Survey Research Methods (2012)
Vol.6, No.1, pp. 37-48
ISSN 1864-3361
http://www.surveymethods.org

A comparison of the invariance properties of the PVQ-40 and the PVQ-21 to measure human values across German and Polish Samples

Jan Cieciuch
University of Finance and Management, Warsaw

Eldad Davidov
University of Zurich, Switzerland

The Schwartz theory of basic human values has promoted a revival of empirical research on values in cross-cultural, social, and personality psychology. A central topic in the application of values is comparing them across cultures. This study examines the measurement invariance of the full 40-item version of the Portrait Values Questionnaire (PVQ) as compared to the 21-item version as applied in the European Social Survey (ESS). The research was carried out on a total sample of 1,204 individuals in Poland and Germany. Measurement invariance was tested using two-group confirmatory factor analyses. Analyses revealed that all the values measured by the PVQ-40 displayed not only configural and metric but also partial scalar invariance for all 10 values with the exception of the stimulation value, thus allowing the mean comparison of nine out of the 10 postulated values. However, the 21-item version achieved partial scalar only for seven values. We conclude that the PVQ-40 is more appropriate for cross-cultural research.
Keywords: human values, configural, metric, and scalar invariance, PVQ-40, PVQ-21

1 The Human Values Theory

The proposal of Schwartz (1992, 2006) is currently one of the best theoretically and empirically grounded theories of values. The theory has received empirical support in a large number of cross-cultural studies (e.g., Bilsky, Janik and Schwartz 2011; Fontaine et al. 2008; Schwartz 2006; Schwartz and Rubel-Lifschitz 2009) using data collected in different countries from around the globe.

Schwartz (1992) defines values as desirable, transsituational goals, which vary in importance and serve as guiding principles in people's lives. He differentiates between the following 10 values (followed by the motivation they express): 1) *conformity*: restraint of action, inclinations, and impulses likely to upset or harm others and violate social expectations or norms, obedience, self-discipline, respect for elders; 2) *tradition*: the acceptance of the customs and ideas that one's culture or religion imposes on the individual; 3) *benevolence*: caring for the welfare of people with whom one is in frequent personal contact, friendship, love; 4) *universalism*: the protection of the welfare of all people as well as nature and environmental protection, justice, wisdom, peace; 5) *self-direction*: independence in thought and action, creativity, freedom, autonomous selection of one's own goals; 6) *stimulation*: novelty seeking, excitement and variety in life; 7) *hedonism*: striving to achieve pleasure, seeking to satisfy one's own needs; 8) *achievement*: seeking personal success through demonstrated competence in accordance with social standards; 9) *power*: striving to attain social status and prestige, control and dominance over other people and resources; and finally 10) security: harmony, social order, personal, family and national safety.

Contact information: Jan Cieciuch, University of Finance and Management, Warsaw, Poland, e-mail: jancieciuch@gmail.com

An important proposition of Schwartz concerns the structure of the abovementioned values. Theoretical deliberations supported by the results of empirical research have led Schwartz to the hypothesis that his values are organized around a circle (Schwartz 1992, 2006). Values located close to each other on the circle are jointly preferred whereas values on opposite sides of the circle exclude each other. Thus, for instance, the universalism and benevolence values can be jointly realized but are in conflict with realizing power and achievement values. Values located next to each other on the circle are expected to correlate positively whereas values located on opposite sides of the circle are expected to correlate negatively or not at all (Schwartz 1992, 2006; Vecchione, Casconi and Barbaranelli 2009; see also the study of Steinmetz et al. in this volume).

Finally, Schwartz proposed that the values can be described as being organized along two bipolar dimensions. The first, self-transcendence, which includes the values universalism and benevolence, opposes self-enhancement, which includes the values power and achievement. The second, openness to change, which includes the values stimulation and selfdirection, opposes conservation, which includes the values tradition, conformity, and security. Hedonism is located between openness to change and self-enhancement.

The theoretical proposal of Schwartz constitutes an excellent foundation for performing cross-cultural comparisons. From the very conception of this theory, Schwartz himself has carried out cross-cultural research (Schwartz 1992, 2005, 2006). One of the scales that Schwartz developed to measure his values is the 40-item Portrait Values Questionnaire (PVQ: Schwartz et al. 2001). A shortened version (21 items) of Schwartz's measurement instrument has been included in the semiannual European Social Survey (ESS) since 2002. A central topic in the application of values data as measured by the PVQ in comparative studies

is the comparability of the value items across nations. After all, measurement invariance of the values is required before cross-cultural comparisons may be carried out meaningfully (Billiet 2003; De Beuckelaer 2005; Steenkamp and Baumgartner 1998; Vandenberg 2002). If a sufficient level of measurement invariance is not established, the results obtained in different cultures or countries cannot be compared in a meaningful way (Byrne at al. 2009; Chen 2008).

2 Measurement Invariance

If we compare scores of values across different groups, we make the important and critical assumption that the scale measures the same construct in all groups. It is not enough to assume this – the instrument must be tested statistically for its equivalence across the studied groups. This is particularly important in cross-cultural or cross-country research (Chen 2008; Steenkamp and Baumgartner 1998; van de Vijver and Poortinga 1997; Vandenberg and Lance 2000).

The definition of measurement invariance is "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn and McArdle 1992:117). If invariance is not established, interpretations of comparisons between groups are problematic (Byrne et al. 2009; Chen 2008). It is possible that differences which are found between groups in means or regression coefficients do not correspond with 'real' differences or, in another situation – that the real differences are obscured because of noninvariant measurements.

There are some statistical possibilities for testing measurement invariance. Of these, the most widely used method is multigroup confirmatory factor analysis (MGCFA; Jöreskog 1971; see also Bollen 1989). This method facilitates testing for invariance by setting crossgroup constraints and comparing more restricted models with less restricted ones (Byrne, Shavelson and Muthén 1989; Steenkamp and Baumgartner 1998).

On a given set of data, the test for measurement invariance may be carried out in four steps:
1) Confirmatory factor analysis: To begin with, a confirmatory factor analysis (CFA) should be conducted separately in each group (in our case, in the Polish and German samples). At this preliminary stage, it is recommended to establish a model that fits the data in both groups (Byrne 2004). The analysis is considered complete once an appropriate model fit is reached in both groups.
2) Configural invariance: This is the lowest level of measurement invariance. Configural invariance requires both groups to have the same number of latent variables loading on the same items. The factor loadings should be significant and the correlations between the latent variables should be less than 1. This level of measurement invariance confirms that the factor structure is similar but this finding does not yet reveal whether the measurements are also equivalent across groups (Byrne 2004).

Configural invariance may be tested using MGCFA. In this way, the parameters are estimated for all groups simultaneously. The fit of the tested model provides the baseline for the further analysis of higher levels of measurement invariance.
3) Metric invariance: Establishing metric invariance means assessing whether the factor loadings of the same items are equal across groups. If this is the case, one may assume that people interpret items in both groups in the same way. Metric invariance is a requirement for comparing factor correlates across groups (covariances or unstandardized regression coefficients). It is tested by restricting the factor loadings between the observed items and the latent variable to be equal across the compared groups (Vandenberg and Lance 2000). Chen (2007) proposes cut-off criteria to help decide whether the decrease in model fit compared with the configural invariance model is substantial or not.
4) Scalar invariance: Scalar invariance is required to justify a comparison of construct means; hence, it is crucial for cross-national comparison. If scalar invariance is supported, differences in the observed variables result from disparities in the unobserved variables and are not affected by other parameters such as, for example, bias in scale use. Scalar invariance is tested by constraining not only factor loadings (as done in the case of testing for metric invariance), but also indicator intercepts to be equal across groups (Vandenberg and Lance 2000). Chen (2007) and Byrne and Stewart (2006) propose cut-off criteria to decide whether the decrease in model fit compared with the metric invariance model is substantial or not.

Several authors have suggested that two indicators with equal factor loadings are sufficient to allow comparisons of correlates. This has been termed *partial metric invariance* (Byrne et al. 1989). Similarly, they suggested that two indicators with equal factor loadings and intercepts per construct across countries are sufficient to allow a comparison of construct means. This is known as *partial scalar invariance* (Byrne et al 1989; Steenkamp and Baumgartner 1998).

3 Previous Studies

To date, the analysis of measurement invariance of human values has been performed on shorter versions of the PVQ consisting of 21 items such as in the ESS (Davidov 2008; Davidov 2010; Davidov et al. 2008) or 28 items (Steinmetz et al. 2009). However, to the best of our knowledge, no study has tested the invariance properties of the *full PVQ-40* in a crossnational framework. This is the goal of the current study.

The results of the analysis performed with the ESS data are consistent across the first three ESS rounds (Davidov 2008; Davidov 2010; Davidov et al. 2008). Configural and metric invariance could be established across most participating countries only for seven values, and scalar invariance could not be confirmed across the full set of ESS countries at all. Instead, only a few subsets of countries displayed scalar invariance for all or a subset of the values.

This is a disappointing finding for substantive researchers. After all, it would be highly interesting to compare value means across countries and draw conclusions regarding similarities and differences between them. The methodolog-

ical literature indicates that this may be done only if we can first establish full or partial scalar invariance. It could be the case that the *number* of items in the short version of the PVQ-21, used in the ESS, is insufficient to establish scalar invariance. Thus, a longer and more complete set of questions to measure the values may offer a better position to identify higher levels of invariance. After all, Byrne et al. (1989) suggested that it is sufficient to have two items with equal loadings and intercepts to guarantee partial scalar invariance. With more items per value construct in the full PVQ version, the chances of reaching this goal are higher. In the current study we will employ the full 40-item version of the PVQ to identify a model that fits the data in both countries and to test this proposition and assess the invariance properties across the two samples, the German and the Polish. In a second set of analyses, we test for invariance across the German and Polish samples using a *subset* of the 40 items. This 21-item subset corresponds to the items used in the PVQ-21, the human values scale used in the ESS. In this way we compare the level of invariance across the two countries using the PVQ-40 and PVQ-21. Our hypothesis is that by using the PVQ-40 we reach scalar invariance, at least for some of the values.

4 Method

Sample

Data for analyses were collected from a total of 1,204 individuals in Germany and Poland. The participants were students attending various secondary schools and high schools, as well as college students mostly from Warsaw (for the Polish sample) and Gießen (for the German sample). Participation was voluntary, and anonymity was guaranteed. The questionnaire was administered in both samples by distributing it among the participants who filled it in themselves in the classroom voluntarily after their lectures (data for the Polish sample was collected by the first author of this study). The Polish and German samples consisted of 602 respondents each. The respondents' ages ranged from 18 to 30 years (the mean age in the German sample was 20.6, SD = 1.85; the mean age in the Polish sample was 22.6, SD = 3.36). In the German sample, 110 respondents were male (18.3%), and 489 were female (81.2%). In the Polish sample there were 254 males (42.2%) and 348 females (57.8%) (for further details about the German data, see Schmidt et al. 2007).[1]

The Questionnaire

The PVQ-40, in its German version (cf. Schmidt et al. 2007), was used along with the Polish version prepared by Cieciuch and Zaleski (2011). Both versions were obtained with independent translations and back-translations. The correctness of the Polish and German translation was controlled by Shalom Schwartz, author of the PVQ (for a list of the question items, see Appendix).

The scale contains 40 items presented in the form of short verbal portraits of 40 different people. Each portrait describes a person's goals, aspirations, or desires that point implicitly to the importance of a value. For each portrait, the respondents had to answer the question: "How much like you is this person?". Response categories ranged from 1 to 6: "very much like me" (6), "like me" (5), "somewhat like me" (4), "a little like me" (3), "not like me" (2), and "not like me at all" (1). The number of portraits for each value ranges from three (stimulation, hedonism, and power) to six (universalism), reflecting the conceptual breadth of the values.

5 Results

Before starting with the two-group analysis, we analyzed the data in each sample separately. At first, two covariance matrices were constructed, one for Germany and one for Poland, as input for the models.[2] Pairwise deletion was used to deal with missing values because it is considered a better strategy than listwise deletion and is appropriate in cases where there are relatively low numbers of missing values as was observed here (see also Schnell 1986, pp. 84-85, 99-100). There were, on average, 1.2% missing values in the German sample and almost none (0.1%) in the Polish sample.

We used four types of global fit measures as criteria to decide whether the model is acceptable. Root mean square error of approximation (RMSEA) reflects the degree to which a researcher's model reasonably fits the population covariance matrix while taking into account the degrees of freedom and sample size (Brown 2006). It is a parsimony-adjusted index that favors simpler models. The probability of close fit (Pclose) indicates the probability that RMSEA is below 0.05. When the RMSEA value is smaller than 0.05 and the Pclose value is larger than 0.5, the model can be assumed to perform very well (Browne and Cudeck 1993). When the RMSEA value is 0.08 or below, the model can be assumed to perform reasonably well (Hu and Bentler 1999; Marsh, Hau and Wen 2004). In such a case Pclose may be lower than 0.5 since it reflects the probability that RMSEA is smaller than 0.05. The comparative fit index (CFI) compares the fit of a researcher's model to a more restricted baseline model. CFI values between 0.90 and 0.95 or larger indicate an acceptable model fit (Hu and Bentler 1999). The standardized root mean square residual (SRMR) compares the sample variances and covariances to the estimated ones (Arbuckle 2005). When the SRMR value is smaller than 0.05, the model can be assumed to perform very well and when it is lower than 0.08, the model can be assumed to perform reasonably well (Hu and Bentler 1999; Marsh et al. 2004). Since the p value is sensitive to the sample size, we do not rely on it (Saris, Satorra and van der Veld 2009). To decide whether the fit of more restrictive models deteriorate significantly, we rely on cut-off criteria suggested by Chen (2007). According to this study, if the sample is larger than 300, metric noninvariance is indicated by a change larger than .01 in CFI, supplemented by a change larger than .015 in RMSEA or a change larger

[1] All significance tests in the analyses are based on the assumption of simple random samples, which is not the case in this study. Therefore, all significance tests will give incorrect p-values.

[2] The data and covariance matrices may be provided by the first author upon request.

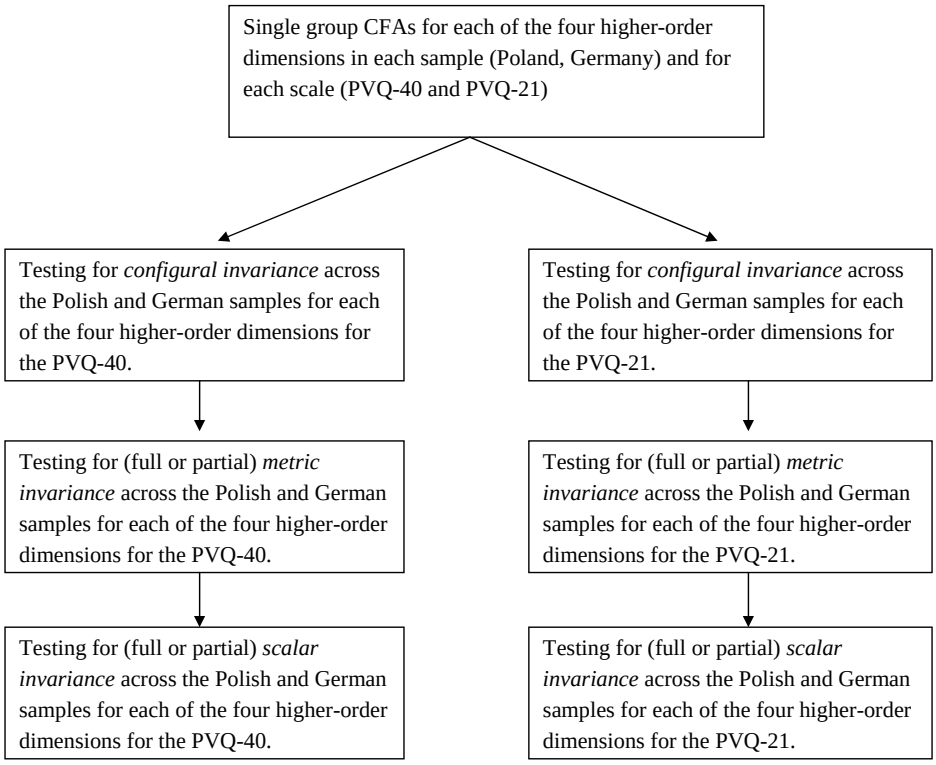

Figure 1. Sequence of model testing

than .03 in SRMR compared with the configural invariance model. Regarding scalar invariance, noninvariance is indicated by a change larger than .01 in CFI, supplemented by a change larger than .015 in RMSEA or a change larger than .01 in SRMR compared with the metric invariance model. As an overall criterion we use the CFI, according to which changes larger than 0.01 indicate noninvariance (Byrne and Stewart 2006). All subsequent analyses were conducted with the program Amos 7.0 (Arbuckle 2005).

We decided to create four models separately, one for each higher-order dimension in each sample. This decision is not uncommon in literature on the topic. For example, Spini (2003) tested measurement invariance of the Schwartz Value Survey for each of the 10 value types separately. In the analysis described here, the theoretical division of value types into four higher-order values as proposed by Schwartz (2006) was followed. Thus, the following values were included in the model of each higher-order dimension: *self-transcendence* included universalism and benevolence; *self-enhancement* included achievement, power, and hedonism; *conservation* included conformity, tradition, and security; and *openness* to change included the values self-direction, stimulation, and hedonism. According to theoretical considerations, hedonism lies between self-enhancement and openness to change. Hence, the decision was made to include hedonism in both dimensions.

Four models were constructed in each sample, one for each higher-order dimension, for which the measurement invariance properties across the German and the Polish samples were tested. We proceeded in the following manner. First, we performed a CFA for each country and for each dimension separately. Where necessary, based on the modification indices and the expected parameter changes, modifications to the models were made. Next, we turned to the two-group CFA to assess configural, metric, and scalar invariance. The two-group CFA across the German and Polish samples was conducted four times, once for each higher-order dimension. If necessary, we opted for the partial invariance model. Since the two samples were different in terms of the gender distribution, we repeated, in a final step, the two-group CFA analyses controlling for the effect of gender, to test whether the results are affected by this uneven gender distribution.

We then repeated the analysis with 21 items from the PVQ-40. We chose the same or the most similar items to the items used in PVQ-21 in the ESS. Figure 1 visualizes the sequence of model testing.

5.1 Single Sample Analyses of the PVQ-40

Self-transcendence. First, separate CFA analyses in each sample demonstrated that the model fitted the data marginally in the German sample, and the fit was acceptable in the Polish sample (in Germany: RMSEA = .086; Pclose = .00; SRMR = .056; CFI = .941; in Poland: RMSEA = .064; Pclose = .077; SRMR = .034; CFI = .971). Some modifi-

Table 1: Global Fit Measures for the Single Sample CFAs with the PVQ-40

		χ^2	df	RMSEA	PCLOSE	SRMR	CFI
Self-transcendence	DE	97.4	18	.086	.000	.056	.941
	PL	63.0	18	.064	.077	.034	.971
Self-enhancement	DE	100.4	23	.075	.003	.059	.960
	PL	122.2	23	.085	.000	.050	.948
Conservation	DE	131.4	41	.061	.060	.049	.913
	PL	114.6	41	.055	.245	.041	.929
Openness to change	DE	100.4	23	.075	.003	.076	.934
	PL	98.3	23	.074	.004	.041	.953

RMSEA = root mean square error of approximation; PCLOSE = probability of close fit; SRMR = standardized root mean square residual; CFI = comparative fit index; DE = Germany; PL = Poland

cations had to be included in the model. The item 'importance of equal treatment' (un8) required a cross-loading to the benevolence construct. Once it was added into the model, its loading on the universalism construct became very small (a standardized loading of .08 in Poland and .11 in Germany). It seemed that it was a better measure of benevolence than of universalism. Thus, we decided to drop this item from further analyses. The item 'importance of forgiving' (be33) had only a small standardized loading on the benevolence latent variable (.27) in Poland; therefore, we decided to drop this item as well from further analyses. According to the modification indices, we added an error correlation between the error of 'importance of caring for nature' (un19) and 'importance of adapting to nature' (un40). It is worth noting that both items concern the protection of nature. Thus, the semantic similarity of the items justifies adding an error correlation. The fit measures are reported in Table 1.

Self-enhancement. Separate CFA analyses of self-enhancement indicated that also these models are acceptable. It was necessary to add one error correlation between the errors of the two items of achievement 'important to be ambitious' (ac24) and 'important to be better than others' (ac32). The two items measure outdoing others and this justifies adding an error correlation. It was also necessary to add a cross-loading between the item 'important to be rich' (po2) and the achievement construct. Therefore, we decided to drop this item. The fit measures are reported in Table 1.

Conservation. In the separate CFA analyses of the conservation metacategory it was necessary to unify the values of conformity and tradition. The two types of values introduced into the model as separate variables generated a nonpositive definite matrix of correlations between the constructs. Unifying conformity and tradition is not only empirically necessary; it is also justified on theoretical grounds. Davidov et al. (2008) and Schwartz (2006) argue that the human values are on a continuum just like the colors of the rainbow: They flow into one another without an abrupt demarcation. In empirical studies, some values may turn out to be too close to each other to be modeled separately. In such a case it may be justified to unify them. However, as Beierlein et al. in this volume demonstrate, the necessity to unify values might also be a result of the choice of items included in the questionnaire and may be avoided if different items are selected to measure the values.

In our data some further modifications were necessary. Two items measuring tradition, 'important not to ask for more than what one has' (tr9) and 'important to be humble and modest' (tr38), had to be dropped for the following reasons: The standardized factor loadings of the item tr9 were too low (in Germany .19 and in Poland .37). The item tr38 necessitated adding several error correlations with other items and had a rather low standardized factor loading in Germany (.31). The two items seemed to measure an additional dimension of humility. In addition, two error correlations were added: one between the security items 'importance of country security' (sc14) and 'importance of social order and stable government' (sc35) and one between the conformity items 'important to behave properly' (co16) and 'important to be polite' (co36). The first pair of items measures societal security and the second measures not upsetting others. Therefore, it is justifiable to allow their errors to correlate. This model had an acceptable fit to the data in both samples. The fit measures are reported in Table 1.

Openness to change. The single-sample CFAs for the openness to change dimension performed relatively well. As noted above, the hedonism value construct was included here as well. To reach a sufficient model fit, we dropped the item 'important to do a lot of different things in life' (st6) because of a low factor loading and a cross-loading to the value construct self-direction. In addition, it was necessary to add one error correlation between two items measuring the selfdirection value construct, 'important to make own decisions' (sd11) and 'important to be independent' (sd34). The two items tap freedom and independence in action. The fit measures for this model are also reported in Table 1.

5.2 Two-group Analyses of the PVQ-40

Now that the models in the single samples have been established, we turn to the two-group analyses to assess measurement invariance for each higher-order dimension across the German and Polish samples. The global fit measures are reported in Table 2.

Self-transcendence. An inspection of Table 2 reveals that both the configural and metric invariance models for the

Table 2: Global Fit Measures for the Two-Group CFA across Poland and Germany with the PVQ-40

	Level of invariance	χ^2	df	RMSEA	PCLOSE	SRMR	CFI
Self-transcendence	configural	160.4	36	.054	.222	.034	.957
	metric	164.4	42	.049	.536	.036	.958
	scalar	206.8	48	.053	.270	.037	.945
	partial scalar	174.6	47	.048	.683	.036	.956
Self-enhancement	configural	222.7	46	.057	.067	.050	.954
	metric	252.4	52	.057	.053	.060	.948
	scalar	321.4	58	.062	.002	.063	.932
	partial scalar	273.9	57	.057	.055	.061	.944
Conservation	configural	246.0	82	.041	.995	.041	.921
	metric	268.8	91	.040	.998	.047	.914
	scalar	599.4	100	.065	.00	.070	.760
	partial scalar	291.4	96	.041	.996	.050	.906
Openness to change	configural	198.7	46	.053	.267	.076	.945
	metric	237.2	52	.055	.137	.072	.934
	scalar	419.1	58	.072	.00	.075	.870
	partial scalar	245.6	55	.054	.172	.070	.932

RMSEA = root mean square error of approximation; PCLOSE = probability of close fit; SRMR = standardized root mean square residual; CFI = comparative fit index

universalism and benevolence values may be accepted based on the global fit measures. The change in RMSEA, SRMR, and CFI is below the recommended criteria (Chen 2007). However, we rejected full scalar invariance. Whereas SRMR and RMSEA displayed only a small change, CFI decreased slightly more than the recommended .01 criteria (ΔCFI = .013). Next, we tested for partial scalar invariance by releasing the equality constraint for the intercept of the universalism item 'importance that everyone be treated justly' (un29). The equality constraint on this intercept was released because it displayed the most severe violation of invariance compared to the other indicator intercepts in the model. After releasing this intercept, partial scalar invariance could be guaranteed.

On the basis of the analysis performed, one may conclude that a sufficient level of invariance is present for the self-transcendence values enabling a comparison of their means between the German and the Polish samples.

Self-enhancement. The test of configural and metric invariance for the power, achievement, and hedonism values resulted in a satisfactory model fit, as can be observed in Table 2. The decrease in the various fit measures in the metric invariance model compared to the configural invariance model was below the recommended criteria (Chen 2007). However, full scalar invariance could not be supported by the data. The decrease in CFI compared with the metric invariance model was beyond the recommended .01 criteria (ΔCFI = .016). Next, we tested for partial scalar invariance. The equality constraint on the intercept of the hedonism item 'important to have fun' (he10) was released because it displayed the most severe violation of invariance compared to the other indicator intercepts in the model. After releasing this intercept, partial scalar invariance was supported by the data. Thus, the means of the self-enhancement values may be compared between the German and the Polish samples.

Conservation. An inspection of Table 2 reveals that both the configural and metric invariance models for the conformity, tradition, and security values may be accepted based on the global fit measures. The difference in RMSEA, SRMR, and CFI between the two models is below the recommended criteria (Chen 2007). However, we rejected full scalar invariance. The differences in the RMSEA, CFI, and SRMR between the scalar and the metric invariance models are very large and beyond the recommended criteria. Next, we tested for partial scalar invariance. The equality constraints on the intercepts of the conformity items 'important to behave properly' (co16) and 'important to be polite' (co36) and on the security items 'importance of country security' (sc14) and 'important to avoid getting sick' (sc31) were released because they displayed the most severe violations of invariance compared to the other indicator intercepts in the model. After releasing these intercepts, partial scalar invariance was supported by the data (Byrne et al. 1989). The unified conformity-tradition construct remained with four items with constrained intercepts and the security value remained with three fully constrained items. Thus, also the means of the conformity-tradition and security values may be compared across the German and the Polish samples.

Openness to change. Finally, we tested for invariance of the self-direction, stimulation, and hedonism values. The test of configural and metric invariance for the openness to change values resulted in a satisfactory model fit, as can be observed in Table 2. The decrease in most of the global fit measures in the metric invariance model compared to the configural invariance model was below the recommended criteria (Chen 2007). The change in CFI was only slightly beyond the recommended criteria (ΔCFI = .011; see Byrne and Stewart 2006). However, the scalar invariance model

Table 3: Global Fit Measures for the Single Sample CFAs with PVQ-21

		χ^2	df	RMSEA	PCLOSE	SRMR	CFI
Self-transcendence	DE	28.3	4	.101	.007	.043	.933
	PL	8.2	4	.042	.562	.020	.991
Self-enhancement	DE	34.5	6	.089	.011	.033	.963
	PL	15.3	6	.051	.432	.022	.988
Conservation	DE	36.8	9	.072	.059	.046	.921
	PL	37.9	9	.073	.050	.042	.920
Openness to change	DE	25.4	6	.074	.079	.039	.956
	PL	25.9	6	.074	.073	.027	.972

RMSEA = root mean square error of approximation; PCLOSE = probability of close fit; SRMR = standardized root mean square residual; CFI = comparative fit index; DE = Germany; PL = Poland

was rejected. The decrease in the various fit measures was beyond the recommended criteria compared with the metric invariance model. Next, we tested for partial scalar invariance. The equality constraints on the intercepts of the hedonism item 'important to have fun' (he10), self-direction item 'important to be independent' (sd34) and stimulation item 'important to have an exciting life' (st30) were released. After releasing these intercepts, partial scalar invariance was supported by the data for the hedonism and self-direction value constructs. Each of them had at least two items with constrained factor loadings and intercepts across samples, thus allowing the comparison of their means across the German and Polish samples (Byrne et al. 1989). However, the means of the stimulation value may not be compared across the samples. After releasing the equality constraints of the intercept of one of its indicators, it remained with only one fully constrained indicator. This does not allow a meaningful mean comparison across the two samples.

5.3 Controlling for Composition Effects of Gender

We mentioned previously that there was an uneven distribution of gender groups across the two samples. In order to examine whether this had any influence on the results, we repeated the two-group CFAs for each model controlling for the effect of gender on the values included in the models by regressing it on each of the values. The conclusions with respect to the level of invariance established did not change in all the analyses, thus confirming that the uneven distribution of gender did not influence our results.

5.4 Analyses of the PVQ-21

We repeated the analysis using 21 items from the PVQ-40. We chose those items which are identical or most similar to the items used in PVQ-21 in the ESS and tested the four models described above. In the Appendix, the items selected for the analysis are marked with an asterisk. We will not report the results in detail as we did for the analysis with the PVQ-40 because of space limitations. Detailed outputs may be obtained from the first author upon request.

At first, separate CFA analyses of each higher-order value dimension indicated that these models also fitted the data reasonably well. The global fit measures are reported in Table 3.

Next, we turned to the two-group analyses to assess measurement invariance for the values in each higher-order dimension across the German and Polish samples. The global fit measures are reported in Table 4.

Although we were able to reach satisfactory global fit measures, these models required releasing the equality constraints of several items for various values. The analysis of the self-transcendence values required releasing the equality constraints of the factor loadings and intercepts of one of the indicators of benevolence (be18 – important to be loyal). Thus, benevolence remained with only one fully constrained indicator (be12 – important to help people around). Analyses of the self-enhancement and openness to change values required releasing the factor loadings and intercepts of one indicator of hedonism (he26 – enjoying life's pleasures) and one indicator of stimulation (st15 – taking risks). Consequently, partial or full scalar invariance could not be established for the two values hedonism and stimulation. Thus, comparing the means of the values hedonism and stimulation as well as those of the benevolence value may be not meaningful.

6 Summary and Discussion

Data analyses of the PVQ-40 revealed that all the values displayed partial scalar invariance with the exception of the stimulation value. Thus, the means of nine out of the 10 values may be compared across the samples. By contrast, partial scalar invariance was attained for only seven out of 10 values as measured by the PVQ-21. Therefore, we conclude that the full version of PVQ may have better properties for cross-cultural research.

This is good news for survey researchers interested in a more elaborate study of values in the two countries. However, this result is not surprising: After all, we have to remember that guaranteeing partial scalar invariance requires having two items with equal loadings and intercepts per construct. With more items per value construct in the full PVQ version, there is a better chance to reach this goal. Merely having more questions per value construct may guarantee higher levels of invariance which in turn enables researchers to engage in more elaborate cross-national comparisons of values in a meaningful way.

Thus, the findings encourage the collection of representative data with the full PVQ-40. However, this conclusion

Table 4: Global Fit Measures for the Two-Group CFA across Poland and Germany with the PVQ-21

	Level of invariance	χ^2	df	RMSEA	PCLOSE	SRMR	CFI
Self-transcendence	configural	36.6	8	.055	.30	.020	.965
	metric	61.0	11	.062	.09	.033	.939
	partial metric	44.2	10	.053	.33	.027	.959
	partial scalar	46.8	12	.049	.50	.024	.958
Self-enhancement	configural	49.8	12	.051	.41	.022	.976
	metric	62.4	15	.051	.41	.023	.970
	scalar	145.8	18	.077	.00	.050	.918
	partial scalar	65.5	17	.049	.41	.026	.969
Conservation	configural	74.7	18	.051	.41	.042	.920
	metric	86.5	23	.048	.60	.048	.911
	scalar	154.8	28	.062	.02	.054	.822
	partial scalar	86.9	25	.046	.75	.049	.913
Openness to change	configural	51.3	12	.052	.37	.027	.966
	metric	76.2	15	.058	.13	.034	.947
	partial metric	58.8	13	.054	.29	.029	.960
	partial scalar	67.4	16	.052	.26	.032	.955

RMSEA = root mean square error of approximation; PCLOSE = probability of close fit; SRMR = standardized root mean square residual; CFI = comparative fit index

may not be generalized to other samples in other countries. Invariance analysis of the full PVQ across other countries should be repeated to find out whether the longer scale provides higher levels of invariance in other cases as well.

Once scalar invariance is guaranteed, there are different strategies to compare means across samples. First, one may compute a simple average index out of the reported values and in the next step compute the so-called 'simple means' (Coromina, Saris and Oberski 2009:13; Saris and Gallhofer 2007:340-41). However, Schwartz (2006) argued that simple means are problematic because respondents differ in their use of the response scale. Some people distribute their importance ratings across the whole rating scale, while others tend to rate most value items as quite important or tend to give lower ratings to most items. Because the scale should measure the values' priorities and not the way the scale was used, Schwartz (2006) proposed to center the responses on the individual means. Centering can be done by subtracting the overall individual mean from the score of each of the value items. In the next step, one may compute an index with the centered items. In this way, centered value means can be obtained. Finally, a third way to compare values is to compare the means of the latent variables. This is the preferred way as it can account for partial scalar invariance and control for measurement errors (Coromina et al. 2009:13). Sörbom (1974) has shown that to compare means of latent variables, one of the groups should be defined as the reference group. Alternatively, one item in each value should be defined as the reference item for all groups, and its intercept should be constrained to zero. Thus, the latent means are only arbitrary and one should compute the latent mean *differences* across the groups rather than directly interpret the absolute means in each group (for alternatives, see Little, Slegers and Card 2006).

Finally, in this study we followed certain steps in the assessment of invariance. Such steps may be followed not only in cross-national studies of values but also in cross-national studies involving any other theoretical construct of interest. Eventually, different language use, different cultures, or a dissimilar scale use may bias the cross-cultural validity of the results and prevent meaningful cross-cultural comparisons.

Acknowledgements

The research was supported by grants from the Polish Ministry of Science and Higher Education in the years 2009-2010 to the first author. The authors would like to thank Lisa Trierweiler for the English proof of the manuscript.

References

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. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (p. 247-264). New York, NY: John Wiley.

Bilsky, W., Janik, M., & Schwartz, S. H. (2011). The structural organization of human values - Evidence from three rounds of the European Social Survey (ESS). *Journal of Cross-Cultural Psychology*, *42*(5), 759-776.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (p. 136-162). Newbury Park, CA: Sage.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS Graphics: A road less traveled. *Structural Equation Modeling*, *11*(2), 272-300.

Byrne, B. M., Leong, F. T. L., Hambleton, R. K., Oakland, T., van de Vijver, F. J. R., & Cheung, F. M. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Educational Professional Psychology*, *3*(2), 94-105.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287-321.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504.

Chen, F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparison in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005-1018.

Cieciuch, J., & Zaleski, Z. (2011). Polska adaptacja Portretowego Kwestionariusza Wartości Shaloma Schwartza [Polish adaptation of Portraits Value Questionnaire of Shalom Schwartz]. *Czasopismo Psychologiczne [Psychological Journal]*, *17*(2), 251-262.

Coromina, L., Saris, W. E., & Oberski, D. (2009, June). *The quality of the measurement of interest in the political issues presented in the media in the ESS.* Paper presented at the QMSS2 seminar of cross-national comparisons on the Quality and Comparability of Measures for Constructs in Comparative Research: Methods and Applications, Bolzano (Bozen), Italy.

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. (2010). Testing for comparability of human values across countries and time with the third round of the European Social Survey. *International Journal of Comparative Sociology*, *51*(3), 171-191.

Davidov, E., Schmidt, P., & Schwartz, S. (2008). Bringing values back in. The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, *72*(3), 420-445.

De Beuckelaer, A. (2005). *Measurement invariance issues in international management research.* Unpublished doctoral dissertation, Limburgs Universitair Centrum, Limburg, the Netherlands.

Fontaine, J., Poortinga, Y., Delbeke, L., & Schwartz, S. (2008). Structural equivalence of the value domain across cultures: Distinguishing sampling fluctuations from meaningful variation. *Journal of Cross-Cultural Psychology*, *39*(4), 345-365.

Horn, J., & McArdle, J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 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), 1-55.

Jöreskog, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409-426.

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.

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*(3), 320-341.

Saris, W., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. New York: Wiley.

Saris, W. E., Satorra, A., & van der Veld, W. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, *16*(4), 561-582.

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 für Sozialpsychologie*, *38*(4), 249-263.

Schnell, R. (1986). *Missing-Data-Probleme in der empirischen Sozialforschung [Missing Data Problems in social Research].* Unpublished doctoral dissertation, University of Bochum.

Schwartz, S. (2006). Basic human values: Theory, measurement, and applications. *Revue française de sociologie*, *47*(4), 929-968.

Schwartz, S., & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: Effects of gender equality. *Journal of Personality and Social Psychology*, *97*(1), 171-185.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, p. 1-65). New York: Academic Press.

Schwartz, S. H. (2005). Robustness and fruitfulness of a theory of universals in individual human values. In A. Tamayo & J. B. Porto (Eds.), *Values and behavior in organizations* (p. 56-95). Brasilia: Editora Universidade de Brasilia.

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., & Harris, M. (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*(5), 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*(2), 229-239.

Spini, D. (2003). Measurement equivalence of 10 values types from SVS across 21 countries. *Journal of Cross-Cultural Psychology*, *34*(1), 3-23.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78-90.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. (2009). Testing measurement invariance using multigroup cfa differences between educational groups in human values measurement. *Quality and Quantity*, *43*, 599-616.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Toward an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, *13*(1), 29-37.

Vandenberg, R. J. (2002). Towards a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, *5*(2), 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*(1), 4-69.

Vecchione, M., Casconi, T., & Barbaranelli, C. (2009). Assessing the circular structure of the Portrait Values Questionnaire. *European Journal of Psychological Assessment*, *25*(4), 231-238.

Appendix

Table A1: Question Items (PVQ-40, PVQ-21 as used in the ESS, and the 21 items from the PVQ-40 taken for our PVQ-21 analysis)

The PVQ-40 question items (*Items used for the PVQ-21 analysis)	The original PVQ-21 question items as used in the ESS (Items slightly different from those we used in the analysis are in italics)
Conformity (CO)	
co7*. 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.	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.
co16*. It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong.	16. It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong.
co28. He believes he should always show respect to his parents and to older people. It is important to him to be obedient.	
co36. It is important to him to be polite to other people all the time. He tries never to disturb or irritate others.	
Tradition (TR)	
tr9. He thinks it's important not to ask for more than what you have. He believes that people should be satisfied with what they have.	
tr20. Religious belief is important to him. He tries hard to do what his religion requires.	
tr25*. He thinks it is best to do things in traditional ways. It is important to him to keep up the customs he has learned.	*20. Tradition is important to him. He tries to follow the customs handed down by his religion or his family.*
tr38*. It is important to him to be humble and modest. He tries not to draw attention to himself.	9. It is important to him to be humble and and modest. He tries not to draw attention to himself.
Benevolence (BE)	
be12*. It's very important to him to help the people around him. He wants to care for their well-being.	12. It's very important to him to help the people around him. He wants to care for their well-being.
be18*. It is important to him to be loyal to his friends. He wants to devote himself to people close to him.	18. It is important to him to be loyal to his friends. He wants to devote himself to people close to him.
be27. It is important to him to respond to the needs of others. He tries to support those he knows.	
be33. Forgiving people who have hurt him is important to him. He tries to see what is good in them and not to hold a grudge.	
Universalism (UN)	
un3*. He thinks it is important that every person in the world be treated equally. He believes everyone should have equal opportunities in life.	3. He thinks it is important that every person in the world should be treated equally. He believes everyone should have equal opportunities in life.
un8*. 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.	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.

Table A1: Continued

un19*. He strongly believes that people should care for nature. Looking after the environment is important to him.	19. He strongly believes that people should care for nature. Looking after the environment is important to him.
un23. He believes all the worlds' people should live in harmony. Promoting peace among all groups in the world is important to him.	
un29. He wants everyone to be treated justly, even people he doesnt know. It is important to him to protect the weak in society.	
un40. It is important to him to adapt to nature and to fit into it. He believes that people should not change nature.	
Self-direction (SD)	
sd1*. Thinking up new ideas and being creative is important to him. He likes to do things in his own original way.	1. Thinking up new ideas and being creative is important to him. He likes to do things in his own original way.
sd11*. It is important to him to make his own decisions about what he does. He likes to be free to plan and to choose his activities for himself.	*11. It is important to him to make his own decisions about what he does. He likes to be free and not depend on others.*
sd22. He thinks it's important to be interested in things. He likes to be curious and to try to understand all sorts of things.	
sd34. It is important to him to be independent. He likes to rely on himself.	
Stimulation (ST)	
st6. He thinks it is important to do lots of different things in life. He always looks for new things to try.	
st15*. He likes to take risks. He is always looking for adventures.	*15. He looks for adventures and likes to take risks. He wants to have an exciting life.*
st30*. He likes surprises. It is important to him to have an exciting life.	*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.*
Hedonism (HE)	
he10*. He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.	21. He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.
he26*. Enjoying life's pleasures is important to him. He likes to 'spoil' himself.	*10. Having a good time is important to him. He likes to "spoil" himself.*
he37. He really wants to enjoy life. Having a good time is very important to him.	
Achievement (AC)	
ac4*. It's very important to him to show his abilities. He wants people to admire what he does.	4. It's important to him to show his abilities. He wants people to admire what he does.
ac13*. Being very successful is important to him. He likes to impress other people.	*13. Being very successful is important to him. He hopes people will recognize his achievements.*
ac24. He thinks it is important to be ambitious. He wants to show how capable he is.	
ac32. Getting ahead in life is important to him. He strives to do better than others.	

Table A1: Continued

Power (PO)	
po2*. It is important to him to be rich. He wants to have a lot of money and expensive things.	2. It is important to him to be rich. He wants to have a lot of money and expensive things.
po17*. It is important to him to be in charge and tell others what to do. He wants people to do what he says.	*17. It is important to him to get respect from others. He wants people to do what he says.*
po39. He always wants to be the one who makes the decisions. He likes to be the leader.	
Security (SC)	
sc5*. It is important to him to live in secure surroundings. He avoids anything that might endanger his safety.	5. It is important to him to live in secure surroundings. He avoids anything that might endanger his safety.
sc14*. It is very important to him that his country be safe. He thinks the state must be on watch against threats from within and without.	*14. It is important to him that the government ensures his safety against all threats. He wants the state to be strong so it can defend its citizens.*
sc21. It is important to him that things be organized and clean. He really does not like things to be a mess. sc31. He tries hard to avoid getting sick. Staying healthy is very important to him. sc35. Having a stable government is important to him. He is concerned that the social order be protected.	

co = conformity; tr = tradition; be = benevolence; un = universalism; sd = selfdirection; st = stimulation; he = hedonism; ac = achievement; po = power; sc = security.
The number next to the abbreviation of each value name corresponds to its order in the questionnaire.