Survey Research Methods (2013)
Vol.7, No.1, pp. 29-44
ISSN 1864-3361
http://www.surveymethods.org

Operationalizing the Theory of Human Values: Balancing Homogeneity of Reflective Items and Theoretical Coverage

Willem E. Saris
Research and Expertise Centre for Survey Methodology,
Universitat Pompeu Fabra

Desirée Knoppen
EADA Business School, Barcelona

Shalom H. Schwartz
The Hebrew University of Jerusalem, Israel and
National Research University-Higher School of Economics, Moscow

Schwartz's theory of human values, as operationalized using different instruments such as the Portrait Values Questionnaire (PVQ), was confirmed by multiple studies using Smallest Space Analysis (SSA). Because of its success, a short version of the PVQ was introduced in the European Social Survey (ESS). However, initial tests using Confirmatory Factor Analysis (CFA) pointed to low discriminant validity of the 10 basic values: The correlations between values next to each other in the two-dimensional space described by SSA were close to or greater than 1. In response, one research stream suggested combining the factors with low discriminant validity. Another stream suggested that the problem was not low discriminant validity but rather misspecifications in the model. Analyses of the short Portrait Values Questionnaire of the ESS confirmed the latter view.

This paper demonstrates that the problems of the short version of the PVQ exist in the full 40-item PVQ as well. Based on SEM analyses of the items of the full PVQ, we propose that it can provide measures of 15 more narrowly defined values with good discriminant validity. Our proposal respects the conceptual complexity of the values theory while avoiding contamination of composite scores. It can be expected that the improved measurement of 15 values will increase their predictive power. The presence of some single items suggests the extension of the value theory and scales to encompass more than 15 values. Implications for further development of the scale are drawn.

Keywords: Human values; Portrait Values Questionnaire; composite scores; validity; CFA

1 Introduction

The values theory developed by Shalom Schwartz has been widely applied in the social sciences since its full publication (Schwartz 1992).[1,2] The theory defines individual values as

> "desirable, trans-situational goals, varying in importance, that serve as guiding principles in the life of a person or other social entity" (Schwartz 1994:21).

Individual values have been used to characterize individuals and social categories (e.g., Schwartz and Sagie 2000), to explore interrelations between values and background variables (e.g., Schwartz and Rubel 2005), and to predict attitudes (e.g., Sagiv and Schwartz 1995) and everyday behavior (e.g., Bardi and Schwartz 2003). This theory of values goes beyond presenting a typology of individual values to specify the structure of relations of conflict and congruence among the individual values. The theory has therefore also been used to analyze how an integrated system of values, rather than single values, relates to other variables (e.g., Schwartz 1996).

The comprehensiveness and widespread validation of the Schwartz theory of human values led to its inclusion in the European Social Survey (ESS[3]), which aims to study changing values, attitudes, attributes and behavior patterns within Europe systematically (Jowell, Kaase, Fitzgerald and Eva 2007). Due to space restrictions, Schwartz (2003) developed a 21-item version of one popular instrument – the 40 item Portrait Values Questionnaire (PVQ). However, extensive empirical testing of the ESS instrument (Davidov 2008; Davidov and Schmidt 2007; Davidov, Schmidt and Schwartz

Contact information: Willem E. Saris, Research and Expertise Centre for Survey Methodology, Department of Political and Social Sciences, Universitat Pompeu Fabra, Barcelona, e-mail: w.saris@telefonica.net

[1] This paper is a revised version of the original paper by Knoppen and Saris presented at the QMSS2 seminar in Bolzano 2009, June 11-12. Since then several studies have followed the approach presented here (Cieciuch and Schwartz 2012, Beierlein et al. 2012 and Lilleoja, 2011) but this paper presents the original approach.

[2] Participation of the third author in preparing this paper was partly supported by the HSE Basic Research Program (International Lab of Sociocultural Research).

[3] http://www.europeansocialsurvey.org

2008) indicates that six out of the 10 values that the theory specifies show low discriminant validity (i.e., the possibility of discriminating between dissimilar values, Campbell and Fiske 1959). Only four of the values (Hedonism, Stimulation, Self-direction, and Security) could be discriminated clearly. Davidov and colleagues suggested grouping the remaining six values into three pairs in order to solve the problem of a lack of discriminant validity.

More recently, Knoppen and Saris (2009a) argued that the low discriminant validity observed in the ESS data set is not an intrinsic characteristic of the values. Rather, it is due to the strategy used to select value items. That strategy maximized coverage of all aspects of the conceptual definition of each value at the cost of the homogeneity of items. The lack of homogeneity among items chosen to operationalize some values meant that some items correlated more with items from other values than with items from their own value. This led to estimated correlations between latent variables (i.e., values) that were close to or even greater than 1.00. Knoppen and Saris (2009a) suggested that an alternative choice strategy, one which balances conceptual coverage and homogeneity of items, could avoid the problem of low discriminant validity.

Heterogeneity within a reflective (rather than formative, Edwards and Bagozzi 2000) set of items is especially problematic when researchers calculate composite scores. This is because the heterogeneity of the items of the composite scores may cause contamination and reduction of the relationships with other variables compared with more specific indicators (Saris 1981, Saris and Gallhofer 2007). Analyses of the ESS instrument by Knoppen and Saris (2009a) showed that an alternative choice strategy, targeted towards a more homogeneous set of reflective items for each individual value – while respecting theoretical coverage – would avoid contamination of composite scores and lead to better prediction.

An initial evaluation of the face validity of the items selected to measure values in the original 40-item PVQ also points to several values that are reflected by items with heterogeneous meanings. Consider, for example, the three items that reflect the Power value:

- It is important to him to be rich. He wants to have a lot of money and expensive things.
- It is important to him to be in charge and tell others what to do. He wants people to do what he says.
- He always wants to be the one who makes the decisions. He likes to be the leader.

The first item refers to being rich, a possible source of power but not power itself. This item could also be interpreted as a goal people would like to achieve. It might therefore reflect the Achievement value. The last two items, in contrast, refer to having control over other people. This example illustrates the problem of heterogeneity that resulted from the choice strategy used to select the PVQ items for each value. There has been no rigorous testing of the homogeneity of the items that operationalize the values in the original PVQ, despite the numerous studies that use the PVQ to estimate relationships between values and other variables. There is thus a risk that these analyses used contaminated composite scores. This paper tests the validity of each individual value as it is measured in the PVQ.

First, we review the Schwartz theory of human values (1992, 1994) and their operationalization in the PVQ. We then suggest an alternative model for analyzing the PVQ and present the method we employ. The method builds upon Confirmatory Factor Analysis (CFA) rather than Smallest Space Analysis (SSA), the method commonly used to evaluate the PVQ. We next present empirical results for each of the 10 basic values. Finally, we draw conclusions and note the strengths and limitations of this study. The paper concludes with methodological issues and implications for further development of the measurement instrument.

2 The Schwartz Values Theory and the PVQ Instrument

Schwartz tested his initial ideas about the structure of basic values with available data from the 36-item Rokeach Value Survey (Rokeach 1973). Rokeach had suggested that values could be distinguished by their goal type into terminal (desirable end-states of existence such as security and wisdom) and instrumental values (desirable modes of behaviour such as honestly and helping). Schwartz ignored this distinction in analyzing the Rokeach data (Schwartz and Bilsky 1987; 1990), arguing that any value could be an end-state for which others are instrumental. Instead, he proposed that values could be organized according to the interests they serve – individual, collective, or mixed. Most importantly, he proposed that basic values could be organized according to their major motivational goal. He suggested that the Rokeach value items could be discriminated into seven motivationally distinct values: enjoyment, achievement, restrictive conformity, security, pro-social, maturity, and self-direction. Applying SSA, both Schwartz and Bilsky studies confirmed that these values were ordered in a circular arrangement in the order listed here. The authors suggested the need to refine these values and to identify additional, motivationally distinct values.

The full theory of basic values presented in 1992 specified ten motivationally distinct, basic individual values that are presumably recognized in all cultural groups (Schwartz 1992). Schwartz derived these ten values from three universal requirements of human biological and social functioning: (1) needs of individuals as biological organisms ("organism"), (2) requisites of coordinated social interaction ("interaction"), and (3) requirements for the smooth functioning and survival of groups ("group"). Each of the ten values was grounded in one or more of these three universal requirements, as shown in Table 1.

Schwartz tested his theory with a new instrument designed explicitly to measure the ten redefined and relabeled values. The last column in Table 1 lists the single value items from the initial 56-item instrument developed to operationalize the ten values (Schwartz 1992). Schwartz speculated that the set of 10 basic value types might be exhaustive:

> "It is possible to classify virtually all the items found in (existing) lists of specific values from

Table 1: The Ten Basic Values, Sources in which they are Grounded, and Specific Items that Operationalize them (from Schwartz 1994)

Values and their Definitions	Sources	Specific Items from the 56-item Instrument
Benevolence: Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.	organism, interaction, group	helpful, honest, forgiving, loyal, responsible
Universalism: Understanding, appreciation, tolerance and protection for the welfare of *all* people and for nature.	organism, group	social justice, equality, world at peace, protecting the environment, unity with nature, world of beauty, broadminded, wisdom
Self-direction: Independent thought and action-choosing, creating, exploring.	organism, interaction	creativity, curious, freedom, choosing own goals, independent
Stimulation: Excitement, novelty and challenge in life	organism	exciting life, varied life, daring
Hedonism: Pleasure and sensuous gratification for oneself.	organism	pleasure, enjoying life
Achievement: Personal success through demonstrating competence according to social standards.	interaction, group	ambitious, successful, capable, influential
Power: Social status and prestige, control or dominance over people and resources	interaction, group	social power, authority, wealth, preserving my public image
Security: Safety, harmony and stability of society, of relationships and of self.	organism, interaction, group	national security, family security, clean, social order, reciprocation of favors
Conformity: Restraint of actions, inclinations and impulses likely to upset or harm others and violate social expectations or norms.	interaction, group	obedient, honoring elders, self discipline, politeness
Tradition: Respect, commitment and acceptance of the customs and ideas that traditional culture or religion provide.	group	respect for tradition, moderate, devout, humble, accepting my portion in life

> different cultures [. . .] into one of these ten motivational types of values" (1994:22-23).

The key feature of the theory of ten basic values is the idea that the values form the circular motivational continuum presented in Figure 1. This theorized circular structure was confirmed initially in samples from 20 countries, using SSA. In addition to the organization of the values by individual versus collective interests, Schwartz noted that they could be described as lying on two orthogonal axes, self-enhancement (power and achievement) versus self-transcendence (universalism and benevolence) and openness to change (self-direction, stimulation, and hedonism) versus conservation (security, conformity, and tradition). The circular structure and two axes identify the conflicting and congruent motivations among the ten basic values. The closer any two values are in either direction around the circle, the more positive the conceptual and empirical association between them; the more distant they are, the less positive their association. If one value is theorized to relate positively to an attitude, behavior, or personal characteristic, its adjacent values should also relate positively to that variable. At the same time, the opposing values in the circle should relate less positively or even negatively to that variable. Substantial research has confirmed this general hypothesis (see Schwartz 2006, for a summary).

In order to understand the proposed circular structure in depth, it is necessary to mention the method for arriving at the structure. Smallest Space Analysis (SSA) is a non-metric technique for mapping items as points in a multi-dimensional space, such that the distance between points reflects the interrelations between items. The greater the conceptual similarity between items, the more related they should be empirically, and hence the closer their locations should be in the multi-dimensional space (Guttman 1968). Schwartz noted the following implications of his theory that SSA can represent:

> "Because values form a motivational continuum, the decisions about exact boundaries are arbitrary. Items near the boundaries of adjacent values inevitably overlap somewhat in meaning. Consequently, in analyses in many samples, value items from adjacent types of values intermix rather then emerge in clearly distinct regions" (Schwartz 2006:942-943).

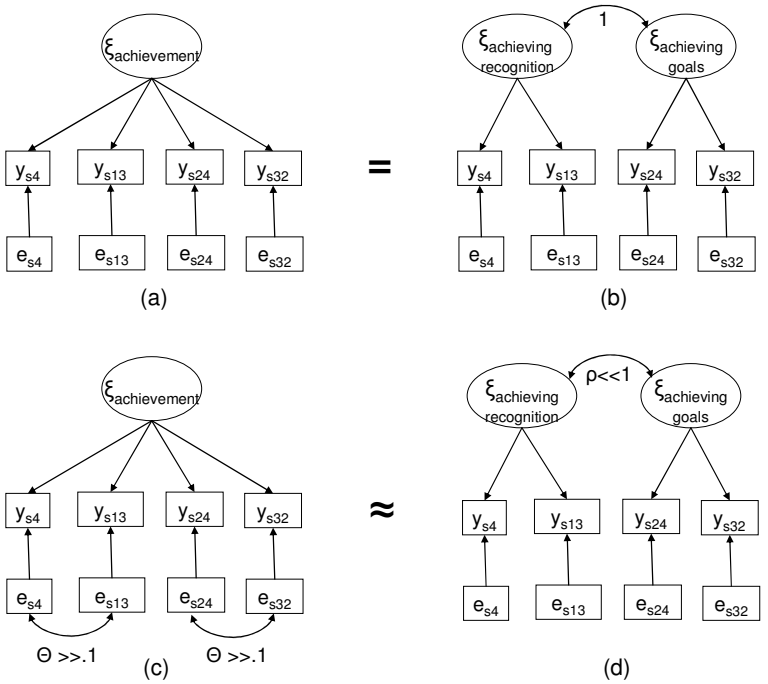

Figure 1. The Structure of Relations among 10 Basic Values according to the Schwartz Value Theory

Furthermore,

> "One could reasonably partition the domain of value items into broader or more fine-tuned distinct value constructs, depending on how finely one wishes to discriminate among motivations" (Davidov, Schmidt, and Schwartz 2008:424).

We will return to these implications below.

2.1 The Portrait Values Questionnaire

The Schwartz values model has been operationalized in various ways. The first instrument, now known as the Schwartz Value Survey (SVS), included 56 (later 57) items (Schwartz 1992). The Portrait Values Questionnaire (PVQ) aimed to reduce the cognitive complexity of the SVS (Schwartz 2006; Schwartz, Melech, Lehmann, Burgess, Harris and Owens 2001). It presents respondents with short verbal portraits of different people in terms of their goals, aspirations, or wishes that point implicitly to the importance of a single value. The portrait is drawn in two sentences. One sentence refers to importance: It is (very) important to him/her to [have an exciting life]. The other sentence refers to an aspiration or wish: He/she likes (or wants or thinks or believes or seeks) [surprises]. For each portrait, respondents answer, "How much like you is this person?" on a 6-point scale with the categories "very much like me, like me, somewhat like me, a little like me, not like me, not like me at all." This comparison focuses them on the specific values rather than on their whole self-concept (Schwartz 2007). Respondents own values are inferred from their self-reported similarity to people described implicitly in terms of their values. The appendix presents the 40-item PVQ.

In developing the PVQ and the shorter ESS instrument, Schwartz sought to construct items for each basic value that would cover all the substantive components in the definition of that broad value. Thus, if a value included diverse components, the distance between the points that represented these components in the two-dimensional space for that value might be quite great. For example, the Universalism value included three related diverse components – tolerance, social concern, and concern for nature – which were expected to form a single, wide region in the space (Schwartz 2006). The decision to consider Universalism a single value with three components or to split it into three values when drawing boundaries in the SSA space was necessarily subjective and arbitrary.

2.2 An Alternative Model for the Measures of the PVQ

Another way of conceptualizing and assessing the aspect of the value theory that specifies a typology of ten values and the PVQ method of measurement is to think of this as a factor model which specifies the items that operationalize each factor (i.e., value). Let us illustrate this approach for two values: Power and Achievement. In this formulation, three items are indicators of the Power factor and four items are indicators of the Achievement factor (see appendix for the specific items). For these two values, the factor model would be the simple factor structure presented in Figure 2.

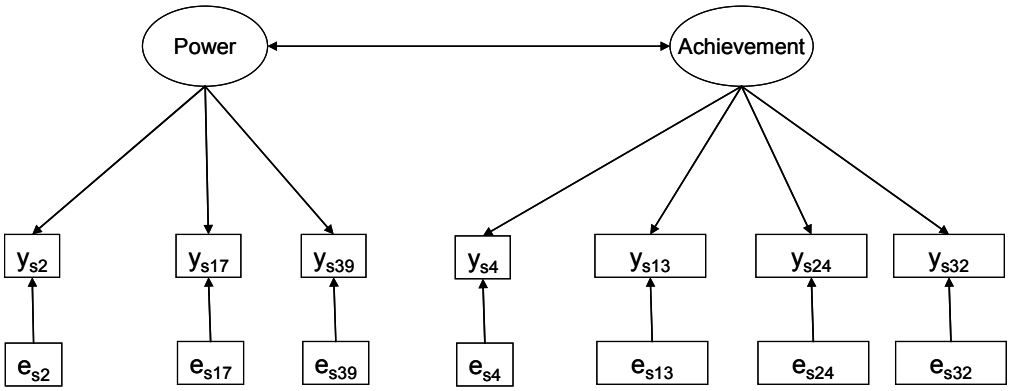

Figure 2. The simple factor structure of Power and Achievement values. The symbol y refers to observed variables; e refers to measurement errors; s_i refers to item numbers

This model could be extended to a ten-factor model in which the factors are the values specified in the Schwartz theory and the indicators for each factor are the items intended to operationalize it in the measuring instrument (e.g., the PVQ in appendix).

As noted, the strategy for selecting items maximized the theoretical coverage at the expense of the homogeneity of the items. We hypothesized that the lack of homogeneity among the selected items is what causes some items to correlate more with items from other values than with items from the same value. These items would consequently fit better in the other value set than in their own set. For example, the "being rich" item may loading more highly on the Achievement factor than on the Power factor. If so, the simple structure factor model would be misspecified and a cross-loading from Achievement to the "being rich" item would be required.

Another type of misspecification that we anticipate is that the lack of homogeneity within the set of items for one value may be so large that the factor splits into two or more sub-factors. That is, a two-factor structure would describe the correlation matrix better than a one-factor structure. This might occur if at least two sets of items that operationalize a value are quite conceptually distinct, as in the example of Universalism that included three related diverse components. In this case, the simple structure, single factor model would be misspecified because there would be correlated errors between items which are more similar to each other than they are to the other items for that value.

In the next sections, we first show that the simple structure factor model does not fit the data in many instances. We try to detect the possible misspecifications mentioned above and discuss their implications for the original value theory. However, it is first necessary to discuss how the original model can be tested and how possible misspecifications in the models for the different values can be detected.

3 Method

3.1 Structural Equation Modeling

The alternative method for modeling the values theory is a confirmatory factor analysis (CFA) model (Bollen 1989) that is a specific case of Structural Equation Model (SEM). Such models can be estimated and tested using SEM software (Jöreskog 1969). It has been common practice to base the accept/reject decision on a range of statistics (e.g., CHI2, AGFI, GFI, SRMR, NFI, CFI, RMSEA). All of these have the common shortcoming of being highly dependent on the power of the test. More precisely, the standard test can only detect misspecifications for which the test is sensitive (high power). A model may therefore be rejected due to very small misspecifications and accepting a model does not necessarily mean that the model is correct.

In response to the increasingly widespread criticism of the (mis)use of test statistics of SEM, Saris, Satorra and Van der Veld (2009) developed an alternative procedure (Jrule[4]) that tests for misspecifications (i.e., relevant parameters that have been omitted from the model, or modeled parameters that are not present in the data). The misspecifications test combines knowledge of (a) the size of the misspecification (Expected Parameter Change, EPC), (b) the impact on the fit if the parameter were included (Modification Index, MI), and (c) the sensitivity of the test in detecting the misspecification (power of the test). Both (a) and (b) are present in the output files of SEM software, and (c) is calculated based on the noncentrality parameter.

Knowing the size of the EPC and the MI also provides a simple way to estimate the power of the test for the size of each misspecification. Consider a specific deviation δ for which one would like to know the power. Hence, δ would be the minimum size of the misspecification that one would like

[4]Jrule presents test statistics for all the restricted parameters, based on the output of LISREL. The program can be requested by sending an e-mail with 'JRule' in the subject line to vdveld @telfort.nl.

the test to detect with a high likelihood (power). By standard theory, under deviation from the null hypothesis, the asymptotic distribution of the MI is non-central χ^2, with the non-centrality parameter (ncp) given by

$$ncp = (MI/EPC^2)\delta^2. \quad (1)$$

This expression of the ncp is a function of statistics provided by the standard software and the user-specified value δ of maximally acceptable misspecification. This ncp can be used to determine the power of the test of a misspecification of for any value of the significance level α of the test and for all restricted parameters. The power of the test can be obtained from the tables of the non-central χ^2 distribution (or using any computer-based routine) as:

$$Prob(\chi^2(1, ncp) > c_\alpha) \quad (2)$$

where c_α is the critical value of an α-level test based on a χ^2 distribution with df=1, and χ^2(1,ncp) is the non-central chi-square distribution with non-centrality parameter ncp.

Table 2 presents four possible outcomes this approach distinguishes. These outcomes result from crossing the significance of the MI test with its power (high/low).

If the MI is significant and the power of the MI test is low there is a misspecification because the MI is significant despite the low sensitivity of the test (low power). If the MI is not significant and the power of the MI is high there is no misspecification because it would have lead to a significant MI given the high power. If the MI is low and the power of the MI test is also low, there is insufficient information to make a decision. If the MI is significant but the power of the MI test is high, there may be a serious misspecification. However, it may also be that the MI is significant due to the high sensitivity of the test for this misspecification. Therefore, in this case, the suggestion is to look at the substantive relevance of the EPC: If the EPC is rather small, there is no serious misspecification. This makes sense because, generally, one does not want to adjust our model for a standardized coefficient of .001, even though this coefficient is significant. However, if the EPC is large, for example larger than .2, one would conclude that there is a relevant misspecification in the model. In sum, Table 2 classifies the different options one may be confronted with in conducting model evaluation. The program Jrule facilitates the procedure.

Although we will mention some of the commonly used fit measures, we will rely mainly on the program Jrule to determine if there are misspecifications in the models and, if so, which corrections should be made. This approach was employed to test the simple structure models for the values as measured in the 21-item PVQ of the ESS (Knoppen and Saris 2009b). In doing so, one has to chose as the critical values (δ) to be detected with high power (.8) and an α level of .05. The critical value chosen for loadings is .4 and for correlations the value is .1. Using this procedure, we examine whether misspecifications are present in the 40-item PVQ.

The above mentioned approach has been used to test the simple structure models for the values in the short version PVQ in the ESS. This time we want to see whether the same misspecifications were present in the original PVQ scale with 40 items.

3.2 Samples

Most studies of the PVQ have been done with students samples. We also used two student samples that were previously the basis of a test of the PVQ by Schmidt, Bamberg, Davidov, Hermann, and Schwartz (2007). For details of the data collection procedure we refer to the paper mentioned. Both samples came from the University of Gießen, Germany. Sample one included 395 students and sample two 321 students.[5] Using students samples is not a serious limitation because we think that the structure of the measurement model for the values will not be very different for students compared to the one for a representative sample of the population. This hypothesis has recently been confirmed as we will discuss later.

3.3 A Test of the Full Model

As a preparatory analysis, we estimated the simple structure for the full model of 10 values in sample 1, using the ML-estimator of LISREL. The test statistics indicate that this model had to be rejected: The model with 695 degrees of freedom had a chi-square value of 2111 (RMSEA=0.072; NFI=0.77; CFI=0.83; AGFI=0.75; RMR=0.085). The Jrule program indicated 78 possible misspecifications in the model. This is much more than one would expect by chance for testing at the .05 level, if the model were correct. Given these results, we concluded that the model was misspecified and had to be rejected. We next looked at sub-models in order to get a detailed picture of the model misspecifications.

Tests of Sub-Models

We tested the sub-models in two steps. First, we separately tested the factor model for each of the basic values, insofar as that was possible. Such tests are possible only if there are more than three items specified for a basic value. Only then there is enough information to test if the items represent one value or if sub-factors are present. In the second step, we tested each value twice, once together with the immediately preceding value in the circular structure defined by the theory and once together with the immediately succeeding value. We thus looked at the fit of the factor models for pairs of adjacent values. We restricted the analysis to these pairs because Davidov (2008) and Davidov et al. (2008) detected the problems of excessively high correlations in these combinations.

Again, we test the models with the JRULE program. This program suggests corrections to improve the fit in response to two types of problems: (a) the presence of cross-loadings, implying that an item reflects more than one value,

[5] Eldad Davidov generously provided these data that were collected by Sebastian Bamberg and Peter Schmidt, University of Gießen, Germany. The data from sample 1 are in the file WERTE5A.SAV and the data from sample 2 in the file 2INSTRUM.SAV.

Table 2: The Decision Table for Assessing Misspecifications

		Power of the Test	
		High	Low
Is the Modification Index Significant?	Yes	Inspect EPC	Misspecification present
	No	No misspecification	Not informative: inconclusive

and (b) the presence of correlated errors between the items of the same basic value, implying the presence of a sub-factor consisting of those items. We further evaluated whether the corrections to the model suggested by JRULE made theoretical sense. If so, we introduced the corrections into the model one by one.

In order to clarify our procedure, we illustrate it with the Power and Achievement values pair and data from Sample 1. In the first step, we tested a one-factor model including the four items (4, 13, 24, 32) for the Achievement value. This test indicated a poor fit (χ^2=19.8, df=2). The JRULE program suggested two sub-factors: Items 4 and 13 share an emphasis on demonstrating one's ability and being admired for it (labeled "achieving recognition"). Items 24 and 32 share an emphasis on ambition and striving (labeled "achieving advancement"). If we introduce correlated errors for these sub-factors, the model fits the data but there is no degree of freedom left. However, one can also specify the same model as a two-factor model without correlated errors, freeing one degree of freedom for testing. Figure 3 illustrates the alternative structures, where (a) and (c) are one-factor models and (b) and (d) are two-factor models. Models (a) and (b) are equivalent if in model (b) we assume that the correlation between the factors is 1, and models (c) and (d) are similar for our purposes (although not equivalent in a strict statistical sense). If the correlation between the factors is very high, we will use model (c) in the analyses; if the correlation is rather low, we will use model (d).

The two factor model for Achievement fits well (χ^2=.19 and df=1) and Jrule no longer indicates any misspecifications. The correlation between the two factors is .8 which is high but far from 1. The two factors are different although they are related.

In the second step, we tested a factor model for Achievement and Power. For Power, we specified one factor for items 2, 17, 39. For Achievement, we drew on the finding in the previous step that it separates into two factors, "achieving advancement" and "achieving recognition". We chose to model this using two correlated errors in the CFA model; this is feasible because, due to the combination of values, the degree of freedom is not zero. The results in Figure 4 clearly show the strength of the two significant correlated errors. This confirms that there are indeed two subfactors, one representing "achieving advancement" and the other "achieving recognition".

JRULE suggested further improving this model by allowing a cross-loading of item 2 on "achieving advancement" (MI=56.43; EPC=.65; power=.99). This was not surprising because, as noted above, item 2 (being rich) differs in meaning from the other two Power items. The correlation matrix also shows that item 2 has a higher correlation with items from the adjacent value than with items from its own value. We therefore modified the CFA model by adopting the cross-loading suggested by JRULE. After this correction, JRULE suggested no further modifications. Figure 5 provides the estimated values of the parameters.

Figure 5 shows that introducing the cross-loading of item 2 on 'achieving advancement' caused its loading on Power to become insignificant. This implies that 'being rich' is perceived more as an Achievement item than as a Power item. The remaining two Power items (17 and 39) refer more precisely to the specific value "control over people". Had we kept item 2 as a Power item, the composite score for the Power value would have been contaminated by an item that measures Achievement. Hence, the correlation between Power and Achievement would have been overestimated. Comparing Figure 4 with Figure 5 also shows a decrease in the strength of this correlation. This is due to the fact that the model is corrected for the misspecification of item 2. Such items should be avoided because they create contamination between the composite scores of the different values.

We followed the same two-step procedure to analyze each of the remaining eight basic values. After an individual test of the basic value for those values with more than three items, we tested a model that paired it with both basic values adjacent to it in the circular structure, one from each side. Thus, for example, we tested not only the Achievement-Power pair but also the Achievement-Hedonism pair, because Hedonism is also adjacent to Achievement in the circular structure of values (Figure 1) and Davidov (2008) and Davidov et al. (2008) found problems only between adjacent values. We repeated this sequence of tests for each basic value. Finally, we tested the corrected model on the second dataset. In order to avoid capitalization on chance, we accepted only those corrections that were supported in both samples.

4 Results

This section describes the results for each of the 10 basic human values. We label and define the sub-factors that emerge in keeping with suggestions in Schwartz et al. (in press).

4.1 Tradition

The first step revealed that the items for Tradition represent two sub-factors. Item 25 clearly measures Tradition and correlates most strongly with item 20 which asks about the importance of religion. The other two items, 9 and 38,

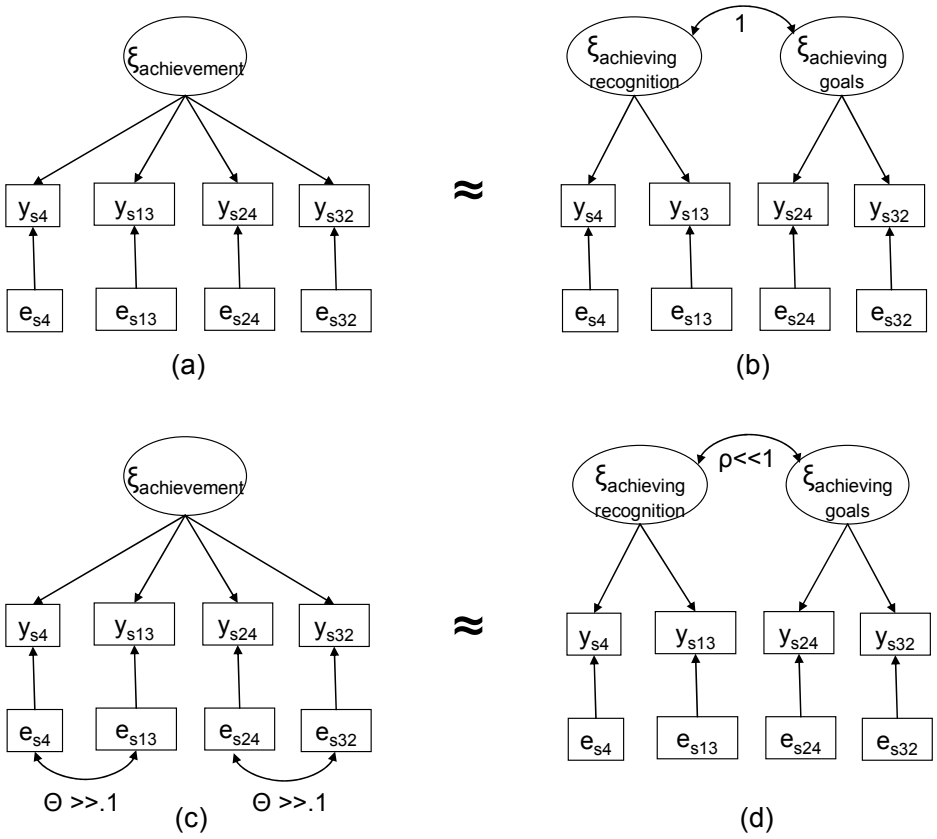

Figure 3. Two similar representations of a one-factor and a two-factor model. ξ refers to latent variables, y to observed variables, e to measurement errors, θ to correlations between error terms, and s_i to item numbers

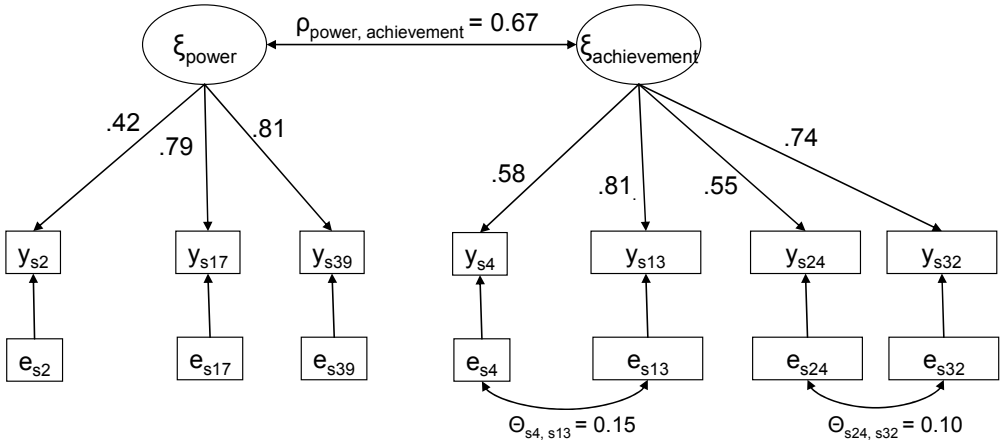

Figure 4. Two factor-structure base model: Power – Achievement. ξ refers to latent variables, ρ to correlation between latent variables, y to observed variables, e to measurement errors, θ to correlations between error terms, and s_i to item numbers

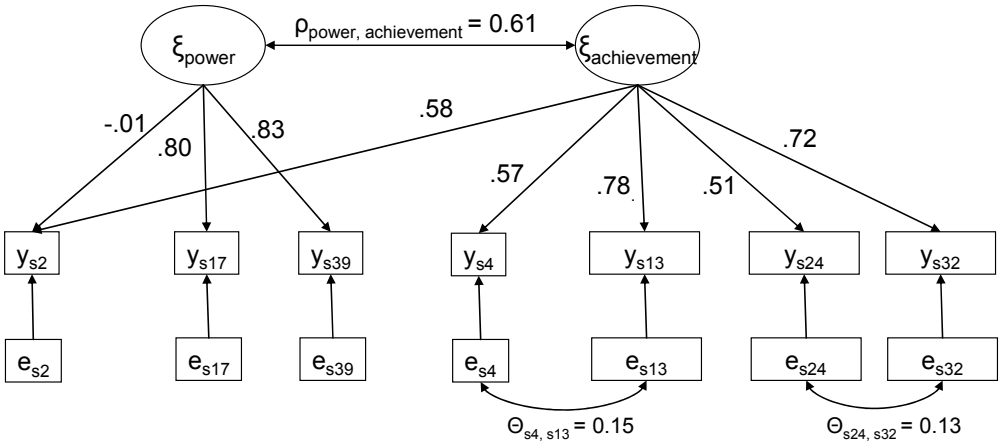

Figure 5. Two factor-structure: Power – Achievement, with cross-loading and correlated errors. ξ refers to latent variables, ρ to correlation between latent variables, y to observed variables, e to measurement errors, θ to correlations between error terms, and s_i to item numbers

measure "humility" (i.e., recognizing and accepting one's insignificance in the universe). The two sub-factors correlated only .29 and .31 in the two different samples. This suggests that these items reflect two factors rather than one. The analysis with all four items in the second step led to serious problems because the correlation between the sub-factors is so low. We therefore decided to continue with the first subfactor alone.

With this restriction of the set of items for Tradition, the simple structure model with Conformity was tested in the second step. This model fit the data after allowing a correlated error between items 16 and 36 for Conformity. We then tested the model for the limited set of Tradition items in combination with Benevolence. After allowing a cross-loading of the religion item (20) on the Benevolence factor, this model fit the data.

This led us to the following conclusions with regard to the Tradition items: Items 9 and 38 measure a distinct concept, Humility, which is only minimally related with Tradition. Item 25 is a direct measure of Tradition. The religion item (20) is a less attractive item for Tradition because it also has a cross-loading with another factor and its correlation with item 25 is relatively low. We therefore conclude that there is only one pure item (25) to measure Tradition.

4.2 Conformity

For Conformity, we also had to adjust the one-factor model to a two-factor model to obtain a good fit. Items 16 and 36 measure a factor which we label "conformity to interpersonal expectations" (meeting informal social expectations in order to avoid upsetting others). Items 7 and 28 measure "conformity to authority" (Doing what authorities or rules explicitly require). The correlations between these two factors were .69 and .95 in the two samples. This was still high enough to continue, for simplicity, in the second step with a one-factor model with two correlated errors (model (c) from Figure 3). In the second step, we first tested the Conformity-Tradition pair of values. Restricting the Tradition items to the two mentioned above (25 and 20), we obtained a model with a good fit. We then tested the Conformity-Security pair, including the correlated errors between the pairs of Conformity items. After this adjustment, the model also fit the data.

We conclude that items 16 and 36 form a "conformity to interpersonal expectations" subfactor and items 7 and 28 form a "conformity to authority" sub-factor. These two distinct values have clearer and narrower meanings than the broad, single Conformity value.

4.3 Security

For Security, we also had to adjust the one-factor model to a two-factor model to obtain a good fit. Items 5, 21, and 31 formed one factor and items 14 and 35 formed another. The three items of the former factor all refer to safety and orderliness in one's personal life and can be labeled "personal security". The two items of the second factor refer to safety and order in the wider society and can be labeled "societal security". The correlations between these two factors were .44 and .65 in the two samples. Schwartz and Boehnke (2004) also identified these two components of Security in their CFA analyses of data from 27 countries.

In the second step, we first tested the Conformity-Security pair of values. Jrule suggested cross-loadings from two of the "personal security" items, 5 and 21, on Conformity. After allowing these cross-loadings, these two items no longer loaded significantly on a separate factor. Item 5 now loaded .34 and .54 on the Conformity factor in the two samples, and item 21 loaded .45 and .39. The "societal security" sub-factor remained, with correlated errors between its component items of .33 and .25 in the two samples. We next tested the Security-Power pair. This test confirmed that the "societal security" sub-factor was present in both samples, with correlated errors between items 14 and 35 of .33 and .30.

In sum, the five items intended to measure Security in the PVQ split into three subsets. Two [14,35] form a factor that

measures "societal security". The two items that refer to living in secure surroundings [5] and [21] have cross-loadings with the value Conformity. Finally, the item [31] that measures avoiding sickness stands alone. The latter three items were intended to measure diverse aspects of 'personal security', but they do not form a factor. Either item 5 or item 31, but not both, could be considered an indicator of 'personal security' on which to build a factor, depending on the aspect one views as more central.

4.4 Power

SEM testing of the one-factor model for Power was not possible because there were only three indicators in the PVQ. The tests therefore started with the second step. We first tested the Power-Achievement pair of values. As anticipated in our earlier discussion, item 2, which refers to "being rich", correlated more highly with items intended to operationalize Achievement (average correlations of .37 and .44 for the two samples) than with the two other Power items, 7 and 39 (average correlations of .25 in both samples). Moreover, there was a significant crossloading of item 2 on Achievement (.51 and .52 in the two samples). When including this crossloading, the original loading of item 2 on the Power factor disappeared (-.01/.12). Including this cross-loading also reduced the correlation between the Power and Achievement factors from .67 to .61 in sample 1 and from .57 to .51 in sample 2.

We then tested the Security-Power pair. In light of the findings of the previous test, we excluded item 2 from the model. The correlations between the Security and Power factors were .29 and .17 in the two samples. The loadings of the items on their respective factors remained approximately the same in this analysis. In sum, items 17 and 39, that refer respectively to "being in charge" and "making the decisions." apparently measure Power well, whereas item 2 that refers to "being rich" correlates more with Achievement than with Power and therefore should not be included as indicator of Power.

4.5 Achievement

For Achievement, two sub-factors were detected in the first step. Items 4 and 13 refer to obtaining recognition from other people for one's performance and can be labeled "achieving recognition". Items 24 and 32 refer to getting ahead through striving and can be labeled "achieving advancement". However, because these two factors correlated quite highly (.77 and .86 in the two samples), we continued for simplicity with a model with one factor and two correlated errors. In the second step, we first tested the Power-Achievement pair of values. After including correlated errors between items 4 and 13 and between items 24 and 35 in the model, the model fit the data and no further corrections were necessary.

We then tested the Hedonism-Achievement pair. This model also fit the data: The loadings of the four Achievement items were .66/.85/.56/.66 in sample 1 and .64/.87/.80/.79 in sample 2; the correlations between Achievement and Hedonism were .18 in sample 1 and .15 in sample 2. In sum, we would argue that it is possible to distinguish two sub-factors of Achievement, "achieving recognition" and "achieving advancement".

4.6 Hedonism

SEM testing of the one-factor model for Hedonism was not possible because there were only three indicators in the PVQ. The tests therefore started with the second step. We first tested the Hedonism-Achievement pair. This model functioned well, as noted above. We then tested the Hedonism-Stimulation pair. This model also functioned correctly: The loadings of the three Hedonism items were .74/.43/.93 in sample 1 and .83/.71/.78 in sample 2; the correlations between Hedonism and Stimulation were .67 in sample 1 and .68 in sample 2. In sum, the Hedonism items function well.

4.7 Self-direction

The test of the one-factor model for Self-direction suggested that a two-factor solution is better. Items 1 and 22 formed one sub-factor that refers to freedom to develop and cultivate one's own ideas and can be labeled "autonomy of thought". Items 11 and 34 formed another sub-factor that refers to freedom to determine one's own actions and decisions and can be labeled "autonomy of action". The correlation between the two sub-factors was rather low (.41 and .53 in the two samples). In the second step, we tested the Self-direction-Universalism pair of values. The loadings of Self-direction were low (only 2 of the 4 items >.4) in both samples. The two "autonomy of action" items (11 and 34) continued to form a sub-factor with correlated errors of .29 and .36 in the two samples.

We then tested the Stimulation-Self-direction pair. Although the two factors separated, item 1, "being creative, original, having new ideas" had low loadings on Self-direction (.34 and .49 in the two samples). Overall, the four Self-direction items were not highly intercorrelated (average of .20 and .26 in the two samples), suggesting that this value is conceptually quite broad and might better be split into the two values mentioned above.

4.8 Stimulation

SEM testing of the one-factor model for Stimulation was not possible because there were only three indicators in the PVQ. The tests therefore started with the second step. In this step, we first tested the Hedonism – Stimulation pair of values. As noted under Hedonism above, this model functioned correctly. We then tested the Stimulation-Self-direction pair. Jrule suggested a cross-loading of item 6 "do different, new things" of .33 on Self-direction in sample 1. Including this cross-loading, the original loading of item 6 on Stimulation decreased from .63 to .25 and the correlation between Stimulation and Self-direction decreased from .55 to .27. Sample 2 confirmed this pattern: There was a cross loading of .29 and a decrease in the correlation between the two values of .06. In sum, after dropping item 6 because of its cross-loading, the remaining two items provide a pure measure of Stimulation.

4.9 Universalism

The test of the one-factor model for Universalism suggested that a two-factor solution is better. Items 3 and 29 formed one sub-factor that we labeled "social equality", which refers to living in a world in which all people are treated equally and justly. Items 19 and 40 formed another sub-factor that we labeled "preserving nature", which refers to protecting the natural environment. The remaining items, 8, "listening to different people", and 23, "promoting harmony and peace", were more closely related to items 3 and 29 and also fit the "social equality" factor, albeit with rather low loadings. The correlation between the factors was .60 and .67 in the two samples, which is small enough to suggest that the two factors measure different components of Universalism.

In the second step, we first tested the Universalism-Benevolence pair of values. Jrule suggested a cross-loading of item 8 on Benevolence in both samples. Adapting this modification resulted in cross-loadings for item 8 of .43 and .50 in the two samples. Moreover, the original loading of item 8 on the Universalism factor disappeared (.17/.08). Jrule also suggested a crossloading from item 29, "treating everyone justly and protecting the weak", on Benevolence. Adapting this modification resulted in cross-loadings for item 29 of .33 and .36, and the original loading of item 29 on the Universalism factor decreased from .73 to .53 and from .74 to .50 in the two samples. We then tested the Self-direction-Universalism pair. After the changes mentioned above, the model fit the data.

In sum, items 3 and 29 measure a "social equality" sub-factor of Universalism and items 19 and 40 measure a "preserving nature" sub-factor, although the loading of item 19 is somewhat low. Schwartz and Boehnke (2004) also identified these two components of Universalism in their CFA analyses of data from 27 countries. The status of items 8 and 23 is less clear. Although both express an emphasis on "social harmony" and are related to the "social equality" sub-factor, item 8 loads more heavily on Benevolence. One could consider the possibility that these two items express two aspects of a separate Social Harmony value – Tolerance and Peace. In all events, it is clear that the broad Universalism values should be split into at least two separate values.

4.10 Benevolence

The test of the one-factor model for Benevolence also suggested that a two-factor solution might be better, albeit that the two factors were highly correlated (.77 and .74 in the two samples). Items 18 and 27 formed a sub-factor that refers to being loyal to and supportive of one's in-group members. Items 12 and 33 formed a second sub-factor whose unique focus is less clear. Moreover, item 33 had a very low loading on that factor. Given the substantial correlation between the factors, the unclear meaning of the second factor and the low loading of item 33 on the second factor, we conclude that it does not make sense to specify two factors in this case. Therefore we continue with a one-factor model with correlated errors. In the second step, we first tested the Benevolence-Tradition pair of values. As explained above for Tradition, we restricted the set of Tradition items and introduced a cross-loading from the religion item (20) to Benevolence. This model fit the data.

We then tested the Universalism-Benevolence pair. We allowed the cross-loadings from Universalism items on Benevolence that were noted in our discussion of Universalism. This model also fit the data. In sum, the items for the Benevolence value function well. We suggest continuing with only one factor for Benevolence but removing item 33 from the set of indicators.

5 Conclusions

Schwartz proposed the PVQ to measure 10 motivational distinct basic values that form four higher-order values. The scale consists of 40 items, with three to six items for each value. Prior studies used SSA to evaluate the instrument and found that the ten values had the same position in the circular structure as posited theoretically in many samples across the globe. That is, the PVQ items intended to represent each value formed spatial regions in the two-dimensional space that represented their intercorrelations, and these regions followed the circular order posited by the theory.

This paper has evaluated the same items with SEM. More specifically, for each value and each of two samples, we tested (a) a one-factor structure (only for values with more than three items) and (b) a two-factor structure (1) with the value that preceded that value in the circular structure and (2) with the value that followed that value in the circular structure. The advantage of performing multiple tests for the same value is that conclusions are based on confirmed and compared results. The adjacent values were chosen for further analysis because the problems that Davidov (2008) and Davidov et al. (2008) detected in the data concerned only adjacent values.

Table 3 summarizes the results of the analyses. It lists the factors and sub-factors we identified in the PVQ scale as well as single, "loose" items. Loose items have either higher crossloadings with other values than with their own value or low correlations with other items from their own value. We found that the items intended to measure several basic values are heterogeneous and that they form multiple sub-factors or exhibit cross-loadings. This reflects the fact that the items selected to measure the values, when constructing the PVQ, were intended to maximize coverage of the diverse substantive components in the definition of each broad value. This led to three outcomes: (a) relatively low correlations among the items intended to measure the more conceptually complex values, (b) relatively high correlations of some items with items of adjacent values, and (c) high correlations between some adjacent values. Knoppen and Saris (2009a) reported similar findings for data from the 21-item ESS value survey.

The analyses revealed that it is possible to avoid the problems we identified by distinguishing 15 narrower values rather than the original ten broad values. Table 3 specifies 15 narrower values that meet the criteria of pure factors: Tra-

Table 3: Overview of results

10 Original Basic Values	PVQ Items	Factors, Sub-Factors, and Items		'Loose' Single Items
Tradition	9, 20, 25, 38	1. Tradition: 25	2. Humility: 9, 38	(20)
Conformity	7, 16, 28, 36	3. Conformity to interpersonal expectations: 16, 36	4. Conformity to authority: 7, 28	
Security	5, 14, 21, 31, 35	5. Societal security: 14, 35		16. Personal security, safety: 5 or 21 17. avoiding sickness: 31
Power	2, 17, 39	6. Power: 17, 39		18. Wealth: 2
Achievement	4, 13, 24, 32	7. Achieving advancement: 24, 32	8. Achieving recognition: 4, 13	
Hedonism	10, 26, 37	9. Hedonism: 10, 26, 37		
Self-direction	1, 11, 22, 34	10. Autonomy of action: 11, 34	11. Autonomy of thought: 1, 22	
Stimulation	6, 15, 30	12. Stimulation: 15, 30		(6)
Universalism	3, 8, 19, 23, 29	13. Social equality: 3, 29	14. Preserving nature: 19, 40	19. Social Harmony, tolerance: 8 peace: 23
Benevolence	12, 18, 27, 33	15. Benevolence: 12, 18, 27		(33)

dition, Humility, Conformity to Interpersonal Expectations, Conformity to Authority, Societal Security, Power, Achieving Advancement, Achieving Recognition, Hedonism, Autonomy of Action, Autonomy of Thought, Stimulation, Social Equality, Preserving Nature, and Benevolence. In working with the PVQ, one may ignore the 'loose' items (they are placed within brackets) or, alternatively, one may consider whether they tap potential value constructs, such as those listed in the last column of Table 3, that could be worth examining. For some of the 'loose' items, we think the latter option merits consideration.

Among the items originally intended to measure Security, item 5 appears to refer to one type of "personal security" (personal safety) and item 31 to another type of "personal security" (avoiding sickness). Among the items originally intended to measure Universalism, items 8 and 23 appear to refer to aspects of "social harmony". Item 8 may measure a potential value of "tolerance toward those who are different" and item 23 may measure "intergroup peace". Finally, item 2, which was intended to measure Power, may measure another potential value, "wealth". In order to determine whether these five potential value constructs are indeed worth adding, it is necessary to include at least one extra item intended to measure each one in the PVQ. It is also necessary to add one item to measure Tradition, which is currently measured by only one item.

Treating the PVQ as a scale that measures 15, and perhaps more, values has two important advantages. First, it avoids correlations between adjacent latent values that are close to 1.00 or higher. This increases discriminant validity. Second, it avoids cross-loadings and contamination of composite scores. The pure factors this approach provides may therefore have greater predictive power. Consider one example. One would expect "preserving nature", a specific value discriminated in the broad, basic value of Universalism, to predict environmentally-friendly behaviors more effectively than the broad, basic value. Table 4, based on data from sample 2, demonstrates that this is indeed the case. For each of ten environmentally-friendly behaviors, the correlation with saying 'yes' was significantly higher for the specific "preserving nature" value than for the composite score of the broad Universalism value.

5.1 Strengths and Limitations

The major strength of this study is that its results and conclusions are based on a combination of different methodological approaches (SSA and SEM). SEM is complemented with an alternative way of testing models (Saris et al. 2009; van der Veld et al. 2009). To the best of our knowledge, only one previous study (Schwartz and Boehnke 2004) has combined insights from SSA and CFA to examine data based on the Schwartz values theory.[6] However, that study did not address the problem of cross-loadings and ambiguous composite scores.

A limitation of this study is that we tested only two samples rather than a larger number and that both samples were from student populations and from only one country. For two reasons, however, this limitation may be less serious than it initially appears. First, Knoppen and Saris (2009b) performed tests similar to those of this paper on ESS data that employed the 21-item short version of the PVQ. The ESS data came from probability samples of the eligible residential populations aged 15+ in three countries. The tests yielded results compatible with the current findings with the full PVQ. Second, recent studies based on non students samples in Poland (Cieciuch and Schwartz 2012) and Germany (Beierlein et al. 2012) confirmed to a large extend the finding presented in this paper. The most recent study was done in Estonia. Also this study comparing two representative samples of the Estonian and Russian speaking populations

[6]This was true when this paper was presented at the conference in Bolzano but since then several researchers followed our approach, see below.

Table 4: Correlations of the Broad, Basic Universalism Value versus the Specific Preserving Nature Value with Environmentally-Friendly Behavior (n=321)

Questions about Environment-Friendly Behavior	Basic Value Universalism	Specific Value Preserving Nature
During the last 5 years, did you ...		
sign a petition which required measures to protect the natural environment? [yes]	.282	.322
donate money to an organization for environmental protection?	.185	.246
boycott or avoid products of a business because you believed it damages the environment [yes]	.319	.347
Are you member of a group that aims to conserve and protect the environment? [yes]	.203	.284
In the last 12 months, did you ...		
read a newsletter, magazine or other publication of an environmental protection group? [yes]	.396	.426
In the last twelve months, how often did you purposely buy ...		
fruit and vegetables which were grown without chemical pesticides and fertilizers? [yes]	.310	.372
paper and plastic products produced from recycled waste material?	.283	.313
environmentally friendly detergents and domestic cleansers? [yes]	.331	.343
In order to protect the environment, I would be willing to ...		
pay more. [yes]	.336	.363
accept a lower standard of living. [yes]	.338	.375

in Estonia applied the approach used here and reported only minimal deviations (Lilleoja 2011).

5.2 Methodological Issues and Implications

This study highlights the importance of combining and contrasting different methodological approaches when developing a theory and its operationalization. The appeal of SSA lies in its power to organize values visually and to describe their conflicting and congruent relationships. It robustly presents the proximity of items from different constructs. It allows items to fill the two-dimensional space and thereby to capture all aspects of a domain whose components are theorized to constitute a continuum, as values are. However, SSA provides no guidance for specifying the boundaries for the different concepts.

The appeal of SEM lies in its capacity to test and derive 'pure' indexes of clearly defined constructs that are not confounded by overlap with other constructs. Consequently, SEM can identify sets of homogeneous items to measure each value. SSA and SEM thus appear to foster contradictory selection principles. In the case of values, SEM can also test the relationships of the values with external variables (e.g., with other values and with behavior), while correcting for measurement errors. This study has sought to demonstrate that it is feasible to find a balance between both theoretical coverage and homogeneity of items.

The results of the analyses reported here imply that it is desirable to compute new composite scores for values based on existing datasets. Measuring the 15 narrower values identified here in these datasets, using the sets of items that permit 'pure' measurement, will avoid contamination of values and may improve their predictive power. The current research further implies that it is worthwhile to develop a revised set of items based on the PVQ that insures maximal continuity with the original scale but covers the larger number of more narrowly defined values that may be distinguishable.[7] In developing a revised scale, the loose items that do not contribute to the calculation of composite scores may be dropped and items may be added for those values that are reflected by fewer than 3 items. Overall, we hope to have provided inspiring ideas for the further development of human values scales.

References

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

Beierlein, C., Davidov, E., Schmidt, P., & Schwartz, S. H. (2012). Testing the discriminant validity of Schwartz' Portrait Value Questionnaire items: A replication and extension of Knoppen and Saris (2009). *Survey Research Methods*, *6*(1), 26-36.

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

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait multimethod matrices. *Psychological bulletin*, *56*, 81-105.

Cieciuch, J., & Schwartz, S. H. (2012). The number of distinct basic values and their structure assessed by PVQ-40. *Journal of personality assessment*, *94*(3), 321-328.

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurement with the second round of the European Social Survey. *Survey Research Methods*, *2*(1), 33-46.

Davidov, E., & Schmidt, P. (2007). Are values in the Benelux comparable? In G. Loosveldt, M. Swyngedouw, & B. Cambré (Eds.), *Measuring meaningful data in social research* (p. 373-386). Leuven (Belgium): Acco.

[7] Such an effort is currently underway. See Schwartz, et al. (in press).

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.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological methods*, *5*(2), 155-174.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrica*, *33*, 469-506.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Jowell, R., Kaase, M., Fitzgerald, R., & Eva, G. (2007). The European Social Survey as a measurement model. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally. Lessons from the European Social Survey* (p. 1-32). London: Sage Publications.

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

Knoppen, D., & Saris, W. (2009b). Measuring Value Orientations Across Nations: Analysis of the Quality of the European Social Survey. *Survey Research Methods*, *3*(2), 91-103.

Lilleoja, L. (2011). *Testing a New Approach for Operationalization of the Basic Value Model.* Paper presented at the ESRA conference 2011 in Lausanne.

Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Sagiv, L., & Schwartz, S. H. (1995). Value priorities and readiness for out-group social contact. *Journal of personality and social psychology*, *69*(3), 437-448.

Saris, W. E. (1981). Different questions, different variables. In C. Fornell (Ed.), *A second generation of multivariate analysis* (Vol. 2, p. 78-96). New York: Praeger.

Saris, W. E., & Gallhofer, I. (2007). *Design, evaluation and analysis of questionnaires for survey research*. Hoboken (New Jersey): Wiley Interscience.

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

Schmidt, P., Bamberg, S., Davidov, E., Hermann, J., & Schwartz, S. H. (2007). Die Messung von Werten mit dem "Portrait Value Questionnaire". *Zeitschrift für Sozialpsychologie*, *38*, 249-263.

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. (1994). Are There Universal Aspects in the Structure and Contents of Human Values? *Journal of Social Issues*, *50*(4), 19-45.

Schwartz, S. H. (1996). Value priorities and behavior: Applying a theory of integrated value systems. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The Psychology of Values: The Ontario Symposium* (Vol. 8, p. 1-24). Hillsdale, NJ: Erlbaum.

Schwartz, S. H. (2003). *A Proposal for Measuring Value Orientations across Nations. Chapter 7 in the Questionnaire Development Package of the European Social Survey.* http://essedunet.nsd.uib.no/cms/topics/1/.

Schwartz, S. H. (2006). Les valeurs de base de la personne: Théorie, mesures et applications [Basic human values: Theory, measurement, and application]. *Revue Française de Sociologie*, *47*, 929-968.

Schwartz, S. H. (2007). 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. 169-203). London: Sage Publications.

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of personality and social psychology*, *53*(3), 550-562.

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, *58*, 878-891.

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

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., et al. (in press). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*.

Schwartz, S. H., Melech, G., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology*, *32*(5), 519-542.

Schwartz, S. H., & Rubel, T. (2005). Sex differences in value priorities: cross-cultural and multimethod studies. *Journal of personality and social psychology*, *89*(6), 1010-1028.

Schwartz, S. H., & Sagie, G. (2000). Value consensus and importance: A cross-national study. *Journal of Cross Cultural Psychology*, *31*, 465-497.

van der Veld, W. M., Saris, W. E., & Satorra, A. (2009). *Jrule 2.0, Judgement Aid Rule for Structural equation models: user manual.* Unpublished document.

Appendix: PVQ Items*

Benevolence

12. It's very important to him to help the people around him. He wants to care for other people.
18. It is important to him to be loyal to his friends. He wants to devote himself to people close to him.
27. It is important to him to respond to the needs of others. He tries to support those he knows.
33. Forgiving people who might have wronged him is important to him. He tries to see what is good in them and not to hold a grudge.

Universalism

3. He thinks it is important that every person in the world be treated equally. He wants justice for everybody, even for people he doesn't know.
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.
19. He strongly believes that people should care for nature. Looking after the environment is important to him.
23.He believes all the worlds' people should live in harmony. Promoting peace among all groups in the world is important to him.
29. He wants everyone to be treated justly, even people he doesn't know. It is important to him to protect the weak in society.
40. It is important to him to adapt to nature and to fit into it. He believes that people should not change nature.

Self-direction

1. Thinking up new ideas and being creative is important to him. He likes to do things in his own original way.
11. 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.
22. He thinks it's important to be interested in things. He likes to be curious and to try to understand all sorts of things.
34. It is important to him to be independent. He likes to rely on himself.

Stimulation

6. He thinks it is important to do lots of different things in life. He always looks for new things to try.
15. He likes to take risks. He is always looking for adventures.
30. He likes surprises. It is important to him to have an exciting life.

Hedonism

10. He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.
26. Enjoying life's pleasures is important to him. He likes to 'spoil' himself.
37. He really wants to enjoy life. Having a good time is very important to him.

Achievement

4. It's very important to him to show his abilities. He wants people to admire what he does.
13. Being very successful is important to him. He likes to impress other people.
24. He thinks it is important to be ambitious. He wants to show how capable he is.
32. Getting ahead in life is important to him. He strives to do better than others.

Power

2. It is important to him to be rich. He wants to have a lot of money and expensive things.
17. It is important to him to be in charge and tell others what to do. He wants people to do what he says.
39. He always wants to be the one who makes the decisions. He likes to be the leader.

Security

5. It is important to him to live in secure surroundings. He avoids anything that might endanger his safety.
14. It is very important to him that his country be safe from threats from within and without. He is concerned that social order be protected.
21. It is important to him that things be organized and clean. He *doesn't* want things to be a mess.
31. He tries hard to avoid getting sick. Staying healthy is very important to him.
35. Having a stable government is important to him. He is concerned that the social order be protected.

Conformity

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.
16. It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong.
28. It is important to him to be obedient. He believes he should always show respect to his parents and to older people.
36. It is important to him to be polite to other people all the time. He tries never to disturb or irritate others.

Tradition

9. 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.
20. Religious belief is important to him. He tries hard to do what his religion requires.
25. He believes it is best to do things in traditional ways. It is important to him to follow the customs he has learned.
38. It is important to him to be humble and modest. He tries not to draw attention to himself.

* Items are numbered according to their order in the questionnaire.