examined regarding all other concepts/instruments used in international large-scale-assessments when one desires to compare several countries.

Tests of cross-country equivalence were performed adopting a factor analytical framework and testing for full measurement invariance of the instruments used to assess the concepts of conventional and social movement citizenship. Confirming our assumption, the results showed a complex picture indicating that the concepts are not necessarily comparable across all countries

involved in ICCS and their use for secondary data analysis can be more or less feasible across all countries depending of the research questions and the methodology applied.

Regarding the two constructs measuring citizenship values, we took two approaches in testing for measurement invariance: assuming one and two factor models. The choice was motivated by observations regarding the practical use of the ICCS data. From our review of literature we could see that researchers tend to use either one or both concepts in cross-country analyses or country comparisons. Moreover, they can involve either all ICCS countries, countries from the same geographical regions (e.g. European) or smaller groups of countries chosen with the reason of either cultural/educational policy similarity or dissimilarity but most of the time from the same region. Furthermore, these analyses can have the purpose of either detecting relationships with other variables (e.g. regression coefficients) or comparing means of latent variables between countries.

Our results showed that there were countries for which the conceptualizations (both one-factor models and two-factor model) did not correspond to the data structure. However, it appeared that the discrepancies observed in one country models were efficiently compensated for in configural and weak measurement invariant models. Furthermore the conventional and social movement citizenship scales considered separately (one-factor model) as well as interrelated (two-factor model) met only the levels of configural and weak measurement invariance, failing to show higher levels of MI such as strong and strict. The configural and weak invariance were nevertheless found also for all sub-groups of countries. The only exception is the group of South and Central American countries where the instrument measuring conventional citizenship showed also strong measurement properties and the test of measurement of invariance for the social movement citizenship scale achieved not only configural, weak and strong but also the highest level of strict MI. The implications of these findings are that the items in the measurement instruments of the conventional and social movement citizenship scales exhibit not only the same configuration of loadings but also the same values of loadings in all the relationships with factors across all countries. It means that each of the citizenship scales has the same measurement unit across countries. This assures that the students in the different countries give a similar meaning to these concepts and tend to respond to the items in the same way. As a result, comparisons involving associations of these concepts and other theoretical construct of interest are possible in analyses conducted across all countries. On the other hand the fact that

results did not support the strong measurement invariance for all countries (except for the South and Central American group) leads to the conclusion that the cross-country differences in the means of the observed items are not a result of the difference in means of the corresponding items. Therefore the instruments measuring citizenship values cannot be used for comparing mean values of the conventional and social movement citizenship across all ICCS countries. These comparisons are nevertheless possible for the South and Central American countries.

Based on these findings researchers can test associations between citizenship values and other concepts (e.g. the relationship between citizenship values and anticipated political and social participation) but may not compare average values of all countries (e.g. students in country X display less positive levels of citizenship values compared to students in other countries). This further extends to models testing associations with methodologies that assume equality of intercepts (e.g. the case of multilevel models) (see Meuleman, 2011). In the case of ICCS these comparisons are informative only when involving South and Central American countries.

Although our study provides researchers with extensive information on assessing the MI of these concepts, it has also some limitations which open possibilities for further research. First, referring to the citizenship values concepts, we mentioned that we have chosen to conduct the analyses on the whole set of countries and on the groups of countries (as in the ICCS data collection modules). The grouping is based on the geographical location that, according to Byrne and Van de Vijver (2010), is often not the most valid one in large scale cross-cultural research. Moreover, this strategy does not exclude the possibility that within or across these groupings smaller subsets of countries might be in fact comparable. Future research could further investigate this scenario. Furthermore, we were only able to show that strong MI does not hold in some cases but we were not able to show why. Provided that such results could be explained by cultural country characteristics (e.g. cultural and historical similarity/dissimilarity, experience with democracy and forms of political participation) further research could examine the reasons for measurement non-invariance by conducting complementary cognitive tests (see Davidov, 2008) and/or examining whether contextual country level variables can explain why measurement invariance was not achieved (see Davidov, Dulmer, Schluter, Schmidt & Meuleman, 2012).

Second, future research may provide further information if alternative techniques to assess MI are applicable. Some of these possibilities could aim at partial MI instead of full MI which means that only a subset of items meeting the invariance criteria are used to estimate group difference (see Gregorich, 2006; Bryne and Van de Vijver, 2010) or apply multilevel SEM techniques (see Davidov, et. al., 2012) which, as aforementioned, would give the possibility of detecting the causes of non-invariance. Moreover, instead of a factor analytical framework one could analyze this type of dichotomous data under an IRT framework (Jansen 2011; see de Jong, Steenkamp, and Fox 2007). Furthermore, the results as well as the concluding statements made in this paper are influenced by our decision to take a less strict approach in evaluating the model fit, especially in the case of the configural invariance verification. However, as stated clearly by Kline (2011, p. 197), the suggested fit criteria are only ‘rules of thumb’ and not ‘golden rules’ and especially in the large scale cross-cultural research should be applied with caution as documented by Byrne and Van de Vijver (2010). Nevertheless, if we had taken a more conservative stand (see for instance Hu and Bentler, 1999) some of our conclusions could change (e.g. both citizenship scales treated separately would have been assessed as not meeting configural or weak MI for all 37 countries).

## References

- Almond, G. & Verba, S. (1963). *The Civic Culture*. Princeton, N.J.: Princeton University Press.
- Asparouhov, T., Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 16, 397–438.
- Baumgartner, H., Steenkamp, J.E.M. (1992). The Role of Optimum Stimulation Level in Exploratory Consumer Behavior. *The Journal of Consumer Research* 19, 434–448.
- Browne, M.W., Cudeck, R. (1993). Alternative ways of assessing model fit. In: *Testing Structural Equation Models*. CA: Sage., Newsbury Park, pp. 136–162.
- Byrne, B.M. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema* 20, 872–882.
- Byrne, B.M., Shavelson, R.J., Muthen, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin* 105, 456–466.
- Byrne, B.M., Van de Vijver, F.J.R. (2010). Testing for Measurement and Structural Equivalence in Large-Scale Cross-Cultural Studies: Addressing the Issue of Nonequivalence. *International Journal of Testing* 10, 107–132.
- Dalton, R. (2006). *Citizenship norms and political participation in America: The good news is ... the bad news is wrong* (CDACS occasional paper). Washington DC: Center for Democracy and Civil Society, Georgetown University.
- Davidov, E., (2008). A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey. *Survey Research Methods* 2, 33–46.
- Davidov, E., H. Dülmer, E. Schlüter, P. Schmidt and B. Meuleman. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance (2012). *Journal of Cross-Cultural Psychology*, 43(4), 558-575.
- Davidov, E., Meuleman, B., Billiet, J., Schmidt, P. (2008). Values and Support for Immigration : A Cross-Country Comparison. *European Sociological Review* 1–17.
- De Jong, M.G., Steenkamp, J.E.M., Fox, J. (2007). Relaxing Measurement Invariance in Cross-National Consumer Research Using a Hierarchical IRT Model. *Journal of Consumer Research* 34, 260–278.
- Gregorich, S.E. (2006). Do Self-Report Instruments Allow Meaningful Comparisons Across Diverse Population Groups? Testing Measurement Invariance Using the Confirmatory Factor Analysis Framework. *Medical Care* 44, 78–94.

- Horn, J.L., Mcardle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process* 18, 117–144.
- Hoskins, B. L., Barber, C., Van Nijlen, D. & Villalba, E. (2011). Comparing civic competence among European youth: composite and domain-specific indicators using IEA civic education study data. *Comparative Education Review*, 55, (1), 82-110. (doi:10.1086/656620).
- Hoskins, B., Janmaat, J. G., & Villalba, E. (2012). Learning citizenship through social participation outside and inside school: An international, multilevel study of young people's learning of citizenship. *British Educational Research Journal*, 38, 419–446.
- Hoskins, B., Villalba, C. & Saisana, M. (2012). The 2011 civic competence composite indicator (CCCI-2): measuring young people's civic competence across Europe based on the IEA international citizenship and civic education study. Ispra, Italy, European Commission, 100pp. (doi:10.2788/67938).
- Hu, L., Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6.
- Isac, M.M., Maslowski. R., Creemers, B., van der Werf, M.P.C. (2013). The contribution of schooling to secondary-school students' citizenship outcomes across countries, *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, DOI:10.1080/09243453.2012.7510352011.
- Jansen, R. (2011). Using Differential Item Functioning Approach to Investigate Measurement Invariance. In: Davidov, E., Schmidt, P., Billiet, J. (Eds.), *Cross-Cultural Analysis: Methods and Applications*. Routledge, Taylor & Francis Group, New York, NY, pp. 415–432.
- Lubke, G.H., Dolon, C. V. (2003). Can Unequal Residual Variances Across Groups Mask Differences in Residual Means in the Common Factor Model? *Structural Equation Modeling: A Multidisciplinary Journal* 10, 175–192.
- Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S., Trautwein, U., 2009. Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal* 16, 439–476.
- Marsh, W.H., Lüdtke, O., Muthen, B., Asparouhov, T., Morin, A.J.S., U., T., 2010. A new look at the Big-Five factor structure through exploratory structural equation modeling. *Psychological Assessment* 22, 471–491.
- Meredith, W. (1993). MI, factor analysis and factorial invariance. *Psychometrika* 58, 525–543.

- Meredith, W., Teresi, J.A., 2006. An essay on measurement and factorial invariance. *Medical Care* 44, 69–77.
- Meuleman, B. (2011). Perceived Economic Threat and Anti-Immigration Attitudes: Effects of Immigrant Group Size and Economic conditions Revisited. In: Davidov, E., Schmidt, P., Billiet, J. (Eds.), *Cross-Cultural Analysis: Methods and Applications*. Routledge, Taylor & Francis Group, New York, NY, pp. 281–310.
- Millsap, R.E., Kwok, O.-M. (2004). Evaluating the Impact of Partial Factorial Invariance on Selection in Two Populations. *Psychological Methods* 9, 93–115.
- Muthén, L.K., Muthén, B.O. (2012). *Mplus User's Guide*, Seventh. ed. Muthen&Muthen, Los Angeles, CA.
- Schulz, W., Ainley, J., Fraillon, J. (Eds.), 2011. *ICCS 2009 Technical Report*. IEA, ACER, nfer, Università degli Studi Roma Tre.
- Schulz, W. & Sibberns, H.(2004). *IEA Civic Education Study: Technical report*. Amsterdam, The Netherlands: IEA.
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam, The Netherlands: IEA.
- Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). *International Civic and Citizenship Education Study: Assessment framework*. Amsterdam, The Netherlands: IEA.
- Schulz, W., Ainley, J., & Fraillon, J. (Eds.). (2011). *ICCS 2009 Technical report*. Amsterdam, The Netherlands: IEA.
- Schulz W. (2009). Questionnaire construct validation in the International Civic and Citizenship Education Study, *IERI Monograph Series Issues and Methodologies in Large-Scale Assessments*, Vol. 2, IEA-ETS Research Institute, ISBN 978-0-88685-404-1, p. 113-136.
- Schulz, W., Friedman, T. (2011). Scaling procedure for ICCS questionnaire items. In: Schulz, W., Ainley, J., Fraillon, J. (Eds.), *ICCS Technical Report*. IEA, ACER, NFER, Università degli Studi Roma Tre, pp. 157–259.
- Selig, J.P., Card, N.A., Little, T.D., (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In: van de Vijver, F.J.R., van Hemert, D.A., Poortinga, Y.H. (Eds.), *Multilevel Analysis of Individuals and Cultures*. Taylor & Francis Group/Lawrence Erlbaum Associates, New York, NY, pp. 93–119.
- Sherrod, L. R., Torney-Purta, J., & Flanagan, C. A. (2010). Research on the development of citizenship: A field comes of age. In L. R. Sherrod, J. Torney-Purta, & C. A. Flanagan

- (Eds.), *Handbook of research on civic engagement in youth* (pp. 1–20). Hoboken, NJ: Wiley.
- Steenkamp, J.-B.E.M., Baumgartner, H., (1998). Assessing measurement invariance in cross-national consumer research. *The Journal of Consumer Research* 25, 78–90.
- Torney, J.V., Oppenheim, A.N., & Farnen, R.F. (1975). *Civic education in ten countries: An empirical study*. Stockholm: Almqvist & Wiksell.
- 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: IEA.
- 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, 129-141.
- Torney-Purta, J., & Richardson, W. (2004). Anticipated political engagement among adolescents in Australia, England, Norway and the United States. In J. Demaine (Ed.), *Citizenship and political education today* (pp. 41–58). London, UK: Palgrave/Macmillan.
- Welkenhuysen-Gijbels, J., Billiet, J., Cambre, B., (2003). Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-Type Score Items. *Journal of Cross-Cultural Psychology* 34, 702–722.
- Winkler, D., and Gershberg, A. (2000). *Education Decentralization in Latin America: the Effects on the Quality of Schooling*, LCSHD Paper Series No. 59. Washington DC: The World Bank.
- Wu, A.D., Li, Z., Zumbo, B.D., (2007). Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration with TIMSS Data. *Practical Assessment Research & Evaluation* 12, 1–26.

Appendix

Table 8. Group of countries

Group	Countries
European countries (26)	Austria, Belgium (Flemish), Bulgaria, Cyprus, the Czech Republic, Denmark, England, Estonia, Finland, Greece, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, the Netherland, Norway, Poland, the Russian Federation, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland,
South and Central American countries (7)	Chile, Colombia, the Dominican Republic, Guatemala, Mexico, Paraguay, Thailand
Asian countries (4)	Chinese Taipei, Hong Kong, Indonesia, the Republic of Korea
	New Zealand