Survey Research Methods (2012)
Vol.6, No.1, pp. 25-36
ISSN 1864-3361
http://www.surveymethods.org

Testing the discriminant validity of Schwartz' Portrait Value Questionnaire items – A replication and extension of Knoppen and Saris (2009)

Constanze Beierlein
GESIS, Germany

Eldad Davidov
University of Zurich

Peter Schmidt
National Research University, Moscow

Shalom H. Schwartz
Hebrew University of Jerusalem/
National Research University, Moscow

Beatrice Rammstedt
GESIS, Germany

Schwartz' theory of ten basic human values has stimulated numerous studies using a variety of instruments. Confirmatory factor analyses (CFA) of the properties of some of the instruments have revealed that three pairs of values were excessively highly correlated. This led Davidov et al. (2008) to propose unifying values. To overcome the problems of loss of precision due to unifying distinct values, Knoppen and Saris (2009a,b) investigated the factorial structure of each of the ten values measured with the PVQ (Schwartz et al. 2001). They identified both cross-loadings and distinct sub-dimensions for the pairs of nondiscriminated values in two German student samples. They concluded that the original strategy for selecting items, maximizing theoretical coverage at the expense of item homogeneity, produced the poor discrimination between values. Our Study 1 examines whether the Knoppen and Saris findings generalize to a representative sample of the German population. With some notable exceptions, our findings replicate theirs. Study 2 uses 33 items from an experimental version of the PVQ to operationalize and test a full model of the 11 basic values. Following Knoppen and Saris, we included only one sub-dimension of each of the 11 values. This CFA model yielded a satisfactory fit with no estimation problems. We conclude that available indicators permit measuring the distinct values without the need to collapse factors. Limitations and implications of the research are discussed.

Keywords: Portrait Value Questionnaire, confirmatory factor analysis, discriminant validity, factor structure

1 Introduction

Recent years have seen a proliferation of studies that apply the Schwartz (1992) theory of basic human values using a variety of measurement scales. Studies have assessed the measurement properties of the scales (e.g., Davidov, Schmidt and Schwartz 2008; Perrinjaquet, Furrer, Usunier, Cestre and Valette-Florence 2007; Vecchione, Casconi and Barbaranelli 2009), factors that influence people's values as measured by the scales (e.g., Meulemann, Davidov, Schmidt and Billiet 2010; Schwartz 2006; Steinmetz, Schmidt, Tina-Booh, Wieczorek and Schwartz 2009), and individual differences in attitudes and behavior traceable to the measured values (e.g., Bardi and Schwartz 2003; Piurko, Schwartz and Davidov 2011). These studies have advanced our knowledge of the measurement of values and their relation to other theoretical constructs of interest. Moreover, the studies have underlined the usefulness of a valid scale to measure human values for researchers who wish to explain attitudes, opinions or behavior. The fruitfulness of this line of research led the designers of the bi-annual European Social Survey (ESS)[1] to include a 21-item short version of one of the scales to measure human values in each of its five survey rounds commencing in 2002.

Studies using confirmatory factor analysis (CFA) on data of the ESS short scale have not supported the full Schwartz model of human values (e.g., Davidov 2008, 2010). Excessively high correlations among three pairs of values (universalism and benevolence, conformity and tradition, power and achievement), prevented separate modeling of each of the factors that represent these values (cf. Knoppen and Saris 2009a). These correlated factors precluded the estimation of the full model of ten distinct values. Davidov and his col-

Contact information: Constanze Beierlein, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany, e-mail: constanze.beierlein@gew.de

[1] see www.europeansocialsurvey.org/; Jowell, Roberts, Fitzgerald and Eva 2007.

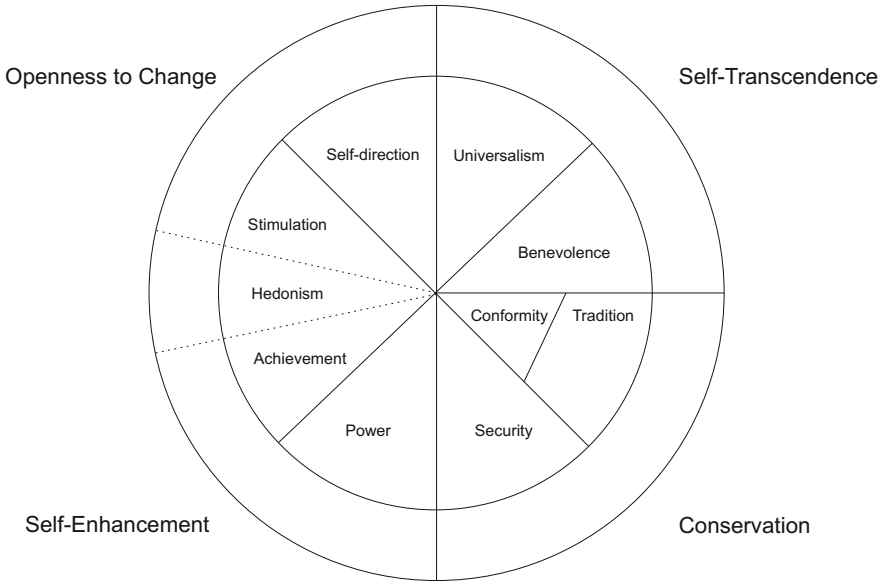

Figure 1. Structural relations among the ten basic human values and the two underlying dimensions (figure adapted from Schwartz, 1992, with permission by the author).

leagues suggested unifying these pairs of values instead of measuring them separately (cf. Davidov 2008; Davidov et al. 2008). But unifying values makes it impossible for researchers to study the predictive and explanatory utility of each of the ten values in the theory. An intensified search for improving measurement of the values is therefore underway.

2 Theory and measurement of Schwartz' basic human values

Schwartz (1992, 2003) defined basic human values as cognitive representations of desirable goals that transcend specific actions and situations. Values serve as standards for evaluating actions, events, and people. Schwartz (1992) derived ten broad basic human values, each of which expresses a different motivational goal. Figure 1 displays the values in a circular structure (a quasi-circumplex) that reflects their conceptual relations with one another (Schwartz 1992; Schwartz and Boehnke 2004). The closer any two values going around the circle, the more compatible the motivational goals they express; the more distant the two values, the less compatible and more opposed their goals. Consequently, the correlations among values should be quite high for values adjacent in the circle and become less positive as the distance between a pair of values increases, possibly becoming negative for opposing values.

Based on his own and other's analyses, Schwartz (1992; 2006; de Clercq, Fontaine and Anseel 2008) claimed that the set of ten values is reasonably comprehensive of the range of motivationally different values that are recognized in all societies.

2.1 Investigations of the discriminant and the content validity of the Portrait Value Questionnaire

The current study focuses on the method of measurement used in the Portrait Values Questionnaire (PVQ), the method also adopted by the ESS (Schwartz et al. 2001; Schwartz 2003, 2006). In order to explain the measurement problems described above and seek ways to remedy them, Knoppen and Saris (2009a,b) employed structural equation modeling (SEM) to investigate the characteristics of the PVQ items in two German student samples. First, they ran a two-factorial analysis for the items of each pair of adjacent values. They concluded that (a) some of the items chosen to measure particular values (e.g., power) had cross-loadings on other values (e.g., achievement) and low loadings on their intended value and (b) some values actually included two subdimensions (e.g., achievement). Their findings contribute to explaining the lack of discriminant validity reported previously.

Second, Knoppen and Saris (2009a,b) considered quantitative properties (e.g., simple structure of factor loadings) and qualitative criteria (their assessment of the content validity of items) in order to select items for further use. They proposed a subset of 22 PVQ items to assess 11 values, splitting the original complex universalism value into two separate values – equality and environment (Knoppen and Saris 2009c).[2] Early explications and analyses had suggested that the universalism value encompasses two or three subdimensions (e.g., Schwartz 1992; Schwartz and Boehnke 2004). In addition, Knoppen and Saris proposed viewing

[2] Knoppen and Saris have revised this unpublished manuscript and now propose a model of 15 rather than 11 values (personal communication, December 2010).

each of the two sentences that constitute a PVQ item as a separate indicator of the respective construct (Saris and Gallhofer 2002; cf. Schwartz 2003).

Knoppen and Saris (2009a,b) provide clear and detailed insights into the factorial structure of the current set of PVQ items. However, their analysis was restricted to the comparison of items belonging to pairs of adjacent value types. They did not test a full model comprising all the 11 values they proposed with their recommended subset of items. Moreover, they tested the factorial structure of the items only in regional student samples, so their findings may only apply to relatively homogeneous, well-educated populations. The factorial structure of the items might differ in more heterogeneous, less educated samples because education influences responses in surveys (Rammstedt, Goldberg and Borg 2010).

2.2 The current research

Study 1 examines whether the Knoppen and Saris (2009a,b) findings generalize to a representative sample of the German population. We also use the results of our model tests to select a set of items from a new experimental value scale that allow modeling all values separately. Study 2 uses these items to investigate the possibility of testing the full model of ten values without unifying pairs of values as done in previous studies. In order to address the underlying estimation problem, we selected items for the full model test that were most similar to those that exhibited discriminant validity in Study 1. The new set of value items comes from an experimental scale currently under development by Shalom Schwartz.

3 Study 1: Replication of Knoppen and Saris (2009a,b)

3.1 Sample

We reanalyzed data from a German, representative national sample (Hinz et al. 2005) collected in 2001 by the commercial vendor USUMA. Respondents were selected using the random route technique and interviewed in their households. The response rate was 65%. We included only native German speakers who responded to all PVQ items. Our sample consisted of 1,966 respondents (53% female) aged 14 to 95 ($\bar{x}$ = 48.0, SD = 17.7).

3.2 Instruments

Values were assessed by use of a German version of the 40-item Portrait Values Questionnaire (Schwartz et al. 2001). Each value is measured by between three and six items, depending on its conceptual breadth. Items contain two sentences describing a person: One sentence describes how important a particular value is for a person (e.g., "Having a good time is very important to him"). The second describes the person's striving or desire for that value (e.g., "He really wants to enjoy life"). Respondents judge how similar the person in the portrait is to themselves. Hinz et al. (2005) used a modified six-point response scale ranging from "very similar" (1) to "very dissimilar" (6) (cf. Schmidt et al. 2007). For a list of the 40 items in English, see Cieciuch and Davidov this issue.

3.3 Statistical analyses

Confirmatory factor analyses were performed with M*plus* 5.0 (Muthén and Muthén, 1998–2007) on the basis of person-centered scores. To scale the latent variables, factor variance in each model was fixed to 1.0 (Byrne 2001:38; Kline 2005:170). Following Knoppen and Saris (2009b), we investigated the underlying measurement model of the PVQ items in three sequential steps: Step 1 assessed model fit and correlated errors which would indicate the presence of sub-dimensions.[3] Where present, we revised accordingly by modeling the sub-factors in the CFA. We ran a separate CFA for each value including all items intended to index it. Step 2 addressed the issue of discriminant validity by testing models for each pair of adjacent values in order to detect cross-loadings. Where found, the models were modified as needed. Step 3 tested the modified models of adjacent values (simultaneous confirmatory factor analysis). We sought a simple (factor) structure in which each item loads strongly on one and only one factor. At each step in the analyses, we considered any substantial modifications that were suggested by Jrule (van der Veld, Saris and Sartorra 2009).[4] Instead of simply drawing on M*plus* modification indices (MI) for model evaluation, Jrule also takes the expected parameter change as well as the power of the MI test into account (Saris, Satorra and van der Veld 2009). Finally, we compared the outcomes of our analysis with those reported by Knoppen and Saris (2009b).

3.4 Results

The following section summarizes results of the three-step analyses described above for each pair of adjacent values.[5] We elaborate the example of power and achievement values and present the other pairings more briefly. Table 1 provides a detailed overview of both the model fit indices in Steps 2 and 3 and information on improvements of model fit.

3.4.1 Power and Achievement. Figures 2 to 4 illustrate the three-step analyses described above. Figure 2 shows the results of the two separate CFAs for the measurement models of achievement and power (Step 1). Jrule revealed that the four achievement items form two highly correlated pairs, suggesting two sub-dimensions. Hence, we distinguished two correlated sub-dimensions of achievement (r = .75) replicating the Knoppen and Saris (2009b) discrimination of separate achievement factors. We label them 'showing success' and 'ambition'.

[3] In the following sections, we will use the term "sub-dimension" when referring to the concept and "sub-factors" when referring to factors in the CFA model and to statistics.

[4] Jrule software by Daniel Oberski available under: http://github.com/daob/JruleMplus/wiki.

[5] More detailed information on the CFA analyses is available upon request (email: constanze.beierlein@gew.de).

Table 1: Fit measures for the different confirmatory factor analyses models of Step 2 and Step 3 with regard to pairs of adjacent values

Pairwise comparison	Model fit – Step 2	Model fit – Step 3	△ CFI
Achievement & Power	Split achievement into two subfactors: 4/13, 24/32 χ^2 = 200.19 (df = 11), CFI = .94; TLI= .89, RMSEA = .09	Added cross-loading of power item 2 on achievement χ^2 = 59.03 (df = 10), CFI = .99, TLI = .97, RMSEA = .05	+ .05
Power & Security	Correlated errors of security items: 35/14; 31/21 χ^2 = 181.30 (df = 17), CFI = .93, TLI = .88, RMSEA = .07	Dropped power item 2 and security item 31 χ^2 = 45.99 (df = 7), CFI = .97, TLI = .95, RMSEA = .05	+ .04
Conformity & Security	Correlated errors of security items: 35/14; 31/21 χ^2 = 167.698 (df = 23), CFI = .90, TLI = .84, RMSEA = .06	Added cross-loading of security item 21 on conformity χ^2 = 136.55 (df = 22), CFI = .92, TLI = .87, RMSEA = .05	+ .02
Tradition & Conformity	χ^2 = 188.98 (df = 19), CFI = .91, TLI = .86, RMSEA = .07	Retained only items 20 and 25 for tradition; added correlated errors for conformity items 7/28, 16/36 χ^2 = 40.11 (df = 6), CFI = .96, TLI = .90, RMSEA = .05	+ .05
Benevolence & Tradition	χ^2 = 104.68 (df = 19), CFI = .92, TLI = .89, RMSEA = .05	Excluded benevolence item 33; added cross-loading of 20 on benevolence χ^2 = 51.23 (df = 12), CFI = .96, TLI = .93, RMSEA = .04	+ .04
Universalism & Benevolence	Split universalism into two subfactors: 3/8/23/29, 19/40 χ^2 = 161.60 (df = 32), CFI = .93, TLI= .90, RMSEA = .05	Excluded universalism item 8 and benevolence item 33 χ^2 = 106.85 (df = 17), CFI = .94, TLI = .91, RMSEA = .05	+ .01
Self-Direction & Universalism	Split universalism into two subfactors: 3/8/23/29, 19/40 χ^2 = 207.56 (df = 24), CFI = .87, TLI = .81, RMSEA = .07	Excluded universalism item 8; split self-direction split into two sub-factors: 11/22, 11/34 χ^2 = 112.45 (df = 23), CFI = .94, TLI = .90, RMSEA = .04	+ .07
Stimulation & Self-Direction	Split self-direction into two subfactors: 11/22, 11/34 χ^2 = 136.09 (df = 11), CFI = .92, TLI = .84, RMSEA = .05	Added cross-loading of stimulation item 6 on self-direction χ^2 = 57.41 (df = 10), CFI = .97, TLI = .93, RMSEA = .04	+ .05
Hedonism & Stimulation	χ^2 = 39.22 (df = 8), CFI = .98, TLI = .96, RMSEA = .05	Excluded stimulation item 6 χ^2 = 4.95 (df = 4), CFI = .99, TLI = .99, RMSEA = .01	+ .01
Achievement & Hedonism	χ^2 = 89.72 (df = 13), CFI = .96, TLI = .93, RMSEA = .06	Split achievement into two sub-factors: 4/13, 24/32 χ^2 = 25.02 (df = 11), CFI = .99, TLI = .99, RMSEA = .03	+ .03

Note. Unless otherwise noted, the two-factor model tested in Step 2 encompasses all PVQ-items for the two values. In case of modifications, specifications are summarized. Model modifications for Step 3 are based on M*plus* and Jrule outcomes (Jrule settings: Power ≥.80, alpha error ≤.05).
△ CFI quantifies the improvement when the Comparative Fit Indices of the two models are compared (see Chen 2007).

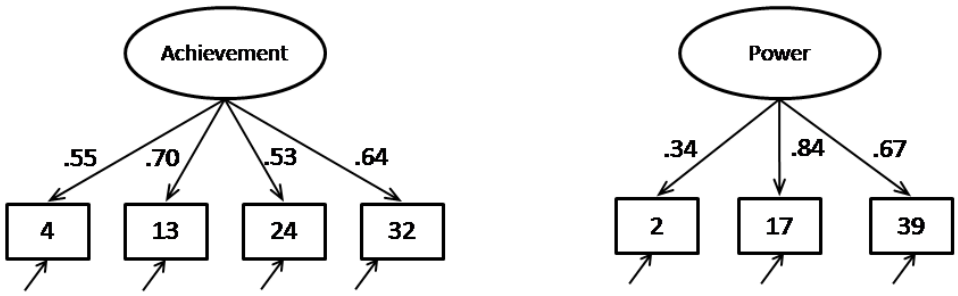

Figure 2. M*plus* model parameters for the two separate single-factor models of the adjacent values achievement (CFI = .95, TLI = .84, χ^2= 76.08, df = 2, RMSEA = .14) and power (just-identified). Standardized path coefficients are reported.

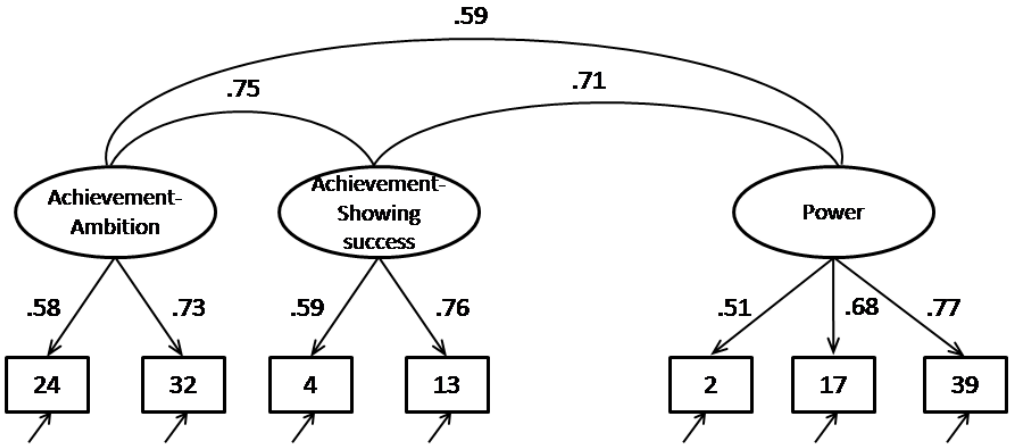

Figure 3. M*plus* model parameters for the model of the adjacent values achievement with two sub-dimensions ("ambition", "showing success") and power (CFI = .94, TLI = .89, χ^2 = 200.12, df = 11, RMSEA = .09). Standardized path coefficients are reported.

Figure 3 presents Step 2 which compared the measurement models of the power and achievement values. We introduced covariances for the three factors because they should be correlated according to the theory (Schwartz 1992). The 'showing success' sub-dimension of achievement shared more common variance with the power factor than the 'ambition' subdimension did (50% vs. 35%). Jrule also suggested a secondary loading of item 2 'being rich' on 'achievement-showing success'.

Step 3 tested a revised model in which power item 2 was allowed a cross-loading on 'achievement-showing success' (see Figure 4). This modified model yielded a good fit in the CFA (Table 1, first panel). The substantial cross-loading of power item 2 (λ = .45) suggests that it taps an aspect of achievement rather than of power. The cross-loading of this item had contributed to the common variance of the original power and achievement indices.

3.4.2 Security and Power. For the single factor model for security Jrule pointed to correlated errors for two pairs of items. One pair concerned 'national security', the other 'personal security'. Step 2 tested the two-factor model including the power items (Table 1, panel 2). Surprisingly, the security and power factors correlated negatively (r = -.74). Jrule indicated that security item 31 (avoiding illness) loaded equally on security and on power. Thus, we did not consider this item in further analyses. In Step 3, even after dropping power item 2 and security item 31, the factor correlation remained strongly negative (r = -.65).

3.4.3 Conformity and Security. The model fit for the single factor model of the conformity items was poor on some indices. Jrule identified no potential misspecifications and the four items had comparable intercorrelations. This contradicts the Knoppen and Saris (2009b) finding of two subdimensions of conformity. The two-factor models in Steps 2 and 3 exhibited satisfactory fit (Table 1, panel 3); one includes the cross-loading of security item 21 on conformity proposed by Jrule. As expected, the conformity and security factors correlated positively (r = .66).

3.4.4 Tradition and Conformity. The single factor model of the tradition items exhibited a good fit. Step 2 revealed a problematic fit for the two factor tradition/conformity model (Table 1, panel 3) with crossloadings for conformity items 16 (behave properly) and 7 (follow rules). The average intercorrelations of items within the factors were rather low while the two factors were highly correlated and difficult to

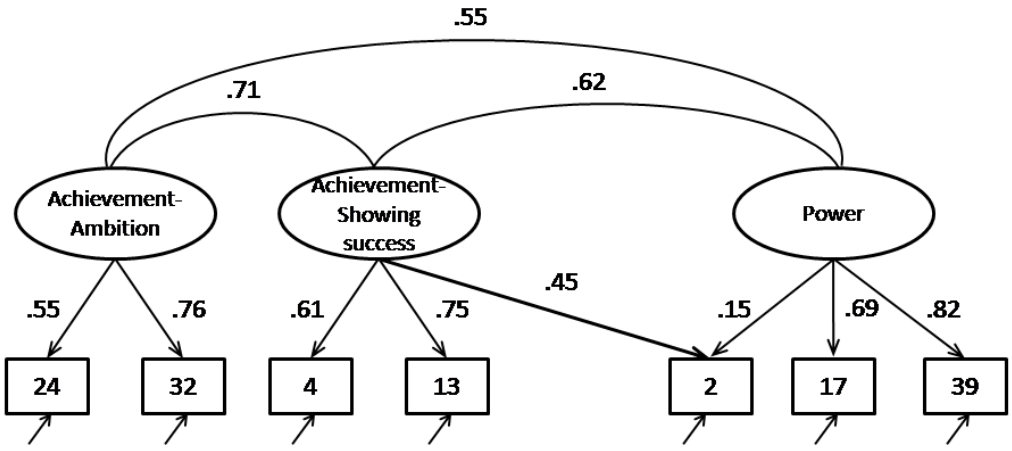

Figure 4. M*plus* model parameters for the model of the adjacent values achievement with two factors ("ambition", "showing success") and power (CFI = .99, TLI = .97, χ^2 = 59.03, df = 10, RMSEA = .05), with a cross-loading for item 2 on "achievement – showing success". Standardized path coefficients are reported.

separate. Confronting the same problem, Knoppen and Saris (2009b) made two recommendations: use only items 20 (being religious) and 25 (traditional ways) to measure tradition and split conformity into two sub-dimensions (obedience and 'not irritating others'). Adopting these proposals in Step 3 yielded a somewhat improved model fit. However, the two value factors were still substantially correlated (r = .76). Interestingly, the conformity sub-factors correlated negatively (r = -.21). This supports the Knoppen and Saris (2009b) suggestion that the sub-dimensions measure different aspects of conformity.

3.4.5 Benevolence and Tradition. The single factor model of the benevolence items yielded a good fit. Step 2 tested a two-factor model that included all four benevolence and all four tradition items. This model yielded an adequate fit (Table 1, panel 5) and a low factor intercorrelation. Knoppen and Saris (2009b) recommended dropping benevolence item 33 (forgiving) and adding a cross-loading of tradition item 20 (religious) on benevolence. In Step 3, the cross-loading of item 20 turned out to be negative rather than positive as in Knoppen and Saris (2009b). Dropping item 33 improved the model fit.

3.4.6 Universalism and Benevolence. Step 1 tested a model with two universalism sub-dimensions (cf. Schwartz 1992; Schwartz and Boehnke 2004): 'societal concern' and 'protecting the environment'. This model yielded a good fit. The model in Step 2 (Table 1, panel 6), including two sub-factors for universalism and one for benevolence, revealed that item 8 (listen to people different from him) had a substantial cross-loading on benevolence. This replicated the Knoppen and Saris (2009b) findings. The Step 3 model dropped 8 and 33 (see above) and included two factors for universalism (3/23/29 for 'societal concern' and 19/40 for 'protecting the environment'). The model fit improved only slightly. The two universalism factors were strongly associated (r = .70), but the 'societal concern' factor shared approximately as much variance with benevolence as with the 'protecting the environment' universalism factor (r = .69).

3.4.7 Self-Direction and Universalism. The single factor self-direction model including four items yielded a poor fit. Jrule results pointed to correlated errors between items 1 and 22 and between items 11 and 34 as was the case in Knoppen and Saris analysis (2009b). The fit of the two factor universalism/selfdirection model in Step 2 was also problematic (Table 1, panel 7). In Step 3, we further split self-direction into two sub-dimensions, 'intellectual openness' (1/22) and 'independence' (11/34) and dropped universalism item 8. This substantially improved the fit. The 'universalism-societal concern' sub-factor correlated negatively with the 'self-directionintellectual openness' (r = -.38) and near zero with the 'self-direction-being independent' (r = -.03) sub-factors. The 'universalism-protecting the environment' sub-factor correlated with the self-direction sub-factors in a similar manner.

3.4.8 Stimulation and Self-Direction. In a single-factor model for the three stimulation items, all of the item loadings exceeded λ = .40 (Step 1). The two factor stimulation/self-direction model in Step 2, splitting self-direction into sub-factors, had a poor fit (Table 1, panel 8). Stimulation item 6 (trying different things) had a substantial cross-loading on 'self-direction-intellectual openness' as in Knoppen and Saris (2009b). Step 3 added this cross-loading. The model fit improved and 6 loaded equally on both factors (λ = .30 and .33).

3.4.9 Hedonism and Stimulation. In a single-factor model for the three hedonism items, all of the item loadings exceeded .40 (Step 1). The two factor stimulation/hedonism model yielded a good fit (Step 2, Table 1, panel 9). The two value factors correlated moderately (r = .51). Dropping stimulation item 6 (Step 3), because of its cross-loading on self-direction noted above, slightly improved the model fit and

hardly affected the correlation between the two value factors (r = .48).

3.4.10 Achievement and Hedonism. We reported the single factor models of Step 1 for both these values above. Step 2 tested a model with one factor for each value (Table 1, panel 10) and yielded an adequate fit. Step 3 tested a model with the two achievement sub-factors ('showing success' and 'ambition'). This model improved the fit slightly. Consistent with the locations of the two values in the value circle of the theory, hedonism correlated moderately with both subdimensions of achievement (r = .29 and .28).

3.5 Discussion

Study 1 followed the procedures of Knoppen and Saris (2009a,b) to investigate the measurement models of the ten values as measured by the 40-item PVQ. Our findings in a representative German national sample largely support those that Knoppen and Saris reported for homogeneous samples of German students. But there were notable exceptions, too. Our findings, like theirs, revealed problems with the factorial structure of some of the values.

Here we summarize the main outcomes of Study 1. The analyses suggest that four of the single values distinguished in the Schwartz' theory of basic human values are better viewed as consisting of two distinguishable sub-dimensions each. Similar sub-dimensions emerged in the Knoppen and Saris (2009b) analyses. We label some of them differently based on our reading of their contents: 1) Achievement: 'showing success' and 'ambition'; 2) Security: 'national security' and 'personal security'; 3) Universalism: 'protecting the environment' and 'societal concern'; 4) Self-Direction: 'intellectual openness' and 'independence'.

Contrary to Knoppen and Saris (2009b), we did not detect sub-factors for the *conformity* value. However, the relatively low loadings of items on the conformity factor suggest that the items are diverse. A more homogeneous set of items would enhance the discriminant validity of the conformity measure.

For the *tradition* value, our analyses did not detect the subsets of items with high intercorrelations found by Knoppen and Saris (2009b). Moreover, item 20 (being religious) had a slight cross-loading on benevolence that is problematic for modeling the factorial structure of this value. Knoppen and Saris recommend only item 25 as an indicator for tradition. It is not clear from our analyses how best to measure the possibly diverse components of tradition with the PVQ.

Except for item 33 (forgiving), all items seem appropriate to measure *benevolence*. Except for item 6, all of the current PVQ *stimulation* items can be considered appropriate indicators for this construct. Except for item 2 (wealth), the other two items are appropriate indicators of power. All three items are appropriate to measure *hedonism*.

To conclude, the results of the current study, based on data of a representative sample of the German population, are to a large extent compatible with those of Knoppen and Saris (2009a,b). Estimation problems reported in earlier studies may be reducible to misspecifications due to the selection of heterogeneous items. Moreover, when using the full PVQ, several sub-dimensions of some of the ten values can be discriminated. The lack of discriminant validity found for some of the PVQ items clarifies the need for a revised assessment tool if researchers wish to model the factorial structure of the ten values.

4 Study 2: Testing the Full Model Using a Different Set of Items

Knoppen and Saris' (2009c) recommended a list of 22 PVQ items to measure 11 values, including the split of universalism into two distinct values.[6] However, our Study 1 analyses suggested that, for some values, it is problematic to find two PVQ items that are likely to yield discriminant validity. In particular, low item loadings, cross-loadings, or heterogeneity of the items measuring the security, tradition, and conformity values implied that there might only be a single valid indicator for each of these values. Hence, it was not feasible to use multiple indicators for each value to control for measurement error and test the full model with items from the current PVQ. Therefore, Study 2 tested the full model using a new set of value items taken from an experimental scale (PVQ5X) currently under development by Shalom Schwartz.[7]

4.1 Sample

The sample consisted of n = 325 students (77% females) at Goethe-University at Frankfurt, Germany (Age: $\bar{x}$ = 23.40, SD = 5.0, Range: 18 to 47 years) enrolled in four different classes in Educational Psychology. They answered the questionnaire during class sessions and received a candy as a reward.[8]

4.2 Instruments

The PVQ5X was translated into German by Eldad Davidov and colleagues at the University of Zurich, Switzerland.[9] The format of its items differs from the format of the original PVQ in only one respect. Each item consists of a single descriptive single sentence rather than two sentences. Respondents rate how similar the person described is to them on a 6-point response scale ranging from "not like me at all" (1) to "very much like me" (6). The items in the PVQ5X were

[6] We included the 'protecting the environment' sub-dimension of universalism as an 11th value because it was suggested in the initial theorizing (Schwartz 1992) and confirmed empirically in analyses across many samples (Schwartz and Boehnke 2004).

[7] The full PVQ5X scale is available from Shalom.Schwartz@mscc.huji.ac.il.

[8] 28% of respondents overlooked item 16 (achievement item 3) due to a layout mistake. We treated this as a random mistake (Kline 2005) and used statistical procedures that allow analyzing data with missing cases (e.g., FIML-Algorithm in SEM, see Schafer and Graham 2002).

[9] We thank Franz Neuberger, Mara Todisco, Rosalina Latcheva, Georg Datler, and Vanita Matta for their support in the translation of the items from English to German.

Table 2: Sample items from the PVQ5X (male version) used to measure the 11 value factors proposed by Knoppen and Saris (2009a)

	Value	Sample item
1.	Universalism-Concern	He thinks that it is important that every person in the world have equal opportunities in life.
2.	Universalism-Nature	He strongly believes that he should care for nature.
3.	Benevolence	It's very important to him to help the people dear to him.
4.	Tradition	It is important to him to maintain traditional values or beliefs.
5.	Conformity	He thinks he should always do what people in authority say.
6.	Security	He avoids anything that might endanger his safety.
7.	Power	He wants people to do what he says.
8.	Achievement	Being very successful is important to him.
9.	Hedonism	Having a good time is important to him.
10.	Stimulation	He thinks it is important to have all sorts of new experiences.
11.	Self-Direction	Freedom to choose what he does is important to him.

Note. Each value was assessed by three items in study 2.

Table 3: Descriptive statistics and construct reliability (H) of the newly developed items (n ≥ 234)

	Descriptive Statistics				Reliability
Items	$\bar{x}$ [1]	SD	Skewness	Kurtosis	H[2]
Universalism-Environment 1	3.61	1.32	0.17	-0.78	.90
Universalism-Environment 2	3.42	1.25	0.23	-0.77	
Universalism-Environment 3	3.87	1.22	0.00	-0.66	
Universalism-Societal Concern 1	4.70	1.01	-0.58	0.02	.98
Universalism-Societal Concern 2	4.83	1.11	-0.88	0.18	
Universalism-Societal Concern 3	4.62	1.11	-0.65	-0.03	
Benevolence 1	5.58	0.64	-1.47	1.90	.52
Benevolence 2	5.44	0.67	-0.98	0.54	
Benevolence 3	5.02	0.86	-0.79	0.48	
Tradition 1	3.38	1.44	0.33	-0.86	.90
Tradition 2	3.54	1.59	0.04	-1.09	
Tradition 3	3.37	1.39	0.32	-0.80	
Conformity 1	4.90	1.11	-1.18	1.52	.71
Conformity 2	3.87	1.28	-0.11	-0.71	
Conformity 3	4.20	1.24	-0.60	-0.13	
Security 1	3.34	1.32	0.32	-0.71	.83
Security 2	4.02	1.27	-0.25	-0.68	
Security 3	4.50	1.06	-0.41	-0.44	
Power 1	3.21	1.22	0.29	-0.65	.93
Power 2	2.36	1.06	0.97	0.89	
Power 3	2.53	1.10	0.69	0.02	
Achievement 1	4.33	1.27	-0.70	-0.03	.65
Achievement 2	3.60	1.18	-0.04	-0.46	
Achievement 3	3.54	1.17	0.01	-0.73	
Hedonism 1	5.34	0.82	-1.44	2.62	.81
Hedonism 2	5.19	0.93	-1.13	0.84	
Hedonism 3	3.90	1.26	-0.27	-0.45	
Stimulation 1	4.31	1.22	-0.39	-0.69	.70
Stimulation 2	4.10	1.27	-0.30	-0.74	
Stimulation 3	4.74	1.01	-0.68	0.02	
Self-Direction 1	5.16	0.80	-1.06	2.26	.62
Self-Direction 2	4.39	1.14	-0.51	-0.10	
Self-Direction 3	5.20	0.90	-1.16	1.20	

[1] Response scale ranging from (1) "not like me at all" to (6) "very much like me".

[2] Hancock's H was calculated separately for each value based on the partial regression coefficients of the single factor CFA.

Table 4: Range of the standardized regression coefficients for the modified full model of 11 values (N = 325)

	Factors										
Items	UNE	UNC	BE	TR	CO	SE	PO	AC	HE	ST	SD
Universalism-Environment	.76 -.91										
Universalism-Societal Concern		.58 -.81									
Benevolence			.61 -.78								
Tradition				.81 -.89							
Conformity					.65 -.69						
Security						.55 -.87					
Power							.59 -.91				
Achievement								.51 -.77			
Hedonism									.79 -.83		
Stimulation										.51 -.76	
Self-Direction											.55 -.62

Note. UNE = Universalism-Environment; UNC = Universalism-Societal Concern; BE = Benevolence; TR = Tradition; CO = Conformity; SE = Security; PO = Power; AC = Achievement; HE = Hedonism; ST = Stimulation; SD = Self-Direction. All coefficients are significant, $p < .05$.

taken from sentences in the original PVQ or constructed to represent sub-facets of the ten basic values more precisely.

Both Study 1 and Knoppen and Saris (2009b) identified two sub-dimensions each for the achievement, security, self-direction, and universalism values. Knoppen and Saris (2009c) recommended measuring these basic values, except universalism, using only one of their subdimensions. Therefore, we chose items from the PVQ5X to operationalize each of the subdimensions they selected. For achievement, we chose 'ambition' items; for security, we chose 'personal security' items; for self-direction, we chose 'independence' items. For universalism, we included both 'protecting the environment' and 'societal concern' items. Although Study 1 did not find sub-dimensions of conformity values, we followed Knoppen and Saris' (2009c) recommendation to choose items that represent the 'obedience' aspect of this value. Table 2 provides exemplary items used in Study 2 to measure the 11 values. We included three items as indicators for each value.

4.3 Statistical analyses

We tested the full model using CFA performed with M*plus* 5.0 (Muthén and Muthén, 1998–2007). In accordance with our analytical strategy in Study 1, we fixed the factor variance in each model to 1.0 in order to scale the latent variables.

4.4 Results

Table 3 summarizes the descriptive statistics for each of the 33 value items (normative scores). It also reports Hancock's H measure of reliability (Hancock and Mueller 2001) which reflects the minimum achievable or anticipated construct reliability associated with the measurement model of each factor (Hancock and Mueller 2006:90). Hancock's H requires fewer assumptions than Cronbach's α (e.g., essential τ-equivalence is not required for factor loadings; cf. Cortina 1993). In our study, Hancock's H ranged from .52 (benevolence) to .98 (universalism – societal concern). Seven of the measurement models for the values achieved or exceeded the recommended minimum of $H = .70$. The item skewness and kurtosis suggested no major violations of normality assumptions.

The confirmatory factor model including three items for each of the 11 factors yielded a low to moderate fit (χ^2 = 879.34, df = 440; CFI= .88; SRMR = .06; RMSEA = .06; PCLOSE = .05) (see Hu and Bentler 1999; Schermelleh-Engel, Moosbrugger and Muller 2003). Jrule identified several model misspecifications: Hedonism item 3 (opportunity to have fun) had a substantial cross-loading on stimulation (λ= .48). Further, achievement item 3 (admire achievements) loaded almost equally on achievement (λ = .31) and power (λ= .35) when permitted to load on both values. Jrule also detected correlated errors for security items 1 (avoid anything that endangers one's safety) and 2 (personal security is extremely important).

We dropped hedonism item 3 and achievement item 3 from the model because our aim was to select items that discriminate sufficiently between different values. In addition, we allowed correlated errors for security items 1 and 2 due to their content overlap. Model estimation terminated normally without any problems of convergence. The modified full model yielded a good fit on most indices (χ^2 = 593.00, df = 350; CFI=.93; SRMR = .05; RMSEA = .05; PCLOSE = .83). Table 4 shows the factor loadings (standardized regression coefficients) of each item for the modified full model. All item loadings exceeded λ = .40, a cut-off-criterion for applied factor analytic research (Brown 2006:130). As expected, the errors of security items 1 and 2 were moderately correlated (λ = .43).

Table 5 presents the correlations between all pairs of the 11 value factors in the modified full model. All factor intercorrelations are well below .85, indicating that there is reasonable discriminant validity between every pair of value

Table 5: Intercorrelations among 11 value factors of the modified full model

Factors	UNC	BE	TR	CO	SE	PO	AC	HE	ST	SD
Universalism – Environment	*.42*	*.27*	.02	.11	.06	.07	.12	.04	.09	*.27*
Universalism – Societal Concern		*.48*	.11	*.43*	*.30*	-.11	.02	.03	.06	*.42*
Benevolence			*.30*	*.46*	*.37*	-.10	*.22*	*.33*	*.21*	*.49*
Tradition				*.34*	*.44*	*.34*	*.27*	-.07	-.04	-.04
Conformity					*.58*	*-.23*	.10	.08	-.08	.02
Security						.06	*.34*	*.15*	.05	*.24*
Power							*.37*	*-.04*	*.16*	*.21*
Achievement								*.18*	*.27*	*.39*
Hedonism									*.57*	*.36*
Stimulation										*.52*
Self Direction										

Note. Coefficients significant at $p < .05$ are *italicized*.

factors (Brown 2006:131). Factor correlations of hedonism and stimulation in the modified model slightly decreased to r= .57 compared to r = .66 in our first full model test. Security and conformity showing the highest factor intercorrelations, r = .58.

5. Discussion and future research

In previous studies, researchers reported estimation problems when including all ten values (Schwartz 1992) simultaneously in a CFA. Solutions to remedy these problems included collapsing some values or dropping some of the items. Using German student samples, Knoppen and Saris (2009a,b) demonstrated that the problems arose because the selection strategy of items sought coverage of the conceptual breadth of each value rather than homogeneity of the items that operationalized each value. This resulted in model misspecifications due to the presence of sub-dimensions within values or due to crossloadings of items with adjacent values. Study 1 built on this research by investigating the factorial structure of the Schwartz values in PVQ data from a representative German sample. It tested all adjacent pairs of values and identified problems in the measurement model similar to though not identical with those identified by Knoppen and Saris (2009a,b).

Knoppen and Saris (2009c) recommended a subset of items that they thought might allow modeling each of the eleven values separately. Differences in the findings of our study and theirs (e.g., the presence or absence of subdimensions for conformity) indicate that some items function differently in student compared with representative samples. Testing value items in demographically heterogeneous representative samples allows selecting value items for wide use.

Study 2 used a set of value items from an experimental version of the PVQ that is undergoing testing in ten countries. It tested the factorial structure of 11 values, using subdimensions of four values that had emerged in Study 1 and in Knoppen and Saris (2009b). For each value, we chose three items. Some of these items resembled those that exhibited discriminant validity in the earlier studies and the others expressed the motivation central to the value quite precisely.

The modified full model test in the German student sample in Study 2 revealed a satisfactory fit for the 11 factor model. There were no estimation problems. We therefore conclude that collapsing factors (e.g., achievement with power) is not required if appropriate indicators are used. However, we had to eliminate two items from the model that showed significant cross-loadings. Achievement item 3 (admire achievements) was associated with power rather than with achievement. In many western societies, personal success and academic performance is linked to a high social prestige. Thus, admiring a person's achievements may be regarded as an aspect of power, too. In our student sample, hedonism item 3 (opportunity to have fun) had a cross-loading on stimulation. Stimulation values refer to the pursuit of "excitement, novelty, and challenge in life" (Schwartz 1992:8). For relatively young samples, having fun may represent motivation for excitement more than for pleasure.

We note two limitations of Study 2. First, it was limited to a single German student sample. Like the Knoppen and Saris studies (2009a,b,c), the sample of Study 2 was relatively homogeneous. Future research with representative samples in multiple countries should assess whether our selection of items will discriminate all 11 values in more diversified populations.

Second, by following Knoppen and Saris (2009c) in choosing only one of the two subdimensions for achievement, security, conformity, and self-direction that they observed in their study, we did not provide a test of the measurement models of the other sub-dimensions. The original value theory (Schwartz 1992) conceptualized each of the ten values as broad constructs. Schwartz (1992, 2006) avers that his partitioning of the motivational circle of values into the ten distinct values in Figure 1 is arbitrary; it could be partitioned into a larger set of more narrowly defined values if scientifically useful. The presence of subdimensions in Knoppen and Saris (2009a,b) and in Study 1 (cf. Schwartz and Boehnke 2004) demonstrated that some more narrowly defined values can be discriminated. Future research should test whether it is possible to discriminate the full set of sub-dimensions found in these studies as well as other sub-dimensions that might be suggested by close content analyses of the ten value constructs.

Acknowledgements

The work of the third and fourth authors on this paper was supported by the Higher School of Economics (HSE) Basic Research Program (International Laboratory of Socio-Cultural Research).

References

Bardi, A., & Schwartz, S. H. (2003). Values and behavior. Strength and structure of relations. *Personality and Social Psychology Bulletin*, *29*(10), 1207-1220.

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

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Hillsdale (New Jersey): Lawrence Earlbaum Associates.

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

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104.

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. H. (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 Clercq, S., Fontaine, J. R. J., & Anseel, F. (2008). In search of a comprehensive value model for assessing supplementary person-organization fit. *Journal of Psychology*, *142*(3), 277-302.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. H. C. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: past and present. A Festschrift in honor of Karl G. Jöreskog* (p. 195-261). Chicago: Scientific Software International.

Hancock, G. R., & Mueller, R. O. (Eds.). (2006). *Structural equation modeling: A second course*. Greenwich (Connecticut): Information Age Publishing.

Hinz, A., Brähler, E., Schmidt, P., & Albani, C. (2005). Investigating the circumplex structure of the Portrait Values Questionnaire (PVQ). *Journal of Individual Differences*, *26*(4), 185-193.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jowell, R., Roberts, C., Fitzgerald, R., & Eva, G. (Eds.). (2007). *Measuring attitudes crossnationally. Lessons from the European Social Survey*. London: Sage Publications.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Knoppen, D., & Saris, W. (2009a). Do we have to combine values in the Schwartz' human values scale? A comment on the Davidov studies. *Survey Research Methods*, *3*(2), 91-103.

Knoppen, D., & Saris, W. E. (2009b). *Evaluation of the Portrait Values Questionnaire using SEM: A new ESS Proposal.* Paper presented at the QMSS2 seminar at Bolzano, Italy, June 11-12.

Knoppen, D., & Saris, W. E. (2009c). *Schwartz' theory of human values: Balancing homogeneity of reflective items and theoretical coverage.* RECSM Working Paper, 9, 1-40. Universitat Pompeu Fabra, Barcelona.

Meulemann, B., Davidov, E., Schmidt, P., & Billiet, J. (2010). *Social location and value priorities. A European-wide comparison of the relation between social-structural variables and human values.* An unpublished manuscript.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus (Version 5.1)*. Los Angeles (California): Muthén and Muthén.

Perrinjaquet, A., Furrer, O., Usunier, J.-C., Cestre, G., & Valette-Florence, P. (2007). A test of the circumplex structure of human values. *Journal of Research in Personality*, *41*(4), 820-840.

Piurko, Y., Schwartz, S. H., & Davidov, E. (2011). Values and political orientation: A crosscountry comparison. *Political Psychology*, *32*, 537-561.

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, *44*(4), 53-61.

Saris, W., & Gallhofer, I. (2002). *Report on the MTMM experiments in the pilot studies and proposals for round 1 of the ESS.* Report to the Central Co-ordinating Team of the European Social Survey, London.

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.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Schermelleh-Engel, K., Moosbrugger, H., & Mller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research – Online*, *8*(2), 23-74.

Schmidt, P., Bamberg, S., Davidov, E., Herrmann, J., & Schwartz, S. H. (2007). Die Messung von Werten mit dem "Portraits Value Questionnaire" [Assessing values using the Portrait Value Questionnaire]. *Zeitschrift für Sozialpsychologie*, *38*(4), 261-275.

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. (2003). *A proposal for measuring value orientations across nations.* ESS Questionnaire Development Report (Chapter 7). Retrieved from: http://www.europeansocialsurvey.org.

Schwartz, S. H. (2006). Value orientations: Measurement, antecedents and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally – lessons from the European Social Survey* (p. 169-204). London: Sage.

Schwartz, S. H., & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, *38*(3), 230-255.

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.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Schwartz, S. H., & Wieczorek, S. (2009). Testing Measurement Invariance using Multigroup CFA: Differences between Educational Groups in Human Values Measurement. *Quality and Quantity*, *43*(4), 599-616.

Van der Veld, W., Saris, W., & Satorra, A. (2009). *Jrule, Judgement Rule Aid for Structural equation models.* Available under vdveld@telfort.nl.

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