# Testing for comparability of human values across countries and time with the third round of the European Social Survey

Davidov, Eldad

# Testing for comparability of human values across countries and time with the third round of the European Social Survey

**Eldad Davidov**

**University of Zurich, Switzerland**

This paper has been accepted for publication in *International Journal of Comparative Sociology* and the final (edited, revised and typeset) version of this paper will be published in *International Journal of Comparative Sociology*, 51/3, June 2010 by Sage Publications Ltd.

DOI: 10.1177/0020715210363534

URL: http://cos.sagepub.com/cgi/content/abstract/51/3/171

Acknowledgement: I would like to thank two anonymous reviewers for very helpful comments on an earlier version of this manuscript. Many thanks to Lisa Trierweiler for the English proof of the manuscript.

# Testing for comparability of human values across countries and time with the third round of the European Social Survey

Abstract

This study tests the compatibility and comparability of the human values measurements from the third round of the European Social Survey (ESS) to measure the 10 values from Schwartz' (1992) value theory in 25 countries. Furthermore, it explains the dangers associated with ignoring non-invariance before comparing the values across nations or over time, and specifically describes how invariance may be tested. After initially determining how many values can be identified for each country separately, the comparability of value measurements across countries is assessed using multigroup confirmatory factor analysis (MGCFA). This is necessary to allow later comparisons of values' correlates and means across countries. Finally, invariance of values over time (2002-2007) is tested. Such invariance allows estimating aggregate value change and comparing it across countries meaningfully. In line with past results, only four to seven values can be identified in each country. Analyses reveal that the ESS value measurements are not suitable for measuring the 10 values; therefore, some adjacent values are unified. Furthermore, a subset of eight countries displays metric invariance for seven values, and metric invariance for 6 values is found for 21 countries. This finding indicates that values in these countries have similar meanings, and their correlates may be compared but not their means. Finally, temporal scalar invariance is evidenced within countries and over time thus allowing longitudinal value change to be studied in all the participating countries.

Key words: Human values; invariance; multigroup confirmatory factor analysis (MGCFA); configural, metric and scalar invariance; latent means comparison

## 1. Introduction

Values play an important role in the social sciences. They may explain opinions, attitudes and behavior both on the individual and aggregate level. On the individual level, they may explain political attitudes, attitudes toward societal groups or social and economic policies and influence opinions and behavior. On the aggregate level, they may be related to cross-country differences in governmental policies, reflect social change and even influence its rate of change. It has been also shown (e.g., Schwartz 2007) that social structure underlies, to a large extent, value priorities. Thus, though little explored, values may be considered mediators of the effect of variables like age, gender, education or economic and professional status on attitudes, opinions and behavior by playing the role of the black box in-between (Hitlin and Piliavin, 2004). Their mean level and their effects may vary across different cultural groups, countries or even time points thus reflecting societal differences and changes.

Values are also an important component of culture. Inglehart (1990, p. 18) defines culture as ‘a system of attitudes, values, and knowledge that is widely shared within a society and transmitted from generation to generation’. He argues that culture is learned and may vary from one society to another. As it is deeply rooted within individuals, it is quite resistant to change. Major shifts in societal conditions may, however, change culture. This process is more likely to take place through intergenerational population replacement. Schwartz (2006a, p. 138) views culture as ‘the rich complex of meanings, beliefs, practices, symbols, norms, and values prevalent among people in a society’. He considers values to be an efficient measure of culture. In general, Schwartz argues, culture is hard to measure. Films, stories, laws, economic institutions, social habits, governmental decisions are all elements and thus indirect measures of culture. However, they all have underlying value emphases that characterize these societies (Weber, 1958; Williams, 1968).

Values have been modeled, conceptualized and operationalized by different scholars early on (e.g., Allport, Vernon and Lindsay, 1960; Kluckhohn, 1951) in various ways (e.g., Feldman, 2003; Halman and de Moor, 1994; Rokeach, 1973). Inglehart (1977; 1990) developed a theory which focused on materialism-postmaterialism values and he later added another dimension to include modernization issues (Inglehart and Baker, 2000). Hofstede (1980, 2001) focused on work values. The Schwartz human values theory (1992, 1994, 2006a, b) was developed later out of his social psychological studies of individual differences in value priorities and their effects on attitudes and behavior (Schwartz, 2006a). The current study will exclusively focus on this theory.

In the last decades there has been an unmistakable increase in the cross-national and longitudinal study of human values (Davidov, Schmidt and Schwartz, 2008a; Hofstede, 2001; Inglehart, 1990; Inglehart and Baker, 2000; Inglehart et al., 2004; Kohn and Schooler, 1983; Rokeach and Ball-Rokeach, 1989; Schwartz, 2006a, b, 2005a,b; Schwartz et al., 2001; Triandis 1993, 1998;). In these studies, value change (Rokeach and Ball-Rokeach, 1989), value levels (Inglehart and Baker, 2000; Schwartz 2006a, 2007), effects of values on different types of attitudes (e.g., toward immigration policies; Davidov et al. 2008b) or behavior (e.g., political; Schwartz, 2005a, 2006a), or relations between values and other exogenous factors (such as religiosity; Schwartz and Huismans, 1995) have been investigated and compared across different nations. Also, the effect of socioeconomic and demographic variables on values has been explored (Schwartz, 2007). These comparative studies raise methodological challenges regarding the validity and comparability of values studied in different contexts such as nations, cultures or time. Even though the same questions are used in the different contexts, people might understand these questions differently. Respondents' use of the scale to answer the value questions might also be dependent on the temporal or cultural context.

Before cross-national and cross-time studies of values are conducted, it is crucial to guarantee that the values are invariant across groups and time points. Absent invariance, comparisons of value mean levels or their correlates are problematic (Billiet, 2003; De Beuckelaer, 2005; Steenkamp and Baumgartenr, 1998; Vandenberg, 2002). There are several statistical tools available to assess invariance and they should be used prior to any cross-national or longitudinal comparisons. Thus, the main goals of this study are:

(1) to explain why testing for invariance is necessary before comparisons are done;
(2) to present how invariance may be tested crossnationally or across time points, and to demonstrate a practical application of such a test with the human values measurements from the European Social Survey; and
(3) to discuss problems arising during the analysis of invariance.

In 2002, questions to measure the human values postulated by Shalom Schwartz (1992) were introduced in the European Social Survey (ESS), a biannual European cross-country survey. They have been included in the first three rounds of the ESS (conducted in the years 2002/2003, 2004/2005 and 2006/2007) and are also going to be included in future rounds of the survey. The addition of these questions provides researchers with the possibility to conduct cross-country comparative studies using the value concept. The methodological challenges of comparing values across nations and over time will be illustrated with these data. The third round of the ESS data will be used to investigate the cross-national comparability of the values. The first and third ESS rounds will be used to asses the intra-country temporal comparability of the values.

Earlier studies have assessed the invariance of values in the first (2002/2003) and the second (2004/2005) rounds of the ESS (Davidov et al., 2008a; Davidov, 2008). These studies suggest

that full invariance of the values (Meredith, 1993) is not supported by the data and, therefore, cannot be assumed and must be tested. Furthermore, it was found that values are rather stable within countries over a period of time of 2-3 years (between 2002 and 2005). The present study illustrates how values may be compared across countries or over time in a meaningful way. It also provides a complementary test of different levels of invariance of the value questions across countries and over time for data from the third round of the ESS (2006/2007). Before beginning with the empirical analysis, a short overview of the theory is provided.

## 2. Human Basic Values

Schwartz defines values as 'desirable, transsituational goals, varying in importance, that serve as guiding principles in people's lives' (Schwartz, 1994: 21). In his theory he proposes 10 basic values with distinct motivations building on earlier approaches (e.g., Rokeach, 1973; Inglehart, 1990). The values are: hedonism, stimulation, self-direction, security, universalism, benevolence, conformity, tradition, power and achievement. Table 1 presents the 10 values and the basic motivations behind them. For example, the motivational goal of power is social status and prestige, with control or dominance over people and resources. The motivational goal of hedonism is pleasure and sensuous gratification for oneself.

In addition, the theory suggests a structural relation between the values. Some values may be closely related to each other but others may oppose each other. In other words, actions to realize one value may be congruent or opposed to actions to realize other values. For example, pursuing power values may conflict with pursuing universalism values. Seeking social status and prestige, the core goals of power values, may obstruct activities that enhance understanding, appreciation and tolerance for other people, the core goals of universalism.

However, pursuing benevolence and universalism values may be compatible. Making efforts to understand, and be tolerant to other people may strengthen and be strengthened by activities directed toward enhancement of the welfare of people with whom one is in frequent personal contact, which is the main goal of benevolence.

Table 1 about here

Figure 1 about here

The circular structure in Figure 1 displays the structural relations among values. Values congruent with each other are close to each other in the circle and the values that are in conflict with each other are opposite to each other in the circle. Strictly speaking, the theory proposes that we distinguish between 10 values. However, it is also suggested that the values form a continuum at a more basic level because the motivational differences of values are continuous rather than discrete (Davidov et al., 2008a). Therefore, in empirical studies, adjacent values often appear as a single value rather than as distinct from each other (e.g., tradition and conformity, universalism and benevolence or power and achievement).

On a higher level, the theory suggests that the values are arranged around two bipolar dimensions. The first dimension contrasts self-transcendence, which includes universalism and benevolence values, with self-enhancement, where power or achievement values are found. The other dimension contrasts conservation, which includes the values tradition, conformity and security, with openness to change, which includes the values self-enhancement and stimulation. The value hedonism is found between the dimensions self-enhancement and openness to change (see Schwartz, 1992, 1994).

3. The questions in the ESS measuring human values

The ESS includes 21 questions to measure the 10 values. Two questions are given for each value and, as an exception, three for universalism because of its broad content. This questionnaire is based on Schwartz' original 40-item portrait values questionnaire (PVQ; Schwartz, Melech, Lehmann, Burgess, Harris and Owens, 2001; Schwartz, 2005b). However, Schwartz shortened this battery of questions to allow its inclusion in the ESS. The questions are double-barrelled and gender matched with the respondent. Schwartz (2003) has shown empirically that the fact that questions are double-barrelled does not affect the quality of the data. The questions describe a fictitious person, and the respondent is asked to rate the extent to which this person **is** or **is not** like him or her. For example, 'Having a good time is important to him. He likes to "spoil" himself' describes a person for whom hedonism is important. Respondents answer on a 6-point rating scale ranging from 'very much like me' (1) to 'not like me at all' (6). Table 2 presents the value questions and their labels, grouped by type of value.

Table 2 about here

The countries participating in the third round of the ESS (with sample size in parentheses) are: Austria (2,405), Belgium (1,798), Bulgaria (1,400), Cyprus (995), Denmark (1,505), Estonia (1,517), Finland (1,896), France (1,986), Germany (2,916), Great Britain (2,394), Hungary (1,518), Ireland (1,800), Latvia (1,960), Netherlands (1,889), Norway (1,750), Poland (1,721), Portugal (2,222), Romania (2,139), Russia (2,437), Slovakia (1,766), Slovenia (1,476), Spain (1,876), Sweden (1,927), Switzerland (1,804), Ukraine (2002), thus making a total of 47,099 participants.[1]

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

## 4. Testing invariance

Several studies in recent years have suggested that guaranteeing the comparability of theoretical constructs in one country to other countries or to other time points is necessary prior to conducting comparative analyses (Billiet, 2003; Cheung and Rensvold, 2000, 2002; Harkness, Van de Vijver and Mohler, 2003; Hui and Triandis, 1985; Meredith, 1993) (for a discussion on the choice of countries as a unit of analysis, see the summary and discussion section). If one does not test for invariance, comparisons of mean levels or correlates 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, p. 117). There have been different techniques forwarded in the literature to test for invariance (for an overview see, e.g., De Beuckelaer, 2005). However, multigroup confirmatory factor analysis (MGCFA: Jöreskog, 1971) is one of the most popular techniques. It provides researchers with tools to decide whether invariance is given or not, which indicators produce incomparability across countries and which types of statistics may be compared (correlates, mean levels or both). Although its use with Likert data (i.e., data that are obviously ordinal and often not normally distributed) has been criticized in the literature (Lubke and Muthén, 2004), researchers have shown that it still works well even when data are not continuous or normally distributed (De Beuckelaer, 2005; Welkenhuysen-Gybels and Billiet, 2002; Welkenhuysen-Gybels, 2004). In these studies, simulations are reported that examine whether assuming normality and continuity of measurement scales when using ordinal categorical scales (like Likert scales) yields different conclusions in a cross-cultural invariance test. The studies generally conclude that the maximum likelihood (ML) parameter

estimates and standard errors are rather robust for small violations of normality (see, e.g., Coenders and Saris, 1995, and Coenders, Satorra and Saris, 1997).

In the analyses, I follow procedural guidelines suggested by several authors (e.g., Cheung and Rensvold, 2002; De Beuckelaer, 2005; Steenkamp and Baumgartner, 1998; Vandenberg, 2002; Vandenberg and Lance, 2000). They describe two strategies to test for invariance. The first is the 'bottom-up' strategy. According to this strategy, one increases the number of equality constraints until no invariance is given. According to the second 'top-down' strategy, one starts with the most constrained model and releases equality constraints until the model is accepted by the data. Both strategies end up with the same conclusions. For the current study with the ESS value data I decided to implement the bottom-up strategy to inquire whether even weak forms of invariance are absent.

The lowest level of invariance is 'configural' invariance; this is sometimes referred to as 'weak factorial invariance' (Horn and McArdle, 1992). Configural invariance requires that the same indicators measure the same theoretical constructs in different groups (i.e., cultures, nations) and time points. Configural invariance is supported if a multigroup model fits the data well, all factor loadings are significant and substantial, and the correlations between the factors are less than one in all nations 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 groups and over time. To test this, a higher level of invariance is required, which presupposes configural invariance. The test of the next higher level of invariance guarantees that the factor loadings between factors and items are similar across groups or time points. It also implies that the constructs have the same content across the groups. This level

of invariance is called ‘metric’ invariance, which is also sometimes referred to as ‘measurement invariance’ or ‘strong factorial invariance’ (Vandenberg and Lance, 2000, p. 12) and is a necessary condition to conduct a comparison of factors’ correlates (i.e., unstandardized regression coefficients, covariances). It is tested by restricting the factor loading of each item on its corresponding factor to be equal across groups and. Metric invariance is supported if such a model fits the data well in a MGCFA and does not result in a significant reduction of model fit. Chen (2007) suggested ‘modern’ indicators for invariance which are especially suitable for large samples. They include differences in the indices comparative fit index (CFI) and root mean square error of approximation (RMSEA). Minimal differences in these global fit measures between the models may support a more restrictive model. Metric invariance is a necessary condition for higher levels of invariance.

A third level of invariance is necessary to allow comparison of constructs’ means. This level is called ‘scalar’ invariance (Meredith, 1993; Steenkamp and Baumgartner, 1998). Scalar invariance is tested by restricting the intercepts of each item to be the same across groups or time points. If they are equal, it implies that mean differences of the latent variables (in this case, the values) are a result of differences in the item scores and not due to differences in factor loadings or intercepts of the items. To assess scalar invariance, one constrains the intercepts of the underlying items to be equal across nations and time points, and tests the fit of the model to the data. Scalar invariance is supported if the model fit is acceptable.[2]

Research with the European Social Survey (ESS) to measure values in the 2002/2003 and 2004/2005 data included a strict test of measurement invariance (Davidov et al., 2008a; Davidov, 2008). Seven value types from the original 10 values postulated by the theory were identified with data in the first round. Three pairs of values had to be unified because they were

[2] Here mean and covariance structure (MACS) analysis is applied (Sörbom, 1974, 1978) because means and intercepts are included in the model (see Steenkamp and Baumgartner, 1998).

interdependent: power with achievement, universalism with benevolence and tradition with conformity. The values that had to be unified are adjacent to each other in the circular theoretical structure. Five additional paths were introduced: (1,2) between the unified factor universalism-benevolence and the items important to be rich and important to have adventures; (3) between the unified factor conformity-tradition and the item important to get respect from others; (4) between the unified value conformity-tradition and the item important to be rich; and (5) between the unified factor power-achievement and the item important to be modest. In the second ESS round, only 14 countries displayed metric invariance with this model and only 4 cross-loadings were significant. In the other countries less than 7 values could be identified. Scalar invariance was demonstrated within countries between the first and the second ESS rounds in all countries. This allowed the study of aggregate value change within countries over the short period between 2002/2003 and 2004/2005. In the next section I will perform measurement invariance tests with the ESS data on human values collected in the third round (2006/2007).

## 5. Data analysis

### a. Single-country analyses

Before testing the invariance of the values across countries and over time it was interesting to test the model in each country separately. Byrne (2001: 175-6) has acknowledged the importance of conducting single-country confirmatory factor analyses (CFAs) (see also Bollen, 1989) prior to the multigroup comparisons (MGCFA). At first, 25 variance-covariance matrices were constructed, one for each participating nation, as input for the models using pair-wise deletion. In the second step, all analyses were repeated using the full information maximum likelihood (FIML) procedure to account for missing values (Schafer and Graham, 2002). Since the two approaches produce similar results when there are less than 5% missing values in the data, conclusions were consistent in this study. However, the results

reported in the current study are based on the FIML procedure because it has been shown that this procedure deals more appropriately with missing values (Schafer and Graham, 2002)[3]. I used the program Amos 16.0 for all subsequent analyses (Arbuckle, 2005). Table 3 provides the results of the single-country tests.

Table 3 about here

Results of the CFA in each country show that it was not possible to identify all the 10 values postulated by the theory in any of the countries with the ESS data. Some values were too strongly related and necessitated unifying them. Between four and seven values could be identified. Seven values were identified in Austria, Denmark, France, Great Britain, Ireland, Norway, Poland, Portugal, Russia, Spain, Sweden and Switzerland. Four values were identified in Cyprus, Estonia and Slovakia. Column 2 of Table 3 reports how many values could be identified in each country. Column 3 reports the values that had to be unified because they were too closely related. It turns out that all unified values belong to the same theoretical dimension. Therefore, unifying them does not refute the theory. However, it suggests that the ESS does not offer enough question items to distinguish between each of the single values. Had there been more than two questions per value with three for universalism in the ESS, maybe more values would have been identified in each country (as was the case, e.g., in Schwartz and Boehnke, 2004). This result is also in line with findings in previous studies (e.g., Davidov et al., 2008a; Davidov, 2008)[4].

b. Multigroup analyses

[3] Since a multiple-group comparison is applied where each country constitutes a single group, it is not necessary to use the population size weight (for further details, see http://essedunet.nsd.uib.no/cms/userguide/weight/). The program Amos does not allow using weights when the FIML procedure is applied.

[4] Knoppen and Saris (2009) suggest another reason: The ESS value measurements do not possess discriminant validity (Campbell and Fiske, 1959). For further discussion, see the final section.

The multigroup confirmatory factor analysis (MGCFA) with 25 countries will enable us to test to what extent value measurements are invariant across countries in the data available from the third ESS round. The model used for the test is the same one that was confirmed for 20 countries in the first round and for 14 countries in the second round (Davidov et al., 2008a). This model included the seven values and five cross-loadings as reported in the previous section. The unified values in this model are universalism-benevolence, tradition-conformity and power-achievement. A test with the third round data will demonstrate how many countries may be compared with this model and whether all five cross-loadings are needed with the new data.

The multigroup analysis indicated that several countries required unifying one or more additional pairs of values because they were related to each other too strongly and could not be modeled separately. These countries did not provide support for the seven-value solution from previous rounds. To retain the seven-value model, those countries were eliminated from the analysis, resulting in a MGCFA with eight countries only.[5] This model was supported by the data as can be seen in the fit measures reported in first row of Table 4. The CFI value was higher than 0.9 and the RMSEA value was lower than 0.05. These fit measures were proposed by different authors to discern between models with a well versus poor fit to the data (Hu and Bentler, 1999; Marsh, Hau and Wen, 2004). In other words, the eight countries displayed configural invariance. Four of the cross-loadings were necessary in all countries. The cross-loading between the unified value conformity-tradition and the item ‘important to be rich’ was significant only in three countries (Denmark, Spain and Russia). However, an additional cross-loading was necessary between the construct self-direction and the item ‘important to be

[5] These countries are: Denmark, France, Norway, Portugal, Russia, Spain, Sweden and Switzerland.

modest' in five countries (France, Portugal, Russia, Spain, Sweden and Switzerland). These modifications are addressed in the summary and discussion part.

Next, I turn to the test of metric invariance. For this purpose, I constrained the factor loadings of the indicators to be equal across the eight countries. The global fit measures displayed in Table 4 supported the metric invariance test as well. The differences in the CFI and RMSEA fit measures between the configural and metric invariance models were below the recommended criteria (Chen, 2007). Thus, we can conclude that the samples display metric invariance.[6] The meaning associated with the values seems to be the same across the eight countries. In this model, only four out of the five cross-loadings were significant. The cross-loading between the unified values conformity-tradition and the item 'important to be rich' was insignificant. The additional cross-loading between the construct self-direction and the item 'important to be modest' was significant in all countries. The determination of metric invariance thus allows the comparison of the values' correlates among the eight countries that are analyzed here. I discuss some implications of these results from an applied point of view in the final section.

Finally, I performed the scalar invariance test. For this test, data are augmented with information about the mean level of the indicators (mean and covariance structure analysis – MACS, Sörbom, 1974, 1978). The intercepts of the indicators across the countries were constrained to be the same. This test resulted in an unacceptable global fit as can be seen in indicators reported in Table 4, suggesting that one should reject the scalar invariance model.

[6] In terms of change in chi-square, the model fit gets significantly worse for the metric invariance model. Thus, some constraints imposed to test the metric invariance of the model do not hold from this point of view. However, the chi-square is extremely sensitive to sample size and small to moderate deviations from normality. Therefore, it tends to reject a model with small discrepancies of no theoretical or practical relevance (Bentler and Bonett, 1980; Byrne and Stewart, 2006; West, Finch and Curran, 1995). Thus, other more pragmatic fit measures such as CFI and RMSEA have been proposed that do not share the disadvantages of the chi-square and may deliver contradictory conclusions to those of the chi-square (Chen, 2007; Hu and Bentler, 1999; Marsh et al., 2004; Byrne and Stewart, 2006). These global fit measures are also applied in this study.

Failure of the model to meet the scalar invariance test implies that the mean values in this data may not be compared across these countries. Several authors have suggested that when full invariance is not guaranteed, one may fall back to partial invariance. Partial invariance requires that only two items per construct possess measurement invariance characteristics (Byrne, Shavelson and Muthén, 1989; Steenkamp and Baumgartner, 1998). Releasing parameters and constraining the parameters of only two items per construct to be the same across countries did not result in any significant improvement of the model fit. To conclude, neither full nor partial scalar invariance were supported by the data.

However, in spite of this finding, it may be the case that a smaller set of countries or a smaller set of values meet the full or partial scalar invariance test. For example, researchers interested in a comparison of values between Great Britain and Ireland using the third round data of the ESS will find out that they display partial scalar invariance for seven values. This allows a latent mean comparison of the value scores across the two countries. It turns out that Irish people on average score lower on hedonism values and higher on self-direction, security and power and achievement values. There are no significant difference in the means of other values between Ireland and Great Britain.[7] Accordingly, to find out whether certain countries and values may be compared, researchers may follow similar steps and conduct tests of configural, metric and scalar invariance sequentially across the values and countries of interest for their specific research question.

As mentioned earlier, only 8 countries could be compared here because I tried to retain the model that was found in previous studies which included seven values (Davidov et al., 2008a). In several countries only six values could be identified. In these countries, another pair of adjacent values had to be unified, stimulation and self-direction, because they were too

[7] A model with seven values and five cross-loadings was applied for the comparison between Great Britain and Ireland.

strongly related to each other and could not be modeled separately. Utilizing this 6-value model allows the comparison of 21 countries.[8] The global fit measures of the configural, metric and scalar invariance tests of this model are presented in Table 4. Results show that the 21 countries display configural invariance. Furthermore, the differences in fit measures between the configural and partial metric invariance models are below the recommended criteria [9],. Thus, one can conclude that data from the 21 countries display partial metric invariance. However, the partial scalar invariance test had to be rejected once again. The reduction in the model fit was too large to allow the acceptance of the model (Chen, 2007; Hu and Bentler, 1999; Marsh et al., 2004).

Table 4 about here

c. Longitudinal invariance test

Finally I turned to the longitudinal invariance test. Here I examine whether values measured in the same country in 2002 (first round) and 2007 (third round) display metric and scalar invariance. Such invariance would allow comparing the values' correlates (covariances with other variables or unstandardized regression coefficients) and means over time, and thus exploring their determinants and consequences over time and studying aggregate value change. Seventeen countries participated in both the first and the third round of the ESS and allow the testing of whether they exhibit longitudinal invariance. Here the test starts with the same model explored in the cross-country test, which was confirmed in previous data with 7

[8] The countries are: Austria, Belgium, Bulgaria, Denmark, Estonia, Finland, France, Germany, Great Britain, Hungary, Ireland, Latvia, Netherlands, Norway, Poland, Russia, Slovenia, Spain, Sweden, Switzerland and Ukraine.

[9] In this model two factor loadings per construct were constrained to be equal across countries. This model suggested, once again, that the cross-loading between the unified value conformity-tradition and the item 'important to be rich' is not significant in most countries.

values and 5 cross-loadings. After reviewing the findings I will consider whether some further modifications are needed in order to achieve invariance.

A total of 17 MGCFA analyses were conducted, one for each country, in which configural, metric and scalar invariance over time were tested for each of the countries separately. In six countries no modifications were needed and data supported full scalar invariance over time for the values. In other countries some adjustments were necessary. Some countries required unifying one or more pairs of values because they were related to each other too strongly and could not be modeled separately. In other words, in these countries, the seven-value model could not be retained. Other modifications included one or more additional cross-loadings or releasing error correlations. After these modifications, the global fit measures suggested that these countries also displayed scalar invariance. Table 5 reports the global fit measures and the necessary modifications in each country. Now country value means may be compared over time in these countries.[10]

Table 5 about here

Sörbom (1974) has shown that to compare means of latent variables they should be constrained to zero in one reference group. As a result one is able to estimate mean differences. Table 6 provides the mean differences over time in each country (empty cells represent no significant change). As one can see, there are 55 significant changes between Round 1 and Round 3. Several changes are medium-sized. Four temporal changes in three countries are larger than 0.2 and 11 changes are higher than 0.15. Only 29 changes are higher than 0.1 (please remember that the values are measured on a 6-point scale).

[10] The statistical necessity to add these modifications and possible resulting changes in the substantive meaning of the values are discussed in the final section.

Such medium to small changes are not surprising, as one does not expect large aggregate value changes over a 5-year period at the country level but rather in the longer run (Hofstede, 2001; Inglehart, 1990; Rokeach, 1979; Schwartz, Bardi and Bianchi 2000; Schwartz, 2006a; Williams, 1979; for a more general discussion see Barber and Inkeles, 1971). In fact, some researchers argue that certain cultural elements need hundreds of years until they change (Schwartz, 2006a; Kohn and Schooler, 1983; Putnam, 1993). Values are more general than attitudes, opinions or norms and, therefore, their change over time takes longer. Especially values which are not related to the emergence or alleviation of major societal problems are expected to remain stable (Rokeach 1979). Societal adaptation to technological developments, increasing gross national product, national and individual wealth, exchange with foreign cultures, media, or other factors may bring about slow and gradual value change (Inglehart and Baker, 2000; Schwartz, 2006a).

Studying whether observed *changes* are meaningful could be accomplished by investigating their relations to other theoretical constructs of interest. Meaningful value change could predict dynamics and variation in other phenomena such as attitudes toward certain groups in society, racism, nationalism, political orientation or voting behavior. Panel data could allow, in addition, studying the individual change and not only the societal (aggregate) one, and whether this change could be linked to other individual characteristics. These analyses are beyond the scope of this paper. However, the observed societal changes in the current study suggest that in spite of a rather high stability, small transitions may be observed across a five-year period. Inspection of substantial value changes requires a longer period of time than the time span between the first and third rounds that the ESS provides. Findings of temporal scalar invariance allow us to interpret these changes within countries meaningfully.

Table 6 about here

## 6. Summary and discussion

In the last decade we have observed a substantial increase in the number of studies that focus on the measurement and application of values both theoretically and empirically. Many of these studies use values in a cross-cultural and longitudinal framework. However, value questions may be understood differently across countries. Furthermore, their meaning might change over time. To guarantee that they are nevertheless comparable both across countries and over time, invariance should be tested.

In this study I explained why testing for invariance is necessary before comparisons of values or other theoretical constructs of interests are done; I introduced how invariance may be tested across nations and time points and presented a practical application to the human values measurements of the European Social Survey. I tested the cross-country and temporal invariance of Schwartz' (1992) human values measurements using data from the third round of the ESS. Furthermore, the compatibility of data from the third ESS round to measure 10 values was assessed. Finally, problems such as necessity to unify values or introducing cross-loadings were presented.

Previous studies have demonstrated that only five to eight values can be distinguished with data from the first and second ESS rounds (Davidov, 2008; Davidov et al, 2008a). These studies have also shown that metric invariance of the values may be guaranteed across all or a subset of the countries. Scalar invariance could not be guaranteed across all countries, but it was shown that it may be reached across small subsets of countries and/or values. Scalar invariance was evidenced between the first and the second ESS rounds in all countries.

This study of data from the third ESS round provided complementary results. In the single country analyses, between four and seven values could be distinguished in each country. Adjacent pairs of values had to be unified because they correlated too strongly and could not be modeled separately. This finding might have been a result of the fact that there are only 21 indicators available in the ESS to measure the 10 theoretically postulated values. Previous studies which used 40 items could identify all the 10 values (see, e.g., Schwartz and Boehnke, 2004). Theoretically speaking, the unified values belong in most cases to the same underlying higher-order dimension: The unified values universalism-benevolence belong to self-transcendence, power and achievement to self-enhancement, tradition and conformity to conservation, and stimulation and self-direction to openness to change. Thus, the findings imply that the ESS value questions allow measuring these higher order dimensions better than the single values, and are therefore useful for investigating research questions related to the higher-order dimensions in the theory.

Metric invariance was established for eight countries and seven values in the third ESS round. These eight countries differ in language and culture, and it is difficult to indicate in what way certain similarities across these countries resulted in metric invariance. Statistically, the results indicate that people in these countries seem to understand the value questions in a similar way because the loadings of the questions on the values are similar. The statistical test conducted in this study, however, is a necessary but not a sufficient condition to guarantee that people understand the questions in a similar way. Cognitive tests may offer a supplementary tool to assess the equivalence of meaning of the values instrument in the different countries and may provide us with additional means to study how the statistical findings are related to the cultural specificities of the countries.

After decreasing the number of values in the model to six, data provided support for partial metric invariance across 21 countries. These findings allow researchers to study how values affect attitudes, opinions and behavior in different countries and to *compare these (unstandardized) effects* meaningfully. Furthermore, the effects of sociodemographic and socioeconomic variables on values may be investigated and compared across countries in a meaningful way.[11]

Scalar invariance was not established across countries.[12] I was able to demonstrate, however, that scalar invariance may be reached across subsets of countries or values. Scalar invariance was also established over time for all countries. This is important because it facilitates studying value transitions within countries meaningfully.

During the analysis it was necessary to add some cross-loadings. Indicators that were originally supposed to measure a certain value had an additional secondary loading on a different, often opposing, value. In the longitudinal analysis it was also necessary to add some error correlations in a few countries. Adding cross-loadings or error correlations brings into question the new meaning of these values. After value indicators are linked to other values by cross-loadings or error correlations the meanings of the values change. One may argue that it is not clear what the new meaning for each country actually is. From a methodological point of view these modifications indicate that convergent and discriminant validity (Campbell and Fiske, 1959) are not always present since some items are related to other items or to other values also directly and not solely via their latent variable. In part, this is a result of the

[11] Metric invariance is also a precondition for using values as predictors in multilevel analyses. Once metric invariance is established, effects of values become comparable. This implies that they may be used as predictors on the micro level. However, standardized effects may still not be compared. To allow comparisons of correlations or standardized regression coefficients, it is necessary to guarantee that also the variances of the values and their predictors or consequences are equal across groups.

[12] This implies that the aggregate value *means* are not comparable across countries. Therefore, one cannot use values as contextual (macro) variables across countries in a multilevel analysis, since such an application requires comparable means across countries.

unification of certain values; indicators originally *intended* to measure a specific value require introducing a correlation between their measurement errors when collapsed with other indicators to measure a single concept. Failing to consider these error correlations or cross-loadings might lead to the rejection of the models and to distorted estimates of model parameters with overestimated factor correlations and distorted structural relations (Marsh et al., 2009). Therefore, in a cross-cultural setting, it is recommended to look for those modifications which are necessary across all or most units of analysis (nations, cultures, time point) and account for them. From a substantive point of view, all the cross-loadings introduced include paths between single indicators and motivationally opposed values that were formed by combining two latent values. Negative cross-loadings indicate that the association between the opposing latent values did not capture all of the opposition for these single indicators. The positive cross-loadings indicate that these associations overestimated the opposition for these single indicators. Furthermore, although significant, the cross-loadings were much weaker than the loadings of the values' original indicators. So the meaning of the values remained at least, for the most part, unchanged. In sum, as Marsh, Hau and Grayson (2005) have argued, apparently almost no multidimensional instrument in practice provides a good fit without some modifications. Obviously, further research is needed regarding *the extent* that the meaning of the values remains unchanged.

These results make it obvious that metric and scalar invariance may not be assumed across countries and time points. This underlines the importance of testing invariance before beginning any further substantive work. Skipping this step and simply assuming invariance of theoretical constructs across countries or over time in comparative studies might lead to severely biased results, as several authors have demonstrated (Billiet, 2003; De Beuckelaer, 2005; Steenkamp and Baumgartner, 1998). Only if invariance is established can researchers confidently carry on their comparative analysis and interpret their results in a meaningful

way.[13] Considering the central role values play in sociological comparative studies, it becomes evident that guaranteeing invariance is indispensable for conducting *meaningful* cross-national and longitudinal comparisons. I hope that the current study is of assistance to researchers interested in conducting comparisons of values or other theoretical constructs across cultures, nations or time to reach this goal.

In this study I focused on countries as the units of analysis. This strategy could be criticized because countries may be heterogeneous and language or cultural groups may also be important units of analysis, especially in value research. Furthermore, countries do not only represent a cultural or a linguistic frame but also other aspects which are more related to the data collection, such as field agencies, national coordinators and the specific type of data collection in the country. Indeed, the study of values may not necessarily be conducted by using the nation as the unit of analysis. I did not make an implicit assumption that nations and cultures are to some extent equivalent. Different nations may include various social groups with a similar underlying culture. By contrast, a single nation often hosts various social groups which are culturally very different from each other. So even when we find similarities across nations, some cultural aspects may vary considerably within a group. However, the study of nations is a useful way to investigate differences and similarities. After all, nations represent societies with unique laws, governments, economic institutions or social norms – all elements of culture that are shaped by underlying common values. From a practical point of view, countries also constitute very important units of analysis for many substantive questions in which, for example, political attitudes and opinions, voting behavior or support of social policies are investigated (De Beuckelaer and Lievens, in press).

[13]Saris and Gallhofer (2007) suggest that the invariance test in this study is too restrictive and provide an alternative test using ESS data (for applications see, e.g., Knoppen and Saris, 2007). However, there is not enough data available in the ESS to perform their proposal for all countries, but only for a selective number of countries. In these countries, split-ballot MTMM experiments were included in the ESS to assess cognitive invariance. For details see Saris and Gallhofer (2007, chapter 16).

The results presented here provide good news for these types of studies because substantial levels of invariance across countries and over time are established. Future investigations may consider assessing invariance of values across language or cultural groups as well. Finally, from a theoretical and practical point of view it is recommended to add more items to the values scale in the ESS, since the 21-item scale does not distinguish between all the 10 theoretically postulated values as the original 40-item scale has done in the past.

References

Allport, G.W., Vernon, P.E. and Lindzey, G.A. (1960) *A Study of Values.* Boston: Houghton Mifflin.

Arbuckle, J. L. ( 2005) *Amos 6.0 User's Guide*. Chicago, IL: SPSS.

Barber, B. and Inkeles, A. (1971) *Stability and Social Change.* Boston: Little, Brown.

Bentler, P. M. and Bonett, D.G. (1980) 'Significance Tests and Goodness-of-fit in the Analysis of Covariance Structures', *Psychological Bulletin, 88*: 588–606.

Billiet, J. (2003) Cross-Cultural Equivalence with Structural Equation Modeling, in J.A. Harkness, F.J.R. Van de Vijver, and P.P. Mohler (eds), *Cross-Cultural Survey Methods,* pp. 247-264. New York, NY: John Wiley.

Bollen, K.A. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

Byrne, B.M. (2001) *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*. Mahwah, NJ: Erlbaum.

Byrne, B.M., Shavelson, R.J. and Muthén, 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. and Stewart, S.M. (2006) 'The MACS Approach to Testing for Multigroup Invarianceof a Second-Order Structure: A Walk Through the Process', *Structural Equation Modeling 13*: 287-321.

Campbell, D.T and Fiske, D.W. (1959) 'Convergent and Discriminant Validation by the Multitrait Multimethod Matrices', *Psychological Bulletin 56*: 81-105.

Chen, F.F. (2007) 'Sensitivity of Goodness of Fit Indices to Lack of Measurement Invariance', *Structural Equation Modeling 14*: 464-504.

Cheung, G.W. and Rensvold, RB. (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. and Rensvold, R.B. (2002) 'Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance', *Structural Equation Modeling 9:* 233-255.

Coenders, G. and Saris, W. E. (1995). Categorization and Measurement Quality. The Choice Between Pearson and Polychoric Correlations. In W. E. Saris and Á. Münnich (eds.), *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, pp. 125-144. Budapest: Eötvös University Press.

Coenders, G., Satorra, A. and Saris, W. E. (1997) 'Alternative Approaches to Structural Modeling of Ordinal Data. A Monte Carlo Study', *Structural Equation Modeling, 4:* 261-282.

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(1)*: 33-46.

Davidov, E., Schmidt, P. and Schwartz, S. (2008a) 'Bringing Values Back In: Testing the Adequacy of the European Social Survey to Measure Values in 20 countries', *Public Opinion Quarterly 72:* 420-445.

Davidov, E., Meuleman, B., Billiet, J. and Schmidt, P. (2008b) 'Values and Support for Immigration. A Cross-Country Comparison', *European Sociological Review 24:* 583-599.

De Beuckelaer, A. (2005) *Measurement Invariance Issues in International Management Research.* Unpublished doctoral dissertation, Limburgs Universitair Centrum, Limburg, the Netherlands.

De Beuckelaer, A. and Lievens, F. (in press) 'Measurement Equivalence of Paper-and-Pencil and Internet Organisational Surveys: A Large-Scale Examination in 16 Countries', *Applied Psychology*.

Feldman, S. (2003). 'Values, Ideology, and Structure of Political Attitudes', in D.O. Sears, L. Huddy, and R. Jervis (eds) *Oxford Handbook of Political Psychology,* pp. 477-508. New York: Oxford University Press.

Halman, L. and de Moor, R. (1994) 'Value Shift in Western Societies', in P. Ester, L. Halman, and R. de Moor (eds), *The Individualizing Society: Value Change in Europe and North America,* pp. 1-20. Tilburg: Tilburg University Press.

Harkness, J.A.,. Van de Vijver, F.J.R and Mohler, P.P. (eds). (2003) *Cross-Cultural Survey Methods*. New York: John Wiley.

Hitlin, S. and Piliavin, J.A. (2004) 'Values: Reviving a Dormant Concept', *Annual Review of Sociology 30*: 359-393.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values.* London: Sage.

Hofstede, G. (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations (2nd ed.)*. Beverly Hills CA: Sage.

Horn, J.L., and McArdle, J.J. (1992) 'A Practical and Theoretical Guide to Measurement Invariance in Aging Research', *Experimental Aging Research 18:* 117-144.

Hu, L. and 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. and 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. (1977) *The Silent Revolution: Changing Values and Political Styles in Advanced Industrial Society.* Princeton, NJ: Princeton University Press.

Inglehart, R. (1990) *Culture Shift in Advanced Industrial Society*. Princeton, NJ: Princeton University Press.

Inglehart, R. and Baker, W.E. (2000) ‘Modernization, Cultural Change, and the Persistance of Traditional Values’, *American Sociological Review 65:* 19-51.

Inglehart, R., Basánez, M., Die´z-Medrano, J., Halman, L. and Luijkx, R. (2004) *Human Beliefs and Values. A Cross-Cultural Sourcebook Based on the 1999-2002 Value Surveys.* Buenos Aires: Siglo xxi editors.

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, and E. Shils, (eds) *Toward a General Theory of Action*. Cambridge, MA: Harvard University Press.

Knoppen, D. and Saris, W. (2007) *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, June, Prague, Czech Republic.

Knoppen, D. and Saris, W. (2009). ‘Do we have to Combine Values in the Schwartz´ Human Values Scale? A Comment on the Davidov Studies’, *Survey Research Methods 3:* 91-103.

Kohn, M. L. and Schooler, C. (1983). *Work and personality.* Norwood, NJ: Ablex.

Lubke, G.H. and 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. and Grayson, D. (2005) ‘Goodness of Fit Evaluation in Structural Equation Modeling’. in A. Maydeu-Olivares and J. McCardle (eds) *Psychometrics: A Festschrift to Roderick P. McDonald*, pp. 275–340. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H.W., Hau, K.-T. and 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.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S. and Trautwein, U. (2009) 'Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching', *Structural Equation Modeling 16*: 439-476.

Meredith,W. (1993) 'Measurement Invariance, Factor Analysis and Factorial Invariance', *Psychometrika 58*: 525-543.

Putnam, R.D. (1993) *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, NJ: Princeton University Press.

Rokeach, M. (1973) *The Nature of Human Values.* New York: Free Press.

Rokeach, M. (1979) 'Change and Stability in American Value Systems, 1968-1971' in Rokeach M. (ed), *Understanding Human Values. Individual and Societal,* pp. 15-46. New York: The Free Press.

Rokeach, M. and Ball-Rokeach, S. (1989) 'Stability and Change in American Value Priorities, 1968-1981', *American Psychologist 44:* 775-784.

Sagiv, L. and 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. and.Gallhofer, I.N. (2007) *Design, Evaluation and Analysis of Questionnaires for Survey Research.* New York: Wiley.

Schafer, J.L. and Graham, J.W. (2002) 'Missing Data: Our View of the State of the Art.', *Psychological Methods 7*: 147-177.

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. (2003) 'A Proposal for Measuring Value Orientations across Nations'. *ESS Questionnaire Development Report* (Chapter 7). http://www.europeansocialsurvey.org

Schwartz, S.H. (2005a) 'Robustness and Fruitfulness of a Theory of Universals in Individual Human Values', in A. Tamayo and J.B. Porto (eds), *Valores e Comportamento nas Organizações* [Values and Behavior in Organizations], pp. 56-95. Petrópolis, Brazil: Vozes.

Schwartz, S.H. (2005b) 'Basic Human Values: Their Content and Structure across Countries', in A. Tamayo and J. B. Porto (eds), *Valores e Comportamento nas Organizações* [Values and Behavior in Organizations], pp. 21-55. Petrópolis, Brazil: Vozes.

Schwartz, S.H. (2006a) 'A Theory of Cultural Value Orientations: Explication and Applications' *Comparative Sociology 5:* 137-182.

Schwartz, S.H. (2006b) 'Basic Human Values: Theory, Measurement, and Applications.' *Revue Française de Sociologie 47*: 249-288.

Schwartz, S.H. (2007) 'Value Orientations: Measurement, Antecedents and Consequences across Nations', in R. Jowell, C. Roberts and R. Fitzgerald (eds), *Measuring Attitudes Cross-Nationally - Lessons from the European Social Survey*, pp. 161-193. London: Sage.

Schwartz, S. H., Bardi, A. and Bianchi, G. (2000) 'Value Adaptation to the Imposition and Collapse of Communist Regimes in Eastern Europe', in S. A. Renshon, and J. Duckitt (eds), *Political Psychology: Cultural and Cross Cultural Perspectives*, pp. 217-237. London: Macmillan.

Schwartz, S.H. and Boehnke, K. (2004) 'Evaluating the Structure of Human Values with Confirmatory Factor Analysis', *Journal of Research in Personality 38*: 230-255.

Schwartz, S.H. and Huismans, S. (1995) 'Value Priorities and Religiosity in Four Western Religions', *Social Psychology Quarterly 58:* 88-107.

Schwartz, S.H., Melech, G., Lehmann A., Burgess, S., Harris, M. and 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.

Sörbom, D. (1978) 'An Alternative to the Methodology for Analysis of Covariance', *Psychometrika 43*: 381-396.

Steenkamp, J.-B.E.M. and Baumgartner, H. (1998) 'Assessing Measurement Invariance in Cross-National Consumer Research', *Journal of Consumer Research 25*: 78-90.

Triandis, H.C. (1993) 'Collectivism and Individualism as Cultural Syndromes', *Cross-Cultural Research*, 27: 155-180.

Triandis, H. C. (1998) 'Vertical and Horizontal Individualism and Collectivism: Theory and Research Implications for International Comparative Management', *Advances in International Comparative Management*, 12: 7-35.

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. and 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.

Weber, M. (1958) *The Protestant Ethic and the Spirit of Capitalism.* New York: Scribners.

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. and Billiet, J. (2002) ‘A Comparison of Techniques for Detecting Cross-Cultural Inequivalence at the Item Level’, *Quality & Quantity 36:* 197-218.

West, S.G., Finch, J.F. and Curran, P.J. (1995) ‘Structural Equation Models with Nonnormal Variables: Problems and Remedies’, in R. H. Hoyle (ed), *Structural Equation Modeling: Concepts, Issues, and Applications,* pp. 56–75. Thousand Oaks, CA: Sage.

Williams, R..M.,J. (1968) ‘Values’, in E. Sills (ed, *International Encyclopedia of the Social Sciences*. New York: Macmillan

Williams, R.M.J. (1979) ‘Change and Stability in Values and Value Systems: A Sociological Perspective’ in M. Rokeach (ed), *Understanding Human Values. Individual and Societal*, pp. 15-46. New York: The Free Press.

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

| Value Type | Core Goal |
|---|---|
| 1. Power | Social status and prestige, control or dominance over people and resources |
| 2. Achievement | Personal success through demonstrating competence according to social standards |
| 3. Hedonism | Pleasure and sensuous gratification for oneself |
| 4. Stimulation | Excitement, novelty, and challenge in life |
| 5. Self-Direction | Independent thought and action - choosing, creating, exploring |
| 6. Universalism | Understanding, appreciation, tolerance and protection for the welfare of all people and for nature |
| 7. Benevolence | Preservation and enhancement of the welfare of people with whom one is in frequent personal contact |
| 8. Tradition | Respect, commitment and acceptance of the customs and ideas that traditional culture or religion provide the self |
| 9. Conformity | Restraint of actions, inclinations and impulses likely to upset or harm others and violate social expectations or norms |
| 10. Security | Safety, harmony and stability of society, of relationships and of self |

Note. Adopted from Sagiv and Schwartz, 1995.

*Table 2: The ESS Human Values Scale in the 3rd 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 (ipcrtiv).<br>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).<br>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).<br>19. He strongly believes that people should care for nature. Looking after the environment is important to him (impenv). |
| Benevolence (BE) | 12. It is very important to him to help the people around him. He wants to care for their well-being (iphlppl).<br>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).<br>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).<br>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).<br>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).<br>17. It is important to him to get respect from others. He wants people to do what he says (iprspot). |
| Achievement (AC) | 4. It is important to him to show his abilities. He wants people to admire what he does (ipshabt).<br>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).<br>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).<br>15. He looks for adventures and likes to take risks. He wants to have an exciting life (ipadvnt). |

Note. Adapted from Davidov, 2008.

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

| Country | Number of Values, ESS Round 3 | Unified Values[C] |
|---|---|---|
| 1. Austria | 7 | UNBE, POAC, COTR |
| 2. Belgium | 6 | UNBE, COTR, POAC, STSD |
| 3. Bulgaria* | 6 | UNBE, COTR, POAC, STSD |
| 4. Cyprus* | 4 | UNBECOTR, POAC, HESTSD |
| 5. Denmark | 7 | UNBE, COTR, POAC |
| 6. Estonia* | 4 | UNBE, SECCOTR, POAC, HESTSD |
| 7. Germany | 6 | UNBE, COTR, POAC, STSD |
| 8. Finland | 6 | UNBE, COTR, POAC, STSD |
| 9. France | 7 | UNBE, COTR, POAC |
| 10. Great Britain | 7 | UNBE, COTR, POAC |
| 11. Hungary | 6 | UNBE, COTR, POAC, STSD |
| 12. Ireland | 7 | UNBE, POAC, COTR |
| 13. Latvia* | 6 | UNBE, POAC, COTR, STSD |
| 14. Netherlands | 6 | UNBE, POAC, COTR |
| 15. Norway | 7 | UNBE, COTR, POAC |
| 16. Poland | 7 | UNBE, COTR, POAC |
| 17. Portugal | 7 | UNBE, COTR, POAC |
| 18. Romania* | 6 | UNBE, POAC, COTR, STSDHE |
| 19. Russia* | 7 | UNBE, COTR, POAC |
| 20. Slovakia* | 4 | SECUNBECOTR, POAC, STSD |
| 21. Slovenia | 6 | UNBE, COTR, POAC, STSD |
| 22. Spain | 7 | UNBE, COTR, POAC |
| 23. Sweden | 7 | UNBE, COTR, POAC |
| 24. Switzerland | 7 | UNBE, COTR, POAC |
| 25. Ukraine* | 6 | UNBE, POAC, COTR, STSD |

[A] Based on data from the third round of the ESS 2006/2007.

[B] Only 17 countries collected data in Round 1 and Round 3. The marked (*) countries did not collect value data in Round 1.

[C] For a full description of the abbreviations of values, see Table 2.

*<u>Table 4:</u> Fit Measures of a Multigroup Confirmatory Factor Analysis of Seven Values, Constraining Configural Invariance, Metric, and Scalar Invariance*

| Model type | CFI | RMSEA | PCLOSE | AIC | BCC | Chi-Square | df |
|---|---|---|---|---|---|---|---|
| A model with 7 values (Davidov et al, 2008a) and 8 countries | | | | | | | |
| 1. Configural invariance | .918 | .018 | 1.00 | 9,573 | 9,589 | 8,133 | 1,296 |
| 2. Metric invariance | .907 | .019 | 1.00 | 10,351 | 10,364 | 9,177 | 1,429 |
| 3. Scalar invariance | .805 | .026 | 1.00 | 18,781 | 18,792 | 17,819 | 1,535 |
| A model with 6 values and 21 countries | | | | | | | |
| 4. Configural invariance | 0.903 | 0.012 | 1.00 | 26,191 | 26,226 | 23,545 | 3,528 |
| 5. (Partial) Metric invariance | 0.899 | 0.012 | 1.00 | 26,848 | 26,879 | 24,482 | 3,668 |
| 6. Scalar invariance | 0.859 | 0.014 | 1.00 | 35,840 | 35,879 | 32,832 | 3,788 |

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; 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).

*Table 5: Metric and Scalar Invariance of a Multiple-Group Comparison of Round 1 and Round 3 in each Country, Global Fit Measures* [A]

| | | | Global fit measures: metric invariance test | | | Global fit measures: scalar invariance test | | |
|---|---|---|---|---|---|---|---|---|
| Country | Values included in the model | Modifications needed | Pclose | RMSEA | CFI | Pclose | RMSEA | CFI |
| 1. Austria | | | 1.00 | 0.039 | 0.921 | 1.00 | 0.040 | 0.911 |
| 2. Belgium | UNBE, POAC, COTR, SEC, HE, STSD | Unifying ST, SD | 1.00 | 0.032 | 0.908 | 1.00 | 0.032 | 0.904 |
| 3. Denmark | | | 1.00 | 0.032 | 0.918 | 1.00 | 0.032 | 0.911 |
| 4. Germany | UNBE, POAC, COTR, SEC, HE, STSD | Unifying ST, SD; UNBE→ipfree | 1.00 | 0.037 | 0.902 | 1.00 | 0.037 | 0.899 |
| 5. Finland | UNBE, POAC, COTR, SEC, HE, STSD | Unifying ST, SD | 1.00 | 0.035 | 0.914 | 1.00 | 0.035 | 0.912 |
| 6. France | | UNBE→ipgdtim, ipbhprp; ST→ipmodst; error(ipudrst) <->error(ipeqopt) | 1.00 | 0.032 | 0.922 | 1.00 | 0.035 | 0.900 |
| 7. Great Britain | | error(ipudrst) <->error(ipeqopt); error(ipadvnt) <->error(impsafe) | 1.00 | 0.035 | 0.914 | 1.00 | 0.036 | 0.905 |
| 8. Hungary | UNBE, POAC, COTR, SEC, HESTSD | Unifying ST, SD; unifying POAC, HE; SEC→impfree; UNBE→impfun; error(ipfrule) <->error(ipudrst); error(ipgdtim) <->error(impfree); error(ipgdtim) <->error(impfun) | 1.00 | 0.037 | 0.903 | 1.00 | 0.038 | 0.893 |
| 9. Ireland | | | 1.00 | 0.040 | 0.912 | 1.00 | 0.039 | 0.910 |
| 10. Netherlands | UNBE, POAC, COTR, SEC, HE, STSD | Unifying ST, SD | 1.00 | 0.035 | 0.911 | 1.00 | 0.035 | 0.906 |
| 11. Norway | | | 1.00 | 0.035 | 0.917 | 1.00 | 0.035 | 0.914 |
| 12. Poland | | | 1.00 | 0.034 | 0.924 | 1.00 | 0.035 | 0.919 |
| 13. Portugal | | UNBE→ipmodst | 1.00 | 0.042 | 0.918 | 1.00 | 0.042 | 0.915 |
| 14. Slovenia | UNBE, POAC, COTR, SEC, HESTSD | Unifying HE, ST, SD; UNBE→ipmodst, impfree; COTR→impfree; error(ipgdtim) <->error(impdiff) | 1.00 | 0.037 | 0.902 | 1.00 | 0.043 | 0.858 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15. Spain | | | 1.00 | 0.036 | 0.930 | 1.00 | 0.037 | 0.924 |
| 16. Sweden | | UNBE→ipmodst; ST→ipstrgv; SD→impsafe | 1.00 | 0.035 | 0.907 | 1.00 | 0.035 | 0.898 |
| 17. Switzerland | | UNBE→ipmodst; | 1.00 | 0.032 | 0.909 | 1.00 | 0.033 | 0.903 |

[A] For 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 Davidov et al., (2008a) 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 longitudinal metric and scalar invariance tests between round 1 and round 3.

→ Signifies that a modification requires releasing the equality constraint on the corresponding factor loading; <-> signifies that a modification requires estimating the covariance.

*Table 6: Aggregate Change in Value Means from Round 1 to Round 3*[A].

| Country | HE | ST | SD | UNBE | COTR | SEC | POAC | HESTSD | STSD |
|---|---|---|---|---|---|---|---|---|---|
| 1. Austria | .240** | | | -.075** | | | | | |
| 2. Belgium | | | | | .071* | | .048* | | |
| 3. Denmark | -.111** | -.092* | -.161** | | .132** | -.098* | | | |
| 4. Germany | .078** | | | | .079** | -.132** | .098** | | |
| 5. Finland | | | | | | -.137** | | | |
| 6. France | | -.162** | -.303** | | -.119** | -.263** | -.166** | | |
| 7. Great Britain | .145** | | -.074* | .107** | .074* | | | | |
| 8. Hungary | | | | .055** | | .079** | | | |
| 9. Ireland | | | | | | -.085* | | | |
| 10. Netherlands | .091** | | | .036* | | -.118** | .075** | | .153** |
| 11. Norway | | .140** | | .106** | | .113** | .073** | | |
| 12. Poland | | | | .035* | | -.120** | .076* | | |
| 13. Portugal | -.174** | | -.136** | | .128** | | | | |
| 14. Slovenia | | | | .065** | .081** | -.072* | | .131** | |
| 15. Spain | | | | .131** | | | -.202** | | |
| 16. Sweden | .126** | .091* | | .042* | | -.153** | | | |
| 17. Switzerland | | | -.104** | .059** | .155** | | .047* | | |

[A] For a full description of the abbreviations of values, see Table 2. Only significant changes are reported. * P < 0.05; ** P < 0.01

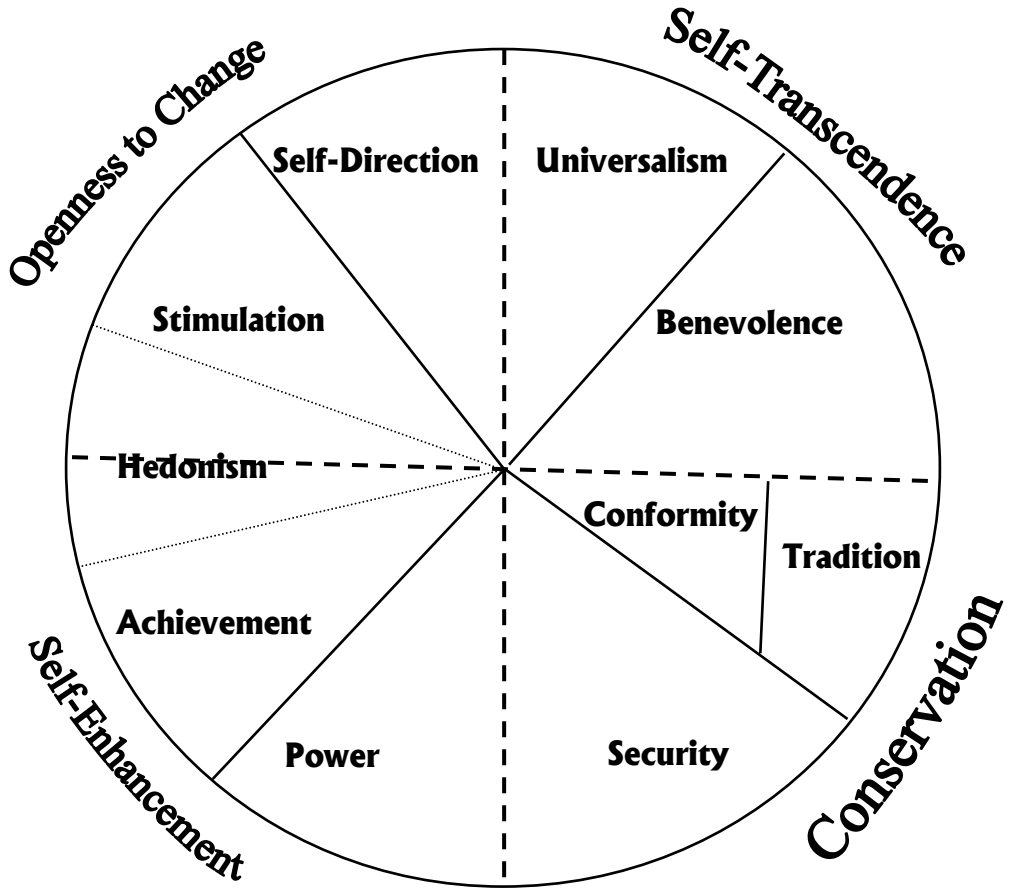


*Figure 1: Structural relations among the 10 values and the two dimensions*