

A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey

Eldad Davidov

University of Cologne/University of Mannheim

The study reported in this paper assesses the fit of a 21-item instrument measuring values in the second round of the European Social Survey (ESS) to the theory of 10 basic values on which it was based (Schwartz 1992). In particular, the measurement invariance of this instrument for studying value priorities across nations and over time was investigated. In the first part of the study, using multi-group confirmatory factor analysis (MGCFA) of data from the second ESS round, configural, metric, and scalar invariance of the values are assessed across 25 countries. Metric invariance is a necessary condition to insure equivalence of the meaning of factors and a precondition for comparing values' correlates. Scalar invariance is a precondition for comparing value means. The MGCFA did not support configural and metric invariance across 25 countries. After reducing the number of countries to 14, the MGCFA supported metric invariance of the same model found for the data of Round 1 (Davidov, Schmidt, and Schwartz in press) with seven distinct values. These value measurements may now be used by researchers to study relationships among values, attitudes, behavior, and sociodemographic characteristics across the 14 nations. Comparing national value means may be possible only across a smaller set of countries where scalar invariance holds. In the second part of the study, metric and scalar invariance were established between the first and the second rounds of the ESS in each of 19 countries separately. In 9 countries, the model found for the data of Round 1 also fitted the longitudinal comparison, but in 10 countries model modifications were needed. Value means may be compared for each of the countries between the first and second ESS rounds (2002-2003 and 2004-2005, respectively).

Keywords: European Social Survey, configural, metric, and scalar invariance, human values, multiple-group confirmatory factor analysis (MGCFA), cross-country comparison

Introduction

The increasing importance of cross-cultural and longitudinal research has advanced the design and implementation of cross-national surveys in the last decades. Surveys such as the European Social Survey (ESS) or the International Social Survey Program (ISSP) are conducted to follow changes across time and to compare countries. The designers of these surveys pay serious attention to exact translations of the items into different languages, to comparable response sets, to comparable sampling procedures, and to similar data collection techniques. Although these steps are necessary for establishing measurement invariance over time or across countries, unfortunately, they cannot guarantee full invariance. Rather, measurement invariance of theoretical constructs should also be assessed during data analysis (Billiet 2003).

The semiannual European Social Survey (ESS) is a major source of data for researchers across all the social sci-

ence disciplines. It employs rigorous methodologies to ensure data comparability across countries and over time (see, e.g., Harkness, Van de Vijver, and Mohler 2003, Ch. 3, for the general standards used for translation). The designers of the European Social Survey chose the Schwartz value theory (1992) as the basis for developing a human values instrument to include in the core of the survey. This instrument was administered to representative national samples in 20 countries in 2002-2003, in 25 countries in 2004-2005, and will be included in future ESS rounds (for details, view <http://www.europeansocialsurvey.org>).

A central question when applying this instrument in cross-country cross-time research is whether it is invariant. The measurement invariance of the value factors may be assessed using various data analysis techniques. Metric invariance is a necessary condition to ensure equivalence of the meaning of the value constructs. Scalar invariance should also be assessed because it is a precondition for comparing national means. Mean comparison is often a central goal in comparative research. Metric and scalar invariance of the value scale have been assessed by Davidov, Schmidt, and Schwartz (in press) with data from the first round of the ESS. However, as of yet the value questions in the second ESS round have not been subjected to such tests. Furthermore, the measurement invariance of the value measurements over time has also not been assessed; this has only become possi-

Contact information: Eldad Davidov, GESIS-Central Archive for Empirical Social Research, University of Cologne and University of Mannheim, Faculty of Social Sciences, A5, 6, 68131 Mannheim, Germany. email: Davidov@za.uni-koeln.de or Davidov@sowi.unimannheim.de

ble since the release of the data from the second round. Such an assessment will answer the question whether estimating value change is legitimate.

In this study, we test whether and to what extent values as measured by the ESS in different European countries are invariant across countries and over time. By doing so, we try to answer the following questions:

1. Does the model found for the ESS data of round 1 (Davidov et al., in press) fit the data in Round 2?
2. Are the values comparable across the full set of countries participating in the second ESS round?
3. Are they comparable over time in each country separately?

To this end, we employ multi-group confirmatory factor analysis (MGCFA) and data from the first and second ESS rounds and assess their invariance. Finding such invariance will allow researchers to compare value means and their correlates and answer questions such as: Can we compare values means of Germany and France, for instance? Is an estimation of aggregate value changes in Germany, for example, legitimate? Thus, the content of this paper is as follows: I begin by introducing the theory of basic human values and their measurement in the ESS. Then I discuss how to test for invariance. Following these discussions is the section describing the data analysis, and finally, this paper concludes with a summary and some closing remarks.

The theory of basic human values

Human values have played an increasingly important role in comparative social science research in the last decades. There are several reasons for this development. Values have become more and more central to public discourse. Social science theories have considered values as important determinants of attitudes and behavior (Allport, Vernon, and Lindsay 1960; Feldman 2003; Halman and de Moor 1994; Kluckhohn 1951; Rokeach 1973; Schwartz 1992; Williams 1968). In 1992, Schwartz introduced a comprehensive new value theory which proposes 10 universal values and four higher-order value dimensions. He developed two instruments to measure them which have been successfully tested in many countries. Furthermore, the European Social Survey (ESS) decided to adopt one of these instruments by including a shortened version of it in its core questionnaire. This instrument was implemented in 20 countries in the first round of the ESS (2002/2003) and in 25 countries in the second round (2004/2005), and it will be included in future rounds of the ESS as well.

Schwartz' theory (1992) defines values as desirable, transsituational goals, varying in importance, that serve as guiding principles in people's lives. It proposes 10, motivationally distinct, basic values, covering values found in earlier theories (e.g., Inglehart 1990; Rokeach 1973). The values put forth by Schwartz are: hedonism, stimulation, self-direction, security, universalism, benevolence, conformity, tradition, power, and achievement. The motivational goal of each basic value, presented in Table 1, defines it. For example, the motivational goal of the security value is safety,

harmony, and stability of society, of relationships, and of self. The motivational goal of the universalism value is understanding, appreciation, tolerance, and protection for the welfare of all people and for nature.

In addition to identifying 10 basic values, the theory postulates the structure of dynamic relations among these values. The main source of the value structure is the fact that values may be interdependent. Actions in pursuit of any value have consequences that will conflict with some values and will be compatible with other values. For example, pursuing conformity values may conflict with pursuing self-direction values. Restraint of actions that are likely to violate social expectations or norms may be in conflict with independent thought and action. But pursuing both tradition and conformity values may be congruent. Commitment and acceptance of the customs and ideas that traditional culture or religion provide may be compatible with restraint of actions, inclinations, and impulses that are likely to upset or harm others and violate social expectations or norms.

The circular structure in Figure 1 portrays the pattern of relations among values. Compatible values are close to each other around the circle; competing value types appear in opposing directions. The theory postulates that the values form a continuum at a more basic level; the motivational differences of values are continuous rather than discrete. As a result, adjacent values often intermix in empirical studies. Thus, the theory specifies the motivational order of value items around the circle and it suggests distinguishing between 10 value factors. However, when there are measurement restrictions such as a low number of items to measure each value, one may not always find 10 distinct values. Instead, pairs of adjacent values may be captured as a single value (e.g., universalism and benevolence).

The opposition between value types can be summarized by viewing values as organized in two bi-polar contrasting higher-order dimensions (Figure 1). The first higher-order dimension contrasts self-enhancement and self-transcendence value types. This dimension opposes power and achievement values – that emphasize one's pursuit of success and dominance of others – to universalism and benevolence values – that involve concern for the welfare and interests of others. The second higher-order dimension contrasts openness to change and conservation value types. This dimension opposes self-direction and stimulation values – that emphasize independent action, thought, and feeling and readiness for new experiences – to conservation values, security, conformity, and tradition – that emphasize self-restriction, order, and resistance to change (for a further discussion see, e.g., Schwartz 1992, 1994). Before conducting the empirical analyses, a short description of the ESS instrument to measure values is provided.

The Human Values Measurement Instrument in the ESS

This instrument originated from the earlier 40-item Portrait Values Questionnaire (PVQ; Schwartz, Melech, Lehmann, Burgess, Harris, and Owens 2001; Schwartz

Table 1: Definitions of the Motivational Types of Values in Terms of their Core Goal

POWER:	Social status and prestige, control or dominance over people and resources
ACHIEVEMENT:	Personal success through demonstrating competence according to social standards
HEDONISM:	Pleasure and sensuous gratification for oneself
STIMULATION:	Excitement, novelty, and challenge in life
SELF-DIRECTION:	Independent thought and action-choosing, creating, exploring
UNIVERSALISM:	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature
BENEVOLENCE:	Preservation and enhancement of the welfare of people with whom one is in frequent personal contact
TRADITION:	Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide the self
CONFORMITY:	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms
SECURITY:	Safety, harmony, and stability of society, of relationships, and of self

Adopted from Sagiv and Schwartz, 1995.

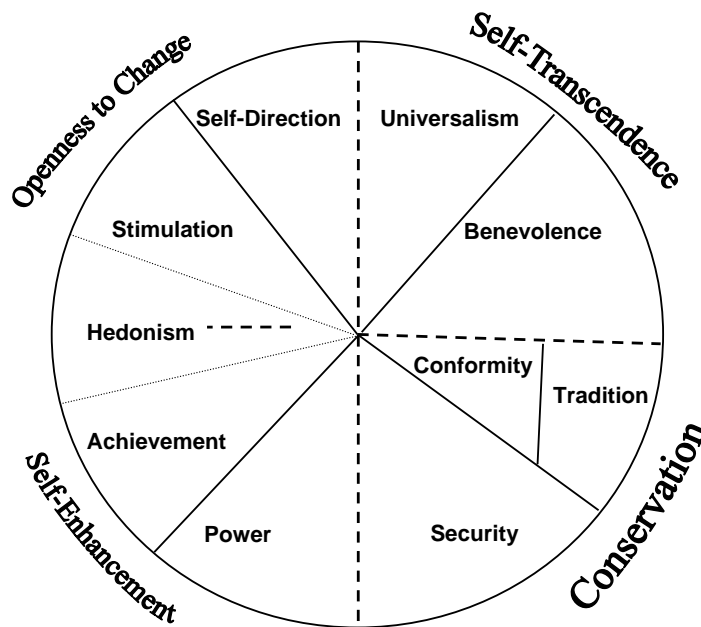


Figure 1. Structural relations among the 10 values and the two dimensions

2005). However, it was reduced to 21 items due to space limitations in the ESS. The ESS scale includes verbal portraits of 21 different people, gender-matched with the respondent. Each portrait describes a person’s goals, aspirations, or wishes that point implicitly to the importance of a value. For example: “He thinks it is important that every person in the world be treated equally. He believes everyone should have equal opportunities in life” describes a person for whom universalism values are important. Respondents’ own values are inferred from their self-reported similarity to people described implicitly in terms of particular values.

Regarding each portrait, respondents are asked to answer: “How much like you is this person?” Responses are recorded on Likert-type rating scales ranging from 1 (not like me at all) to 6 (very much like me). Table 2 presents

the value labels and the items and their names, grouped by type of value. Two portraits are used as reflective indicators to measure each value, and three for universalism because of its very broad content (for a general discussion on measurement of latent variables see, e.g., Bollen and Lennox 1991; Edwards and Bagozzi 2000). Table 2 displays the question wording of the male version for each item.

The countries, with sample size listed in parentheses, are: Austria (2,256), Belgium (1,778), Czech Republic (3,026), Denmark (1,487), Estonia (1,989), Finland (2,022), France (1,806), Germany (2,870), Great Britain (1,897), Greece (2,406), Hungary (1,498), Ireland (2,286), Iceland (579), Luxemburg (1,635), Netherlands (1,881), Norway (1,760), Poland (1,716), Portugal (2,052), Slovenia (1,442), Slovakia (1,512), Spain (1,663), Sweden (1,948), Switzer-

Table 2: The ESS Human Values Scale in the 2nd Round (N = 47,537)

Value	Item # (according to its order in the ESS questionnaire) and Wording (Male Version)
Self-Direction (SD)	1. Thinking up new ideas and being creative is important to him. He likes to do things in his own original way (ipctiv). 11. It is important to him to make his own decisions about what he does. He likes to be free to plan and not depend on others (impfree).
Universalism (UN)	3. He thinks it is important that every person in the world be treated equally. He believes everyone should have equal opportunities in life (ipeqopt). 8. It is important to him to listen to people who are different from him. Even when he disagrees with them, he still wants to understand them (ipudrst). 19. He strongly believes that people should care for nature. Looking after the environment is important to him (impenv).
Benevolence (BE)	12. It's very important to him to help the people around him. He wants to care for their well-being (iphlppl). 18. It is important to him to be loyal to his friends. He wants to devote himself to people close to him (iplylfr).
Tradition (TR)	9. It is important to him to be humble and modest. He tries not to draw attention to himself (ipmodst). 20. Tradition is important to him. He tries to follow the customs handed down by his religion or his family (imptrad).
Conformity (CO)	7. He believes that people should do what they're told. He thinks people should follow rules at all times, even when no one is watching (ipfrule). 16. It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong (ipbhprp).
Security (SEC)	5. It is important to him to live in secure surroundings. He avoids anything that might endanger his safety (impsafe). 14. It is important to him that the government insures his safety against all threats. He wants the state to be strong so it can defend its citizens (ipstrgv).
Power (PO)	2. It is important to him to be rich. He wants to have a lot of money and expensive things (imprich). 17. It is important to him to get respect from others. He wants people to do what he says (iprspt).
Achievement (AC)	4. It's important to him to show his abilities. He wants people to admire what he does (ipshabt). 13. Being very successful is important to him. He hopes people will recognize his achievements (ipsuces).
Hedonism (HE)	10. Having a good time is important to him. He likes to 'spoil' himself (ipgdtim). 21. He seeks every chance he can to have fun. It is important to him to do things that give him pleasure (impfun).
Stimulation (ST)	6. He likes surprises and is always looking for new things to do. He thinks it is important to do lots of different things in life (impdiff). 15. He looks for adventures and likes to take risks. He wants to have an exciting life (ipadvnt).

land (2,141), Turkey (1,856), and Ukraine (2,031), total = 47,537.¹

Testing invariance

In recent years there has been a growing collection of literature on testing for measurement invariance of theoretical constructs in cross-country comparisons (Billiet 2003; Harkness Van de Vijver, and Mohler 2003; Hui and Triandis

1985; Cheung and Rensvold 2000, 2002) and in longitudinal studies (e.g., Chan 1998; Meade, Lautenschlager, and Hecht 2005; Wicherts, Dolan, Hessen, Oosterveld, van Baal,

¹ Details on data collection techniques in each country are documented at the website http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=10&Itemid=123. The data were taken from website <http://ess.nsd.uib.no>.

Boomsma, and Span 2004). It is suggested that assessing the comparability of measurements developed in one country to other countries is an important step prior to conducting cross-country and temporal comparisons of factors means and their correlates. If one does not test for invariance, comparisons are problematic, and conclusions are at best ambiguous and at worst severely biased.

Measurement invariance refers to whether or not, under different conditions of observing and studying a phenomenon, measurement operations yield measures of the same attribute (Horn and McArdle 1992:117). Determining invariance may involve different techniques and different approaches. For example, De Beuckelaer (2005) and De Jong, Steenkamp and Fox (2007) discuss the item response theory (IRT) approach, and Saris and Gallhofer (2007) show how to test for cognitive equivalence. However, multiple group confirmatory factor analysis (MGCFA; Jöreskog, 1971) is one of the most popular techniques used today. It can provide a statistical test for the measurement invariance of factors over groups or temporally. Furthermore, it enables researchers to decide which type of statistics may be compared, given different levels of invariance. MGCFA has been criticized in the literature by several authors. Lubke and Muthén (2004) argued that an analysis of Likert data under the assumption of multivariate normality may distort the factor structure differently across groups. They propose fitting a model for ordered categorical outcomes. However, in simulation studies, De Beuckelaer (2005) demonstrated that using Likert scales and skewed data does not significantly affect the probability of incorrect conclusions in MGCFA. Welkenhuysen-Gybels and Billiet (2002) proposed, as an alternative approach, the proportional odds model (POM) technique which has more statistical power (low type II error) than MGCFA. However, they also found out that this technique produces inflated type I error and that MGCFA is more flexible (Welkenhuysen-Gybels and Billiet 2002; Welkenhuysen-Gybels 2004).

Several sources describe methods for the evaluation of factorial measurement invariance (e.g., Cheung and Rensvold 2002; De Beuckelaer 2005; Steenkamp and Baumgartner 1998; Vandenberg 2002; Vandenberg and Lance 2000). Steenkamp and Baumgartner provide procedural guidelines for assessing measurement invariance in cross-national studies using MGCFA, and in this analysis I draw upon their approach.

The lowest level of invariance is 'configural' invariance (which is sometimes referred to as 'weak factorial invariance', Horn and McArdle 1992). Configural invariance requires that the same pattern of fixed and free factor loadings is specified for each country and at each time point. That is, the confirmatory factor analysis confirms that the items in the measuring instrument exhibit the same configuration of loadings in all countries and time points in the study. Configural invariance must be established in order for subsequent tests to be meaningful. Configural invariance is supported if a multiple-group model specifying which items measure each factor fits the data well, all item loadings are significant, and the correlations between the factors are less than one in all countries and time points. The latter requirement guarantees

discriminant validity between the factors.

Configural invariance does not guarantee that the relationships between factors and items are the same across countries and over time. Even when the same items form a factor that represents each construct, the factor loadings may still be different. The test of the next higher level of invariance assesses a necessary condition for equivalence of meaning. This level, 'metric' invariance, (which is sometimes referred to as 'measurement invariance' or 'strong factorial invariance', Vandenberg and Lance 2000:12) is a necessary condition to conduct a comparison of factors correlates. It requires that the factor loadings between items and factors are invariant across nations and over time. It is tested by restricting the factor loading of each item on its corresponding factor to be the same across groups.

$$\Lambda^1 = \Lambda^2 = \Lambda^3 = \dots = \Lambda^G \quad (1)$$

$G = \text{group number}; \Lambda = \text{factor loading}$

Metric invariance is supported if such a model fits the data well in a multiple-group confirmatory factor analysis. Metric invariance must be established in order for subsequent tests to be meaningful.

A third level of invariance should be assessed to justify comparing the means of the underlying constructs across countries or over time. Such comparisons are meaningful only if 'scalar' invariance of the items is insured (Meredith 1993; Steenkamp and Baumgartner 1998). Scalar invariance, or a test of the null hypothesis in which intercepts of like items regressions on the latent variable are invariant across groups, implies that cross-country and temporal differences in the means of the observed items are a result of differences in the means of their corresponding factors.

To assess scalar invariance, one constrains the intercepts of the underlying items to be equal across countries and time points, and tests the fit of the model to the data.

$$\tau^1 = \tau^2 = \dots = \tau^G \quad (2)$$

$G = \text{group number}; \tau = \text{item intercept}$

In sum, in order to conduct a comparison of factor means across countries and over time and interpret it meaningfully, three levels of invariance are required, configural, metric and scalar. Only if all three types of invariance are supported can we confidently carry out mean comparisons.

Research with the new instrument to measure values in the European Social Survey (ESS) 2002-2003 included a strict test of measurement invariance and evidenced support for its basic structure (Davidov, Schmidt, and Schwartz, in press).² Seven value types from the original 10 values postulated by the theory were identified in 20 countries. Three pairs of values had to be unified because they were interdependent: power with achievement, conformity with tradition,

² Although this study did not test to what extent the basic structure is circular.

and universalism with benevolence. However, the values that had to be unified are adjacent in the circular theoretical structure. Five additional paths were introduced: between the unified factor universalism-benevolence and the items 'important to be rich' and 'important to have adventures'; between the unified factor conformity-tradition and the item 'important to get respect from others'; between the unified factor power-achievement and the item 'important to be modest'; and between the unified value conformity-tradition and the item 'important to be rich'. All values displayed metric invariance across countries. In the next section we will perform measurement invariance tests with the ESS data on human values collected in the second round (2004-2005).

Data analysis

In the first step of the data analysis, 25 variance-covariance matrices were computed for each country using pair-wise deletion. These matrices are used to estimate the model in each country. By using variance-covariance matrices rather than raw data for the analyses, the researcher is provided with modification indices from the program which report about model misspecifications. After necessary model modifications, each model was re-run using the raw data and the full information maximum likelihood (FIML) procedure to deal with the problem of missing values (see Schafer and Graham, 2002). The reported results are based on this second procedure. The program Amos 6.0 is used in all the analyses reported here (see Arbuckle 2005).³ In the first part of this section, the cross-country invariance of the values measurements in the second ESS round is tested. In the second part, the temporal invariance of the values is assessed in each country separately. In both parts, the basic model tested is the final invariant model in Davidov et al. (in press) that was supported by data of the first ESS round. Finally, because temporal scalar invariance is established, aggregate value means are compared in each country between 2002-2003 and 2004-2005.

Cross-Country Invariance

The first question presented here regards how well the data in the second round of the ESS measure the values in the 25 countries participating in this round. Byrne (2001, pp. 175- 176) emphasized the importance of conducting single-country analyses prior to analyzing a multiple-group comparison. Separate CFAs in each country with 10 value factors correlating with each other reveal that all items have a substantial factor loading on their corresponding value factor. However, in each country, at least two pairs of values are strongly related to each other and cannot be modeled separately (i.e., the solution was not admissible). To solve this problem, such value pairs were unified. Table 3 reports the number of value factors found in each country. According to the test of dimensionality, only five to eight distinct values could be identified in each country. 8 distinct values are identified in Belgium, Finland, France, Iceland, Luxembourg, Netherlands, Norway, Poland, Sweden, and Switzerland. Only 5 values are identified in the Czech Republic,

Table 3: Number of Values Found in each Country after Unifying Strongly Related Values in Single-Country CFAs

Country	Number of Values, ESS Round 2
1. Austria	7
2. Belgium	8
3. Czech Republic	5
4. Denmark	7
5. Germany	7
6. Finland	8
7. France	8
8. Great Britain	6
9. Greece	5
10. Hungary	5
11. Ireland	5
12. Netherlands	8
13. Norway	8
14. Poland	8
15. Portugal	6
16. Slovenia	6
17. Spain	5
18. Sweden	8
19. Switzerland	8
20. Estonia	7
21. Iceland	8
22. Luxemburg	8
23. Slovakia	6
24. Turkey	6
25. Ukraine	6

^AFor a full description of the abbreviations of values, see Table 2.

^BOnly 19 countries collected data twice, i.e., in Round 1 and Round 2. The last 6 countries (20-25) started to collect value data only in the second round.

Greece, Hungary, Ireland, and Spain. All unified values are adjacent in the value circle of the theory (Figure 1). Therefore, unifying them does not contradict theory. It suggests, however, that the low number of items per value factor (2 with 3 for universalism) in the ESS may not capture all the fine-tuned distinctions of the different values and requires, therefore, reducing the number of factors.⁴

Next, I turn to the simultaneous MGCFA that includes 25 countries in order to assess configural invariance. Since the theoretical 10 values cannot be identified, the analysis starts with the model found for the data of Round 1 (Davidov et al. in press) to test whether this solution also fits the data

³ In order to identify the CFA models, the factor loading of one indicator per factor was constrained to 1. When mean information was included in the model to test for scalar invariance, the means of the values were fixed to zero in the reference group (for further details, see, e.g., Arbuckle 2005).

⁴ Knoppen and Saris (2007) show that the items measuring universalism or power in the ESS are too different to be constituents of the same value. They suggest dropping these values from empirical tests because of their low face value. These values (e.g., power) may form part of the unified values of this study. Results of the single country analyses are consistent with similar findings from the first ESS round.

from Round 2. This solution includes 7 values and 5 cross-loadings. However, in 11 countries, at least one additional pair of values are strongly related to each other and cannot be modeled separately (i.e., the solution was not admissible). Interestingly, these are the same countries where less than 7 values were identified in the single-country analyses reported in Table 3. The measurement of these values does not capture the minute differences between the values in these countries.

As a result I decided to reduce the number of countries. Thus, the model was composed of the 14 countries for which the 7-value solution fits: Austria, Belgium, Denmark, Estonia, Finland, France, Germany, Iceland, Luxemburg, Netherlands, Norway, Poland, Sweden, and Switzerland. Eleven of these were also included in the first ESS round. Figure 2 shows the model with the seven value factors. In this model, universalism was unified with benevolence, power with achievement, and conformity with tradition. Due to high correlations between them, they could not be modeled separately. These unified values are adjacent in Schwartz' theoretical circle, and they share a broad motivational goal. This finding suggests that the value scale also captures the motivational continuum of values in the theory with the data of the second ESS round, but only for these 14 countries. In addition, the model of the 14 countries achieves configural invariance. In this model, a fifth path between the unified value conformity-tradition and the item 'important to be rich' (which was necessary in the analysis with data from the first ESS round) was not significant with this data. The reason may be the fact that, in part, different groups of countries were included in the multiple-group comparison with data from the first and the second rounds. Interestingly, the other four additional paths were significant in the same manner as with the first round data. It should be noted that they include paths between indicators and motivationally opposed latent factors that were formed by combining two value constructs. The negative paths indicate that the association between the opposing latent value factors did not capture all of the opposition. The positive paths indicate that these associations overestimated the opposition. Probably the unification of three pairs of values into one factor necessitated adding these paths. It could well be the case that with the original 40 items (rather than the reduced version of 21 items) we would have identified 10 values, model fit would have been even better, and secondary paths would not have been needed.

Table 4 presents sequentially the fit indices of the different models. Model 1 is the basic configural invariant model with 14 countries. Its fit indices reveal a good fit to the data (CFI = .914, RMSEA = 0.013, Pclose = 1.00, AIC = 15,488, BCC = 15,523, Chi-square = 12,996, degrees of freedom=2,282).⁵ The fit indices are satisfactory for not rejecting the model (Hu and Bentler 1999; Marsh, Hau and Wen 2004). In other words, configural invariance is supported in this model, and the model configuration can thus be treated as invariant across the 14 countries.

Only 11 countries out of 20 participating in the first ESS round also displayed configural invariance in the second round. Eight other countries did not display configural invariance (the 20th country in the first round, Israel, did

not participate in the ESS the second time). An important implication of this result is that invariance across the same countries does not seem to remain stable over time. This is surprising and worth future research. In-depth interviews may provide a supplementary tool to investigate whether the meaning associated to the value questions has changed across countries and temporally. In future research, alternative techniques to assess cognitive equivalence suggested by Saris and Gallhofer (2007) may provide possible explanations.

Now it is time to turn to the next level of invariance, the metric invariance test. In particular, this test provides an answer to the question of the extent to which the values are related to the question items equally across countries. This is a necessary condition to guarantee that people understand the questions equally across the 14 countries. Establishing metric invariance will allow the comparison of associations between values and other theoretical constructs of interest. In order to test for metric invariance, the factor loadings between the indicators and their corresponding factors were constrained to be the same in each country. This test was conducted on the same (configural invariance) model, since configural invariance is a pre-condition for metric invariance.

Model 2 in Table 4 summarizes the fit indices of the metric invariance model. These indices suggest that one should not reject the model (CFI = .903, RMSEA = 0.014, Pclose = 1.00, AIC = 16,509, BCC = 16,536, Chi-square = 14,511, degrees of freedom = 2,529). Although the BCC and AIC increased, and one may do better by allowing variation in the factor loadings across countries, other indices indicate that the metric invariance model (Model 3) is still good itself⁶ (Hu and Bentler 1999; Marsh et al. 2004). Allowing some of the factor loadings to vary across countries and constraining only two factor loadings per factor to be equal in each country could improve this model. It would imply resorting to partial metric invariance, which is sufficient for cross-country comparison of construct associations (Byrne, Shavelson, and Muthén 1989; Steenkamp and Baumgartner 1998). However, as the fit indices of the full metric invariance model are satisfactory, partial metric invariance is not necessary. The unstandardized invariant factor loadings of the metric invariance model are reported in Appendix A. These factor loadings are very similar to those reported in the analysis with the first round data.

Finally, I tested for scalar invariance. Establishing scalar invariance will allow the comparison of factor means. This test is done by constraining the intercepts of the items to be equal across the 14 countries in addition to the factor loadings between the items and the factors. The fit indices are presented in Model 3 in Table 4. They suggest that this model should be rejected. Only the Pclose and the RMSEA are indicative of a good fit. Thus, one can conclude that the value scale does not meet the scalar invariance test across these countries.

⁵ Because of the large number of cases, the chi-square values are relatively high.

⁶ Because of this very large sample size, we do not apply the chi-square difference test (Cheung and Rensvold 2002).

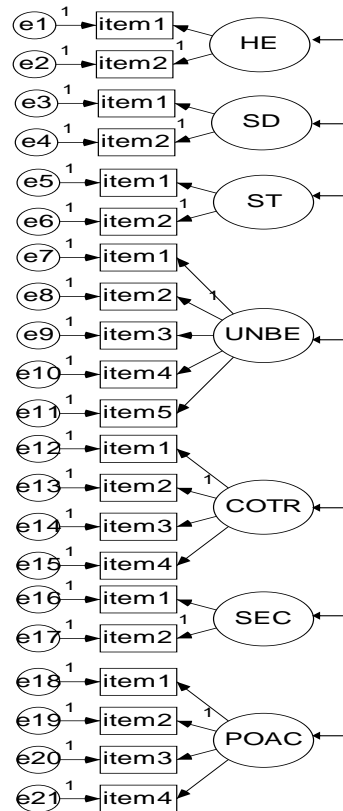


Figure 2. A confirmatory factor analysis with seven value factors for 14 countries. For a full description of the abbreviations of values see Table 2. The unified factors are measured by the items of the respective factors. For example, the unified factor Universalism-Benevolence (UNBE) is measured by the three universalism items and the two benevolence items. For item names see Table 2.

Table 4: Fit Measures of a Multi-Group Confirmatory Factor Analysis of Seven Values, Constraining Configural Invariance, Metric, and Scalar Invariance across 14 Countries

Model type	CFI	RMSEA	PCLOSE	AIC	BCC	Chi-Square	df
1. Basic model: configural invariance only (Davidov et al, in press)	.914	.013	1.000	15,488	15,523	12,996	2,282
2. Metric invariance	.903	.014	1.000	16,509	16,536	14,511	2,529
3. Scalar invariance	.794	.019	1.000	29,778	29,798	28,326	2,802

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; PCLOSE = probability of close fit; AIC = Akaike information criterion; BCC = the Browne- Cudeck criterion; df = degrees of freedom. For details see, e.g, Arbuckle (2005).

This finding suggests that one may not compare value means across the set of 14 countries. However, it may still be possible to compare the national value means across a smaller set of countries. In these countries scalar invariance may hold for some or all values. Sörbom (1974) has shown that in order to be able to compare means across groups, mean levels have to be constrained to zero in one group to get the model identified (for alternatives see Little, Slegers,

and Card 2006). Researchers interested in the comparison of mean levels of particular values in certain countries may follow the same sequence of analyses to test for invariance of the values before carrying out the comparison. Thus, since invariance is not established throughout the full set of countries participating in the second round of the ESS, it is suggested to compare only a few countries that are invariant, that are of interest to the substantive study and that the re-

searchers are familiar with.⁷

Cross-Time Invariance

Now we turn to the invariance test within countries and over time. A total of 19 countries participated in the first two rounds of the ESS and these allow testing for aggregate change of the value means across the two years between the measurements. Once again, the analysis in each country over time starts with the model found for the data from Round 1. Using this model I tested whether this solution also fits for the longitudinal comparison. This model includes seven values and five cross-loadings. If unacceptable fit is determined, then I considered what adjustments have to be made in the analysis.

Sequential configural, metric, and scalar invariance tests were performed in each country. Nine countries passed the metric and scalar invariance tests and no modifications were needed. In 10 other countries, metric and scalar invariance were established after introducing a few modifications to achieve an acceptable model fit. In eight countries, two or three *adjacent* values had to be unified since they could not be modeled separately. In other words, the basic model from Davidov et al. (in press) could not be retained, and less than 7 values could be identified in these countries.⁸ Other modifications included an error covariance, an additional cross-loading, or releasing an equality constraint of the intercepts. The modifications required in each country and the global fit measures are reported in Table 5. One can see that, after the modifications, all countries meet the temporal metric and scalar invariance tests. The meaning of the values identified did not change over the two years between the measurements. Value means may be compared in the same countries. Such a comparison does not allow estimating the absolute mean in each wave, but rather the aggregate mean difference in the values between the waves (Sörbom 1974).

Observing aggregate mean differences between Round 1 and Round 2 reveals a few findings. In a total of 122 values (the total number of values identified in 19 countries, according to the information on the unified values in column 2 of Table 5), 53 values had a significant ($p < 0.05$) aggregate change over time in the same country. However, although significant, most of these changes were practically negligible. Only 13 of the aggregate mean value changes were larger than 0.15, and only 3 of them were larger than 0.2. In five countries (France, Hungary, Netherlands, Slovenia and Sweden) the country mean of only one value changed significantly, and this change was very small. The smallest changes were observed for the values stimulation and self-direction. Table 6 summarizes values changes over time for each of the countries (empty cells represent no change).

Summary and Discussion

The designers of the European Social Survey chose Schwartz (1992) human values theory as the basis for developing the human values measurement instrument included in the core of the survey. In the current study the following three research questions were investigated:

1. Does the model found for the ESS data of Round 1 (Davidov et al., in press) fit the data in Round 2?
2. Are the values comparable across the full set of countries participating in the second ESS round? and
3. Are they comparable over time in each country separately?

In 25 separate confirmatory factor analyses in each country, substantial factor loadings were found between the value factors and their indicators. However, several pairs of values displayed very high correlations. As a result, these values had to be unified. After repeating the CFAs in each country with the unified value factors I identified between 5 and 8 values in each country. The unified values are adjacent in the theoretical value circle and represent closely linked motivations. Therefore, they do not violate the theorized circular motivational structure. Unifying them may have been a consequence of the fact that the original number of 40 items was reduced to 21 items in the ESS for measuring the 10 values. The need to unify them suggests that the ESS value scale may not capture all of the fine-tuned distinctions in the theory. Other studies find that by using a larger number of items, one can identify more values (e.g., Schmidt, Bamberg, Davidov, Herrmann, and Schwartz 2007; Schwartz and Boehnke 2004).

Next, tests were performed assessing the invariance of the value measurements with the Round 2 data. This started with the assessment of configural invariance by computing the confirmatory factor analysis for the 25 countries simultaneously. The MGCFA did not support configural invariance of a model with 25 countries, because in 11 countries at least one additional pair of values had to be unified and less than seven values could be identified. Therefore, the number of countries was reduced to 14. With this set of countries the model was acceptable and it displayed configural invariance across the 14 countries with seven values. However, only 11 out of 20 countries participating in the first round displayed configural invariance with the new data, suggesting that cross-country invariance may not be stable over time. In future research, alternative techniques to assess cognitive equivalence, suggested by Saris and Gallhofer (2007), may provide possible explanations for this finding.

The current analyses suggest that one can use the scale in the second ESS round to measure only seven distinct values with confidence across these countries. A higher number of values may be identified if a smaller number of countries is compared. For example, in France and Belgium, two neighboring countries that partially share the same language (French), where eight distinct values were identified in the single CFA analyses, one can identify the same eight values also in a MGCFA.

⁷ Future studies should repeat the current analyses to reassess invariance for particular countries of interest. Furthermore, cognitive interviews offer a supplementary tool to assess the equivalence of meaning of the values instrument in various countries.

⁸ One can think of an alternative strategy: dropping the unified values since their meaning is questionable.

Table 5: Metric and Scalar Invariance of a Multiple-Group Comparison of Round 1 and Round 2 in each Country, Global Fit Measures ^A

Country	Values included in the model	Modifications needed	Global fit measures: metric invariance test			Global fit measures: scalar invariance test		
			Pclose	RMSEA	CFI	Pclose	RMSEA	CFI
1. Austria			1.00	0.037	0.925	1.00	0.038	0.916
2. Belgium		Equality of intercepts of ipfrule and ipshabt released	1.00	0.032	0.909	1.00	0.034	0.901
3. Czech Republic	UNBE, POAC, COTR, SEC, STSDHE	Unifying HE, ST, SD; UNBE → ipgdtim, ipsuces, impfun	1.00	0.037	0.910	1.00	0.037	0.906
4. Denmark			1.00	0.032	0.916	1.00	0.032	0.914
5. Germany			1.00	0.034	0.917	1.00	0.034	0.914
6. Finland			1.00	0.033	0.925	1.00	0.033	0.924
7. France			1.00	0.034	0.915	1.00	0.033	0.913
8. Great Britain	UNBE, POAC, COTR, SEC, HE, STSD	Unifying ST, SD; Error(ipadvnt) ↔Error(impsafe)	1.00	0.037	0.903	1.00	0.038	0.894
9. Greece	UNBE, POAC, COTR, SEC, STSDHE	Unifying HE, ST, SD; Error(imprich) _{j-ζ} Error(ipgdtim); POAC ipudrst;	1.00	0.042	0.903	1.00	0.042	0.899
10. Hungary	UNBE, POAC, COTR, SEC, HE, STSD	Unifying ST, SD; SEC → impfree	1.00	0.037	0.904	1.00	0.037	0.902
11. Ireland	UNBE, POAC, COTR, SEC, HE, STSD	Unifying ST, SD	1.00	0.034	0.907	1.00	0.034	0.903
12. Netherlands			1.00	0.033	0.920	1.00	0.033	0.919
13. Norway			1.00	0.036	0.909	1.00	0.036	0.906
14. Poland			1.00	0.034	0.926	1.00	0.035	0.921
15. Portugal	UNBE, POAC, COTR, SEC, HEST, SD	Unifying ST, HE	1.00	0.045	0.912	1.00	0.045	0.908
16. Slovenia	UNBE, POAC, COTR, SEC, STSDHE	Unifying ST, SD, HE; UNBE → impfree, ipmodst; SDSDHE → iplyfr	1.00	0.037	0.903	1.00	0.038	0.895
17. Spain	UNBE, POAC, COTR, SEC, HE, STSD	Unifying ST, SD	1.00	0.039	0.911	1.00	0.040	0.905
18. Sweden		Error(ipmodst) ↔ Error(ipudrst); ST → ipstrgv	1.00	0.035	0.909	1.00	0.034	0.908
19. Switzerland			1.00	0.034	0.904	1.00	0.034	0.901

^AFor a full description of the abbreviations of values and value indicators, see Table 2. If not otherwise indicated in columns 2 and 3, the model in the test is the same model tested in the cross-country analyses with 7 values (HE, ST, SD, SEC, and the unified values UNBE, POAC and COTR) and 5 cross-loadings. All countries passed the metric and scalar invariance tests for the values compared across rounds. →Signifies that a modification requires releasing the equality constraint on the corresponding factor loading; Intercept(indicator name): a modification requires releasing the equality constraint on the intercept; ↔ signifies that a modification requires estimating the covariance.

Table 6: Aggregate Change in Value Means from Round 1 to Round 2.^A

Country	HE	ST	SD	UNBE	COTR	SEC	POAC	STSD	STSDHE
1. Austria	.226**			-.059**					
2. Belgium	-.073*	-.089*			.188**	.078*	-.110**		
3. Czech Republic				-.202**	-.156**	-.156**	.357**		.105**
4. Denmark			-.136**		.119**				
5. Germany	.062*		-.083**		.103**	-.168**			
6. Finland		-.076*				-.073*			
7. France							-.066*		
8. Great Britain	.091*			.106**	.097**	.071**			
9. Greece				.061**	.090**	.072**			-.075**
10. Hungary				.087**					
11. Ireland				.094**	.178**				
12. Netherlands				.047**					
13. Norway		.072*			.100**		.060*		
14. Poland				.037*			.079*		
15. Portugal			-.192**	-.165**	.118**	-.110**	.135**		
16. Slovenia							-.072*		
17. Spain	-.135**				-.167**	-.188**	-.138**	-.108**	
18. Sweden					.073*				
19. Switzerland	-.060*			.048**	.186**				

^AFor a full description of the abbreviations of values, see Table 2. Only significant changes are reported. * $P < 0.05$; ** $P < 0.01$

Next I tested for metric invariance. Establishing metric invariance led to the conclusion that the meaning of the values, as measured by the indicators of the ESS, is probably the same in these 14 countries. In spite of cultural differences, people appear to understand the meaning given to the values by their indicators in a similar manner. Cognitive tests in each country could supplement these findings (Prüfer, Rexroth, and Fowler 2004). However, scalar invariance was not supported by the data. Researchers may not use the value instrument to compare value means across these countries. Smaller sub-sets of countries may meet the scalar invariance test, at least partially. When scalar invariance for some or all values is found, national means may be compared.

Finally, we tested the invariance of the values in the same country over time. Such invariance is a pre-condition for comparing the means of the values over time and inspecting value change. The MGCFA analyses in each of the 19 countries in which values were measured in both ESS waves supported metric and scalar invariance. However, the model found for the data of Round 1 fitted only in nine countries. In eight countries, two or three *adjacent* values had to be unified since they could not be modeled separately. In these countries the value measurements do not capture the minute differences between the values, and the meaning of the unified values is questionable. In spite of this limitation, the longitudinal metric and scalar invariance found in the ESS data allows researchers to investigate value change over the two years between the first and the second rounds of the ESS.

Based on the current findings, researchers may also use the ESS value scale of the second round to confidently carry out comparisons of the relationships between individual values and other variables such as socio-demographic character-

istics, attitudes, and behavior across 14 countries. Similarities or differences across countries in the relations of value priorities to behavior and attitudes may be meaningfully interpreted.

The increasing availability of cross-national data sets nowadays allows researchers to conduct comparisons across countries and over time. Guaranteeing invariance provides the critical legitimacy for comparative work. Without first establishing measurement invariance, any type of comparison is questionable. Establishing invariance should also be done with panel data (for an application see, e.g., Schlüter, Davidov, and Schmidt 2006). In this study I illustrated how invariance may be tested across countries and over time. Researchers may also be interested to compare mean values of *different* countries over time. In such a case, similar procedures may be used.

Recent studies suggest that when full or partial measurement invariance is not guaranteed, it may still be the case that factors are equivalent. Saris and Gallhofer (2007, ch. 16) indicate that the test of measurement invariance is too strict and may fail even when functional equivalence holds. In other words, although there may be cognitive invariance, the measurement invariance test could be rejected due to differences between measurement features of the questions in the different countries. For instance, different reactions to the method used in various countries could cause the measurement invariance test to fail, although the link by definition and by intuition is invariant (for a demonstration see, e.g., Van der Veld 2007). An alternative procedure would be to conduct complementary cognitive tests in various countries. However, such additional tests are expensive and may require a lot of work. As long as such measures are not available, es-

tablishing invariance should be routinely practiced to maintain meaningful cross-country and longitudinal comparisons.

Acknowledgements

An earlier version was presented at the 2nd ESRA meeting in Prague, in June 2007. The author would like to thank Shalom H. Schwartz, Peter Schmidt, and two anonymous reviewers for very helpful comments on this paper. The author would also like to thank Lisa Trierweiler for the English proof of the text.

References

- Allport, G. W., Vernon, P. E., & Lindzey, G. A. (1960). *A study of values*. Boston: Houghton Mifflin.
- Arbuckle, J. L. (2005). *Amos 6.0 user's guide*. (Chicago, IL: SPSS)
- Billiet, J. (2003). Cross-cultural equivalence with structural equation modeling. In J. A. Harkness, F. J. R. V. de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (p. 247-264). New York: John Wiley.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *11*, 305-314.
- Byrne, B. M. (2001). *Structural equation modeling with amos. basic concepts, application, and programming*. London: Lawrence Erlbaum Associates.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (lmcs) and multiple indicator latent growth modeling (mlgm). *Organizational Research Methods*, *1*, 421-483.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*, 187-212.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.
- Davidov, E., Schmidt, P., & Schwartz, S. (n.d.). *Bringing values back in: Testing the adequacy of the european social survey to measure values in 20 countries*. (Public Opinion Quarterly)
- De Beuckelaer, A. (2005). *Measurement invariance issues in international management research*. (Unpublished doctoral dissertation, Limburgs Universitair Centrum, Limburg, the Netherlands)
- De Jong, M. G., Steenkamp, J.-B. E. M., & Fox, J.-P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical irt model. *Journal of Consumer Research*, *34*, 260-278.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155-174.
- Feldman, S. (2003). Values, ideology, and structure of political attitudes. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (p. 477-508). New York: Oxford University Press.
- Halman, L., & de Moor, R. (1994). Value shift in western societies. In P. Ester, L. Halman, & R. de Moor (Eds.), *The individualizing society: Value change in europe and north america* (p. 1-20). Tilburg: Tilburg University Press.
- Harkness, J. A., Van de Vijver, F. J. R., & Mohler, P. P. (2003). *Cross-cultural survey methods*. New York: John Wiley.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117-144.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, *16*, 131-152.
- Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.
- Kluckhohn, C. K. M. (1951). Values and value orientations in the theory of action. In T. Parsons & E. Shils (Eds.), *Toward a general theory of action*. Cambridge: Harvard University Press.
- Knoppen, D., & Saris, W. (2007, June). *Measuring value orientations across nations: Analysis of the quality of the european social survey for measuring values*. (Paper presented at the 2nd European Survey Research Association (ESRA) conference, Prague, Czech Republic)
- Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in sem and macs models. *Structural Equation Modeling*, *13*(1), 59-72.
- Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, *11*, 514-534.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing hu and bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341.
- Meade, A. W., Lautenschlager, G. J., & Hecht, J. E. (2005). Establishing measurement equivalence and invariance in longitudinal data with item response theory. *International Journal of Testing*, *5*, 279-300.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.
- Prüfer, P., Rexroth, M., & Fowler, F. J. (2004). *Quest 2003. questionnaire evaluation standards*. (ZUMA-Nachrichten Spezial, Band 9, Mannheim, Germany: ZUMA)
- Rokeach, M. (1973). *The nature of human values*. New York: Free Press.
- Sagiv, L., & Schwartz, S. H. (1995). Value priorities and readiness for out-group social contact. *Journal of Personality and Social Psychology*, *69*, 437-448.
- Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation and analysis of questionnaires for survey research*. New York: Wiley.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177.
- Schlüter, E., Davidov, E., & Schmidt, P. (2006). Applying autoregressive cross-lagged and latent growth models to a three-wave panel study. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (p. 315-336). Mahwah, NJ: Lawrence Erlbaum Publishers.
- Schmidt, P., Bamberg, S., Davidov, E., Herrmann, J., & Schwartz,

- S. (2007). Die messung von werten mit dem 'portraits value questionnaire' [the measurement of values with the 'portraits value questionnaire']. *Zeitschrift fr Sozialpsychologie*, 38(4), 249-263.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, 25, 1-65.
- Schwartz, S. H. (1994). Are there universal aspects in the content and structure of values? *Journal of Social Issues*, 50, 19-45.
- Schwartz, S. H. (2005). Robustness and fruitfulness of a theory of universals in individual human values. In A. Tamayo & J. B. Porto (Eds.), *Valores e comportamento nas organizações [values and behavior in organizations]* (p. 56-95). Petrópolis, Brazil: Vozes.
- Schwartz, S. H., & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, 38, 230-255.
- Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross Cultural Psychology*, 32, 519-542.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research*, 25, 78-90.
- van der Veld, W. M. (2007, June). *The quality of social trust items across europe*. (Paper presented at the 2nd European Survey Research Association (ESRA) conference, Prague, Czech Republic)
- Vandenberg, R. J. (2002). Towards a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5, 139-158.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Welkenhuysen-Gybels, J. (2004). The performance of some observed and unobserved conditional invariance techniques for the detection of differential item functioning. *Quality & Quantity*, 38, 681-702.
- Welkenhuysen-Gybels, J., & Billiet, J. (2002). A comparison of techniques for detecting cross-cultural inequivalence at the item level. *Quality & Quantity*, 36, 197-218.
- Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., Baal, G. C. M. van, & al., D. I. B. et. (2004). Are intelligence tests measurement invariant over time? investigating the nature of the flynn effect. *Intelligence*, 32, 509-537.
- Williams, R. M. J. (1968). Values. In D. L. Sills (Ed.), *International encyclopedia of the social sciences (vol. 16)*. New York: Macmillan.

Appendix A: Unstandardized Regression Coefficients for 14 Countries in the Metric Invariant Model (Model 2 in Table 4).

	SD	UNBE	COTR	SEC	POAC	HE	ST
ipertiv	1.00						
impfree	.74						
ipeqopt		1.00					
ipudrst		1.15					
impenv		1.13					
iplylfr		1.03					
iphlppl		1.28					
imptrad			.89				
ipmodst			.84		-.40		
ipfrul			1.00				
ipbhprp			1.16				
impsafe				1.00			
ipstrgv				.93			
imprich		-.52	.01		1.00		
iprspt			.37		.87		
ipshabt					1.30		
ipsuces					1.38		
ipgdtim						1.00	
impfun						1.10	
impdiff							1.00
ipadvnt		-.99					1.35

The path coefficients added to the original model in the modification process are in boldface type. The path from COTR to imprich is not significant. All other coefficients are significant, ($p < 0.01$). Empty cells represent no direct relation between the values and the indicators.