JOURNAL OF PERSONALITY ASSESSMENT
2024, VOL. 106, NO. 5, 651–664
https://doi.org/10.1080/00223891.2024.2311193

Measuring the Four Higher-Order Values in Schwartz's Theory: Validation of a 17-Item Inventory

Clemens M. Lechner[1], Constanze Beierlein[2], Eldad Davidov[3,4] and Shalom H. Schwartz[5]

[1]Department of Survey Design and Methodology, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany; [2]Department of Psychology, Hamm-Lippstadt University of Applied Sciences, Hamm, Germany; [3]Department of Sociology, University of Cologne, Koln, Germany; [4]URPP 'Social Networks', University of Zurich, Zurich, Switzerland; [5]Department of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel

ABSTRACT

Schwartz's theory of basic human values is the dominant framework for assessing values. One of its strengths is that it allows for different levels of analysis. The 10 basic values can be reliably assigned to four higher-order dimensions: Openness to Change, Conservation, Self-Transcendence, and Self-Enhancement. In this paper, we examined the psychometric properties of the Higher-Order-Value Scale-17 (HOVS17), an inventory that economically assesses these higher-order values. We analyzed data from the GESIS Panel, an ongoing large-scale probability-based panel study that fields HOVS17 annually since 2013 and for which HOVS17 was originally developed. We found HOVS17 to have satisfactory psychometric properties. The 17 items were located in the two-dimensional multi-dimensional scaling (MDS) space as hypothesized. All four subscales were unidimensional, showed good fit when modeled as reflective latent variables, and had acceptable reliabilities as well as one-year test–retest stabilities (.65 to .69). The subscales correlated in theoretically plausible ways with a wide range of correlates and criteria, such as personality traits and well-being. This demonstrates that HOVS17 provides a sound basis for studying the development, precursors, and consequences of the higher-order values in the GESIS Panel and in future surveys that adopt HOVS17. We also discuss suggestions for further improvements of the inventory.

ARTICLE HISTORY
Received 7 July 2023
Revised 15 January 2024
Accepted 18 January 2024

Measuring values

The study of values is of long-standing interest in psychology, sociology, and adjacent fields. In the past three decades, Schwartz's (1992) theory of human values has become the leading framework for value research. His original framework distinguishes 10 basic human values whose quasi-circumplex structure emerges with great regularity across cultures and languages (Schwartz et al., 2012). Based on their position in two-dimensional space, these basic human values can be assigned to four more abstract, higher-order values that reflect fundamental tensions between Openness to Change vs. Conservation and between Self-Enhancement vs. Self-Transcendence.

Drawing upon Schwartz's theory, value research has become a vibrant field spanning multiple social and behavioral science disciplines and addressing a broad diversity of topics such as the predictive power of values for attitudes and behaviors (e.g., Lee et al., 2022), links between values and well-being (e.g., Grosz et al., 2021; Sortheix & Schwartz, 2017), life-span development of values (e.g., Gouveia et al., 2015; Milfont et al., 2016), and cultural or contextual influences on values (e.g., Sagiv & Schwartz, 2022; Sortheix et al., 2019).

Concomitantly, several national and international survey programs such as the European Social Survey (ESS) have included measures of the Schwartz values. These surveys provide a wealth of secondary data for value research and several well-validated value inventories for future survey programs from which to choose. Most of these inventories aim to assess all 10 basic human values. However, assessing all the basic human values may not always be feasible for surveys because of constraints on survey length. In some cases, parsimoniously assessing the four higher-order values with fewer items is sufficient or even desirable.

To meet the needs of surveys and individual researchers who seek to measure only the four higher-order values, we here evaluate a 17-item inventory developed to assess these higher-order values economically. We developed a German version of this inventory for the GESIS Panel, an ongoing probability-based panel study run by GESIS—Leibniz Institute for the Social Sciences (Bosnjak et al., 2018). The inventory, henceforth called the Higher-Order Value Scale-17 (HOVS17), has since been fielded in three cohorts and nine waves of the GESIS Panel so far. However, the psychometric properties of the inventory have not yet been thoroughly investigated and

CONTACT Clemens M. Lechner clemens.lechner@gesis.org GESIS – Leibniz Institute for the Social Sciences B6, 4–5, 68163 Mannheim, Germany.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00223891.2024.2311193.

documented. To fill this gap, we provide an in-depth analysis of the psychometric properties of HOVS17 including descriptive statistics, structural/factorial validity, reliability and test–retest stability, measurement invariance across socio-demographic groups, and nomological networks/criterion validity. Our aim is to provide benchmarks for secondary users of the GESIS Panel but also for researchers who aim to include the measure in their own future surveys.

A brief overview of Schwartz's framework

Schwartz (1992; see also Schwartz, 1994) defined values as beliefs pertaining to desirable end states or modes of conduct that are ordered in importance, transcend specific situations, and serve as guiding principles in life which influence the selection or evaluation of behavior, people, and events. According to this theory, basic human values form a motivational continuum, represented in a quasi-circumplex or "circle" model (Schwartz & Boehnke, 2004). Values with compatible underlying motivations are located closer to each other in the circle, whereas those with incompatible or conflicting underlying motivations are located on opposite sides. By segmenting the two-dimensional MDS space, Schwartz (1992) distinguished 10 motivationally distinct basic human values. These 10 values can be assigned to the four broader, higher-order values: Openness to Change (i.e., the importance of following one's own intellectual and emotional interests, especially in novel situations), Conservation (i.e., the importance of self-restriction, preserving the past, order, and resistance to change), Self-Transcendence (i.e., the importance of transcending selfish concerns and promoting the welfare of other people), and Self-Enhancement (i.e., the importance of enhancing one's own interests, even at the expense of others).

Figure 1 depicts, in idealized form, the structure of human values according to Schwartz's theory. The higher-order value Conservation includes the basic values Tradition, Conformity, and Security, whereas the opposing higher-order value Openness to Change includes Self-Direction and Stimulation. Self-Enhancement is comprised of Achievement and Power, whereas its opposing higher-order value Self-Transcendence includes Benevolence and Universalism. The basic value Hedonism is not clearly assigned to any of the four higher-order values because its motivational content combines elements of both Openness to Change and Self-Enhancement as reflected in its MDS position in most cultures. As in any quasi-circumplex, one could combine adjacent individual values into different higher-order combinations. However, these four higher-order values are widely used and readily interpretable. We therefore chose them as the focus of HOVS17.

The need for a short higher-order value scale

The most widely used inventories assessing basic human values in Schwartz's theory are the 57-item Schwartz Value Survey (SVS; Schwartz, 1992), the 40-item Portrait Value Questionnaire (PVQ; Schwartz, 2003), and the 57-item Revised PVQ (PVQ-RR; Schwartz & Cieciuch, 2022). These inventories may be too lengthy for some surveys, and assessing all the 10 values may not be feasible with a reasonable number of items. For example, the PVQ-21 in the ESS was not able to distinguish between all 10 values in the theory (Davidov et al., 2008). More generally, for some research questions involving values, the four more abstract higher-order values may provide a more appropriate level of abstraction. For example, if the goal is to study the lifespan development of,

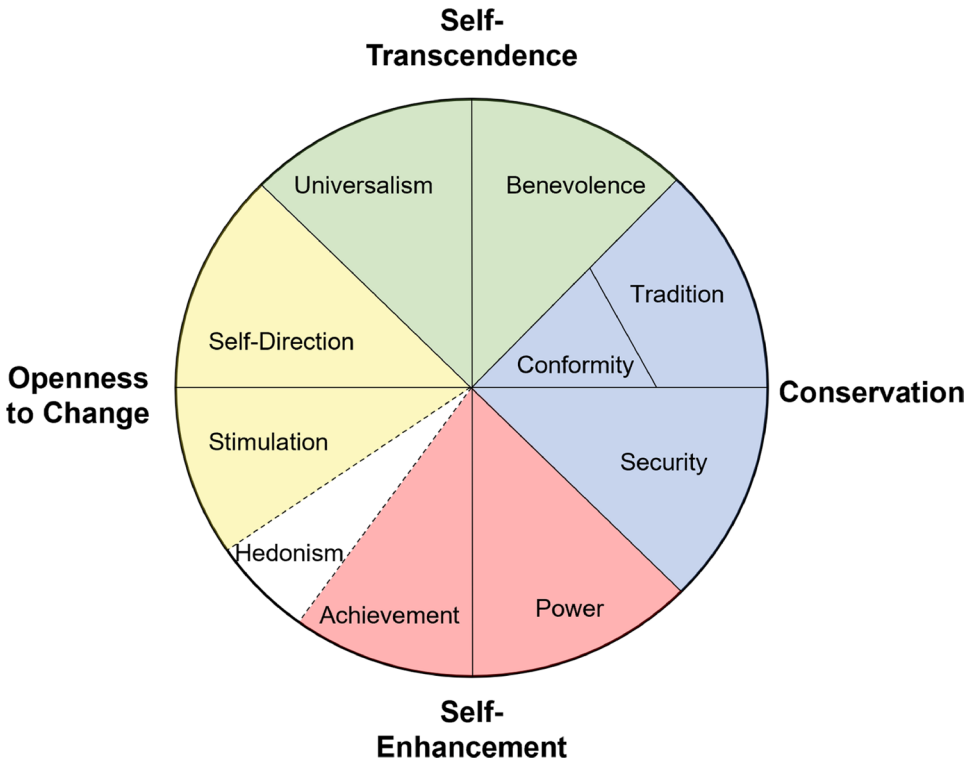

Figure 1. Idealized quasi-circumplex structure of Schwartz's value theory.

or group differences in, broad value orientations, the higher-order values may suffice (e.g., Grosz et al., 2021).

The motivation for developing HOVS17 was to provide a short and economic, yet reliable and valid inventory that enables measuring the four higher-order values with as few items as possible. There exist various short value inventories, such as the 21-item version of the PVQ included in the ESS (PVQ-21; Davidov et al., 2008). However, the ESS inventory is slightly longer than HOVS17 and, more importantly, includes two sentences per item; such double-barreled items increase response burden and may reduce validity. HOVS17 avoids such double-barreled items and, akin to PVQ-RR, uses a single sentence per value item.

Another aim in developing HOVS17 was to allow for a wide variety of statistical approaches that secondary data users might adopt. In particular, the goal was to enable analyses based not only on observed scores but also on a latent variable for each higher-order value. Observed scores are the most widely used way of including values in statistical analyses. Typically, unit-weighted mean scores are computed as unit-weighted sum or mean scores for each value dimension. Often, this is done after centering all items at the within-person mean to obtain value *priorities* (relative scores instead of absolute scores). Importantly, centering also eliminates individual differences in scale use that reflect acquiescent responding, socially desirable responding, the overall importance of values to the person, or some combination of these and other influences (e.g., Rudnev, 2021; Schwartz et al., 1997).

This standard scoring approach based on observed scores treats values as *formative* constructs. However, some researchers (e.g., Davidov et al., 2008; Grosz et al., 2021; Schwartz & Boehnke, 2004) prefer latent-variable models such as confirmatory factor analysis (CFA). Using latent variables corresponds to conceptualizing the values as reflective measures in which a common latent variable is thought to influence all items that measure a given value. In set theoretical terms (see McCrae, 2015), formative indices are the union of the variances of the constituent items (i.e., total variance), whereas reflective latent variables correspond to their intersection (i.e., common variance).

Because the four higher-order values, by definition, are abstract, heterogeneous constructs, subscales that measure them would ordinarily not conform to the unidimensionality and internal consistency standards of reflective measurement. Therefore, one goal in constructing HOVS17 was to enable researchers to apply both conceptualization of values (i.e., formative and reflective) and attendant methodologies.

Initial construction of the inventory

To develop a new inventory to measure the four higher-order values as economically and validly as possible, the second, third, and fourth authors set out in 2012 to select items from a revised version of the 57-item Portrait Values Questionnaire (i.e., PVQ-R3 and PVQ-RR; Schwartz et al., 2012). All items in the original item pool, shown in Table OS-1 in the Supplementary Online Material (SOM 1). The goal was to select good indicators for the four higher-order values according to substantive (i.e., representing the constructs in breadth) and statistical (i.e., position in MDS; measurement invariance) criteria. The primary selection criterion was the items' position in MDS space, prioritizing items that clearly represented a single higher-order value in that they were close to one of the four poles in two-dimensional MDS space.

Using (unpublished) preliminary analyses in a sample from seven countries (N=8,062), which we detail in SOM 1, we identified 31 candidate items based on MDS and on CFA-based measurement invariance tests. We ultimately selected 17 items for the short scale to represent the four higher-order values according to these criteria. The items selected for measuring the four higher-order values represent nine of the 10 basic human values; we omitted Hedonism because, as noted earlier, the Schwartz theory does not assign hedonism unequivocally to one of the four higher-order values (see Figure 1). The 17 selected items (1) measure each of the four higher-order values with at least three items and (2) represent each of the nine values with at least 1 item. They also reflect the subdivision of Self-Direction into Thought and Action introduced by Schwartz et al. (2012). Table 1 shows the wording of the 17 items in English, along with the higher-order value they were selected to measure and the basic human value they represent.

To fit the needs of mixed-mode surveys, each item consisted of a short verbal portrait of a person. The portraits described a person's desires, wishes, and aspirations that expressed the goal of one of the basic values. Respondents were asked to rate how similar the person described in each portrait was to themselves on a fully labeled six-point scale. Unlike the original PVQ, items were not gender-matched but contained both male and female pronouns ("he/she is"). Although this makes the items slightly less readable, it permits the items to be fielded in identical ways in both online surveys and paper-pencil surveys, a requirement for the mixed-mode GESIS Panel.[1]

We piloted these items in a heterogeneous online sample of adults ($N = 520$). Despite the global nature of the higher-order values and the small number of items per subscale, internal consistencies were satisfactory ($.62 \leq \omega \leq .70$), as was the factorial validity of each subscale. The scale also yielded the expected relations with established value scales in terms of convergent and discriminant validity. We provide further details on the item selection steps in the cross-national sample and the key results from the unpublished pilot study conducted in Germany (Beierlein et al., 2014) in SOM 1. However, results from this pilot study are unpublished. The large-scale, high-quality data from the GESIS Panel have not yet been used to validate HOVS17. We evaluated the psychometric properties of HOVS17 in the present study.

The present study

To evaluate the psychometric properties of HOVS17, we analyzed secondary data from the GESIS Panel, for which the inventory was originally developed. These analyses reveal

[1]Although not yet been implemented in the GESIS Panel, items could be adapted to include a third gender pronoun or only one, gender-neutral pronoun (e.g., "they/them"). Given the aim of maximizing respondents' identification with the individual in each portrait, it is best to use gender-matched language when possible.

Table 1. Wording of the 17 items and assignment to the (higher-order) values.

Item	Wording (English)	Basic value
Conservation		
12CO	It is important to her/him to obey all laws.	Conformity (rules)
7CO	It is important to her/him that the state is strong and can defend its citizens.	Security (societal)
4CO	It is important to her/him to maintain traditional values and ways of thinking.	Tradition
Openness to Change		
3OC	It is important to her/him to develop her/his own opinions.	Self-direction (thought)
8OC	It is important to her/him to expand her/his knowledge.	Self-direction (thought)
14OC	It is important to her/him to be free to choose by herself/himself what s/he does.	Self-direction (action)
17OC	It is important to her/him to figure things out herself/himself.	Self-direction (action)
10OC	It is important to her/him to have all kinds of new experiences.	Stimulation
Self-Enhancement		
2SE	It is important to her/him to show that her/his performance is better compared to the performance of other people.	Achievement
15SE	It is important to her/him that people recognize what s/he achieves.	Achievement
6SE	It is important to her/him to be rich.	Power (resources)
11SE	It is important to her/him to be the one who tells others what to do.	Power (dominance)
Self-Transcendence		
9ST	It is important to her/him to help the people dear to her/him.	Benevolence (caring)
13ST	It is important to her/him to concern herself/himself with every need of her/his dear ones.	Benevolence (caring)
1ST	It is important to her/him to care for nature.	Universalism (nature)
5ST	It is important to her/him to be tolerant toward all kinds of people and groups.	Universalism (tolerance)
16ST	It is important to her/him that everyone be treated justly, even people s/he doesn't know.	Universalism (concern)

Notes: The item labels refer to the item's order in the inventory (number) and the higher-order value it measures (letter). E.g., 12CO is the 12th item and measures conservation. The table shows the English wording of the item, whereas the GESIS Panel, which we analyzed in this study, used the German translations (see SOM).

how appropriate HOVS17 is for use both by secondary data users of the GESIS Panel and by researchers who wish to include a measure of higher-order values in their primary data collection.

For each subscale, we tested its unidimensionality, reliability (including one-year test–retest reliability), and factorial validity (fit of unidimensional CFA models). We also evaluated the structural validity of the inventory as a whole by means of MDS (the preferred approach to testing the structure of values in this research tradition).

Moreover, we computed correlations of the four scales with several correlates and criteria that are prominent in value research and neighboring fields. Specifically, we report associations between values and the following correlates:

1. Sociodemographic characteristics (age, gender, educational attainment, and income) to test the sensitivity to group differences of the four value scales (e.g., Gouveia et al., 2015; Milfont et al., 2016)
2. Big Five personality traits, to test the discriminant validity of the value scales (e.g., Parks-Leduc et al., 2015). We expected that the four higher-order values are only moderately correlated with Big Five personality traits, which would attest to the discriminant validity in relation to personality. Based on Parks-Leduc et al. (2015), we expected stronger associations ($|r| > .20$) for of Openness to Change (positive) and Conservation (negative) with Open-Mindedness; and of Self-Transcendence (positive) and Self-Enhancement (negative) with Agreeableness.
3. The importance attached to different life areas, to test the convergent validity with specific values.
4. Political orientation (e.g., Piurko et al., 2011; Schwartz et al., 2010), to assess whether values are related to how individuals place themselves on the left-right continuum,
5. Subjective well-being (e.g., Grosz et al., 2021; Sortheix & Schwartz, 2017) to test their convergent and criterion validity.

Although this step was mostly exploratory, we preregistered specific hypotheses for some of these associations. The SOM 2 lists these hypotheses, and the full preregistration can be accessed at OSF at https://osf.io/whqfv. Because Schwartz's model is a quasi-circumplex model, we expect that opposite values (i.e., Openness to Change vs. Conservation and Self-Enhancement vs. Self-Transcendence) generally have opposite associations with these outcomes. Such patterns would lend support to the validity of value measurement by the HOVS17.

Methods

Open science statement

All data used in this study are publicly available after registration from https://www.gesis.org/gesis-panel/data. The complete R code used for this project is available from the first author's GitHub repository at https://github.com/CleLec/HOVS17. Moreover, full details on all measures used can be found in the GESIS Panel documentation at https://www.gesis.org/en/gesis-panel/documentation. The scale construction process that led up to the inventory in its form analyzed here are documented in SOM 1, and preregistered hypotheses are documented in SOM 2.

Data

The GESIS panel is an ongoing, probability-based, longitudinal mixed-mode survey of German-speaking adults aged 18

to 70 years who reside in private households that are registered in Germany (Bosnjak et al., 2018). Informed consent and ethical approval were obtained in accordance with German law from the institutional review board at GESIS. Sampling was done in a two-stage procedure; municipalities were randomly sampled in the first stage and individuals from these municipalities in the second stage. All members of the probability sample ($N = 21{,}870$) were first contacted through a pre-notification letter and then personally by an interviewer (minimum of four contact attempts). Of the probability sample, 7,599 participated in a computer-aided face-to-face recruitment interview. Those (more than half) who expressed their willingness to continue their participation (more than half) were taken into the panel. They could choose to participate *via* an online questionnaire or *via* a paper-and-pencil questionnaire sent *via* postal mail. Thirty-eight percent of the panelists chose to participate in the survey *via* mail (for more details on sampling, recruitment rate, and sample composition, see Bosnjak et al., 2018).

The GESIS Panel now comprises three cohorts, an original cohort and two refreshment cohorts. We analyzed Cohort A, the panel's first and largest cohort. We focused mainly on data from Wave "bd" (2014; $n = 4{,}035$) and Wave "cd" (2015; $n = 3{,}519$), the first and second wave in which HOVS17 was fielded. All our analyses are based on the first of these waves, except for our analyses of the one-year test–retest stability of values, which required the second wave. We took additional measures (e.g., basic demographic information and correlates) from the initial screening interview (Wave "a12"; 2013/2014) and from waves in between the two main annual waves which measured values (for details, see Measures and the R code in the first author's GitLab repository).

Measures

HOVS17

Since 2014, starting with wave "bd," the GESIS Panel uses letters as wave codes. The first letter denotes survey years, e.g., "b" refers to waves fielded in 2014, "c" to waves fielded in 2015. The GESIS Panel includes HOVS17 as part of its annual core module, which is typically fielded in September of the respective survey year. Table S1 in the Appendix provides the German version of HOVS17 fielded in the GESIS Panel.

The instructions read as follows: "We will now describe different people to you. Please read each description and think about the extent to which each person is similar to you or not. Please indicate how similar the person in each description is to you." Respondents indicated how similar the person described in the item was to them on a fully labeled six-point rating scale (1 = *is not at all similar to me*; 6 = *is very similar to me*).

Correlates and criteria

Sociodemographic characteristics. We included age in years, educational attainment (1 = university entrance certification (German Abitur); 0 = no university entrance certification), gender (1 = male; 0 = female), and monthly net income (measured in 15 steps that yield an approximately normal distribution).

Big Five. Personality traits were measured with the Big-Five Inventory-10 (BFI-10, Rammstedt & John, 2007). The BFI-10 measures each Big Five dimension–Open-Mindedness (or Openness to Experience), Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (i.e., the reverse of Neuroticism)–with 2 items. Despite its brevity, the BFI-10 performs well in terms of predictive validity compared to longer inventories (Rammstedt et al., 2021). We used data from the same wave in which values were assessed (wave "bd"). Test–retest reliabilities across six weeks have been shown to range from .58 for Agreeableness to .84 for Extraversion (Rammstedt & John, 2007).

Importance of life domains. Respondents indicated how important each of seven life domains were to them on a scale ranging from 1 ("very unimportant") to 7 ("very important"). The life domains were work, family, leisure, friends, neighbors, financial situation, and health. We used data from wave "ba."

Political orientation. Respondents indicated their political orientation on the left—right spectrum on a scale from 1 ("Left") to 7 ("Right") in wave "bc."

Subjective well-being. We used two indicators of subjective well-being: General life satisfaction and happiness. For life satisfaction, respondents in wave "bd" were asked to indicate, on a scale from 0 ("extremely dissatisfied") to 10 ("extremely satisfied"), how satisfied they were with their life at present. For happiness, respondents in wave "bb" indicated how happy they were, all things considered, on a scale from 0 ("extremely unhappy") to 10 "extremely happy").

Analyses

Our analyses of HOVS17 comprised four main steps. First, we computed descriptive statistics on the subscale and item level as well as intercorrelations of the four subscales. Second, we assessed the internal consistency and one-year test–retest stability reliability of the four scales.

Third, we comprehensively tested the structural validity of each subscale and the full inventory. Toward that end, we first assessed the underlying dimensionality of each subscale and the full item set. We then tested the structural (factorial) validity of the four subscales by fitting single-factor CFA models to each. We assessed the structural validity of the inventory as a whole using MDS, the preferred methodology in this research tradition (Schwartz, 1992, 1994;

Schwartz et al., 2012). MDS imposes no constraints on the associations among variables and allows to analyze (as well as visually inspect) the interdependencies between values. We additionally tested the measurement invariance of the four scales across age groups, gender, and educational strata by means of multiple-group CFA.

Fourth and finally, we assessed the nomological network and criterion validity of the scales by correlating them with the abovementioned correlates and criteria. We used R (Version 4.2.1; R Core Team, 2022) for all our analyses. The Results section provides additional details on the procedures.

Results

Descriptive statistics and scale intercorrelations

We first computed descriptive statistics of the observed scale scores. The density plots in Figure 2 show the distribution of these scores. Table S2 in SOM 3 lists detailed descriptive statistics for both the raw and centered scale scores. Additionally, Table S4 provides descriptive statistics on the item level. Given the high quality of this sample, these descriptive statistics can be used as benchmarks against which to compare future data.

Openness to Change was the most highly endorsed higher-order value, whereas Self-Enhancement was the least highly endorsed. Self-Enhancement and Conservation were not only less highly endorsed but also showed greater variance than Self-Transcendence and Openness to Change. All higher-order values were negatively skewed except for Self-Enhancement. This pattern of means and skew likely reflects the fact that all values, per definition, are personally and socially desirable end states, although some (especially Self-Enhancement) less so than others.

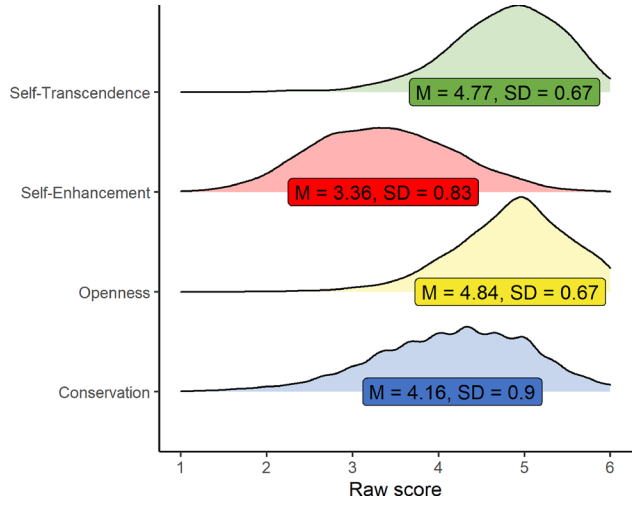

Figure 2. Distributions of the observed raw scale scores in wave "bd" (2014; $n = 4{,}035$).

Table 2 shows the correlations between the observed scores based on centered value scores (upper triangle) and uncentered scores (lower triangle). Correlations among the raw scale scores, as is common, were all positive, albeit mostly moderate in size, in this case, even between the theoretically opposed Openness to Change and Conservation values (r = .30). This indicates the presence of a strong general (response style) factor. This response style factor, common with value measurement, is understood to reflect socially desirable responding due to the desirability of values, content-independent individual differences in the use of the response scale (acquiescence), the overall importance of values to individuals, or a combination of these influences (e.g., Rudnev, 2021; Schwartz et al., 1997). Importantly, after centering the items at the within-person mean, the familiar pattern emerged with great clarity (upper triangle), as is typical with values data (Schwartz et al., 1997, 2012): Openness to Change correlated negatively with Conservation, its opposing higher-order value, and Self-Transcendence correlated negatively with Self-Enhancement, its opposing higher-order value. The changes in the patterns of correlation reflect the fact that centering removes bias introduced by the general response style (see, e.g., Rudnev, 2021; Schwartz et al., 1997, 2012).

Reliability

We computed two measures of scale reliability, Cronbach's alpha (α) and McDonald's omega (ω). Both are measures of the reliability of a unit-weighted scale score–the type of score most value researchers employ. Coefficient α assumes at least an essentially τ-equivalent model, whereas ω assumes a τ-congeneric model and can be used even if there are correlated errors (Widaman & Revelle, 2023; Zinbarg et al., 2006). We mainly focused on ω because it is more adequate for heterogeneous and τ-congeneric scales. Additionally, we estimated test–retest stability over one year. Unlike internal reliability, test–retest stability does not assume reflective measurement. It indicates the rank-order stability of scale scores when scales are administered repeatedly.

Table 3 shows the reliabilities of the four higher-order value dimensions. Internal reliabilities (ω) ranged from 0.55 for Conservation to 0.81 for Openness to Change. One-year test–retest stabilities were more similar across the four dimensions, ranging from 0.65 to 0.69. Test–retest reliability

Table 2. Correlations among the four scale scores based on centered (upper triangle) and uncentered (lower triangle) value scores.

	Openness to Change	Conservation	Self-Transcendence	Self-Enhancement
Openness to Change	–	**–.46**	.02	–.42
Conservation	.30	–	–.26	–.19
Self-Transcendence	.58	.34	–	**–.65**
Self-Enhancement	.24	.26	.04	–

Notes: The upper-right diagonal shows the correlations of the centered scale scores obtained by subtracting the with-person mean across all items. Correlations between theoretically opposed values are marked in bold. The lower-left diagonal shows the correlations of the raw (uncentered) scale scores, which are biased by response style variance. Total $n = 4{,}035$.

was higher than internal reliability for Conservation but similar or lower for the other dimensions.

Structural validity

Dimensionality

Although higher-order values were conceived as regions in MDS space (Schwartz, 1992, 1994) rather than unidimensional constructs, one goal in constructing HOVS17 was to obtain internally consistent, ideally unidimensional subscales for each higher-order value because some researchers may be interested in using the value scales from that perspective (e.g., Grosz et al., 2021). We therefore tested the unidimensionality of the subscales. Unidimensionality holds if a single latent variable underlies a set of items (Hattie, 1985; Ziegler & Hagemann, 2015). Only if unidimensionality holds can scale scores be unambiguously interpreted as reflecting the target construct.

To test unidimensionality, we used traditional parallel analysis (PA; Horn, 1965), the minimum average partial test (MAP; Velicer et al., 2000), and the empirical Kaiser criterion (EKC; Braeken & van Assen, 2017). These tests offer complementary information for assessing dimensionality and all perform well according to simulation studies (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017). Although none of them necessarily performs better than the other two in all empirical scenarios, we gave EKC the greatest weight because EKC performs consistently as well as or better than other indices for short scales like HOVS17 (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017). Additionally, we inspected the first eigenvalue and the ratio of the first to second eigenvalue because larger values are generally preferable, although there are no universal cutoffs for how large eigenvalues or their ratio should be.

The results in Table 4 suggest that all four scales are unidimensional. Our preferred indicator, EKC, always suggested a single factor. There was always a dominant first eigenvalue. For Conservation, MAP indicated no dominant factor, which is likely due to the small number of items (i.e., three), but all other indices pointed to one factor.

We applied the same tests to the dimensionality of the full item set. Here the goal was to test whether the full item set reflected four factors, because HOVS17 was intended to measure four higher-order values. Results applying EKC (our preferred indicator) and PA suggested four factors, whereas MAP suggested only two major factors. This may reflect differences in the three indices' sensitivity to minor factors and the strengths of the expected first general factor (i.e., the response style factor).

Table 3. Reliabilities of the four higher-order value scales.

	Internal consistency		Test–retest
Dimension	α	ω	(One year)
Self-Transcendence	.68	.76	.65
Self-Enhancement	.65	.69	.69
Openness to Change	.76	.81	.66
Conservation	.55	.55	.67

Note. We computed alpha and omega in wave "bd" (2014; $n = 4{,}035$) and the test–retest stability from this wave to wave "cd" (2015; $n = 3{,}519$).

In sum, 17 items appear to reflect four unidimensional higher-order factors. This suggests that the item selection for HOVS17 achieved relatively homogeneous dimensions, which is a prerequisite for modeling each value with a single-factor CFA, which we tested next.

CFA models for individual subscales

We fit a single-factor CFA model to each of the four subscales, conceived as unitary reflective constructs, to test their factorial validity. We estimated τ-congeneric models in which items can have varying loadings. Because Conservation was measured by only three items, such a model would be just-identified ($df = 0$). To obtain a testable model for Conservation, we therefore constrained the loadings of two items, 12CO and 4CO, to equality. These items measure the basic values of Conformity and Tradition, respectively, which occupy essentially the same segment in the value circle, justifying the choice to give them equal loadings.[2]

We estimated all models with a robust maximum likelihood estimator and full-information maximum likelihood to handle missing data (e.g., Graham, 2009). In line with current conventions for judging model fit (Hu & Bentler, 1999), we relied on the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) to assess model fit. We judged model fit to be acceptable according to the following rules of thumb: CFI > .90 ("adequate") or > .95 ("good"), RMSEA < .05 ("good") or at least < .08 ("adequate"), and SRMR < .05 ("good") or at least < .10 ("adequate"). We stipulated that a model was acceptable when at least two of the three indices passed the cutoffs.

Table 5 presents the fit indices of CFA models for each higher-order value. Model fit was acceptable for three of the value dimensions. For Self-Transcendence (Model D), the model showed borderline fit as judged by CFI and unacceptable according to RMSEA. Closer inspection of the modification indices and residual correlation matrix indicated that the two items measuring benevolence, 9ST and 13ST, violated local stochastic independence. This may reflect shared wording (in the German version, both items contain the exact same phrase to refer to people who are dear to one's heart) and/or that benevolence items share more variance with each other than they share with the Universalism items. We therefore estimated an additional model (Model E) in which we let the two benevolence items covary. Permitting this residual covariance greatly improved model fit. We thus concluded that CFA model fit was acceptable for all higher-order values, especially after permitting one residual covariance for Self-Transcendence.

MDS for the full inventory

Next, we applied MDS to test whether the 17 selected items were located in the two-dimensional space as proposed by the theory. For this purpose, we converted the correlation matrix into a distance matrix. As is conventional, we ran an

[2] A τ-equivalent model in which all items had equal loadings also showed good fit.

Table 4. Dimensionality of higher-order value scales and the full item set.

Dimension	EKC	MAP	PA	First eigenvalue	Second eigenvalue	Eigen ratio
Self-Transcendence	1	1	1	2.21	0.92	2.38
Self-Enhancement	1	1	1	1.98	0.84	2.36
Openness to Change	1	1	1	2.55	0.75	3.38
Conservation	1	0	1	1.58	0.72	2.19
Full item set	4	2	4	4.36	2.07	2.10

Notes: EKC: Empirical Kaiser Kriterion; MAP: Minimum Average Partial Test; PA: Parallel Analysis. EKC is the guiding indicator because it is expected to be most performant for the current data based on simulation studies (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017). MAP and PA are ancillary indicators.

Table 5. Fit of CFA models for each of the higher-order value scales.

Dimension	Model	χ^2	*df*	*p*	CFI	RMSEA	SRMR	BIC
Conservation	A	0.33	1	.568	1.00	0.00	0.00	38,066.79
Openness to Change	B	94.31	5	.000	0.97	0.08	0.03	50,203.66
Self-Enhancement	C	42.13	2	.000	0.98	0.08	0.02	48,133.55
Self-Transcendence	D	234.68	5	.000	0.91	0.12	0.04	54,153.23
Self-Transcendence[a]	E	55.59	4	.000	0.98	0.06	0.02	53,945.09

Note. CFI: comparative fit index; RMSEA: root mean square error of approximation; SRMR: standardized root mean square residual; BIC: Bayesian Information Criterion.

[a]Model E contained one residual covariance between the two items measuring benevolence, 9ST and 13ST.

interval MDS and evaluated the solution using the stress-1 value (e.g., Mair, 2018).

Figure 3 shows the location of the 17 items (colored according to the higher-order value they measure) in a two-dimensional MDS space. The stress value of the two-dimensional configuration was acceptable at .206. The items intended to measure the same higher-order value were consistently located close to each other, forming two sets of opposing poles reflecting the motivational tension of Openness to Change vs. Conservation and of Self-Enhancement vs. Self-Transcendence.

Thus, despite the small number of items (which can lead to less clear MDS solutions), the selected items covered the value circle reasonably well.

Measurement invariance

Using multiple-group single-factor CFAs, we tested measurement invariance (Steenkamp & Baumgartner, 1998) of the subscales across age groups, gender, and educational strata. For Self-Transcendence, we based these tests on Model E in Table 5 (permitting one residual covariance between the two benevolence items). We fitted configural, metric (i.e., equal factor loadings), and scalar (i.e., equal factor loadings and intercepts) models. Metric ensures that the meaning of the four latent variables is identical across all groups. Scalar invariance enables unbiased comparisons of latent means.

We compared the three levels of invariance based on the criteria suggested by Chen (2007) and on the Bayesian Information Criterion (BIC). Chen's criteria stipulate that differences of $\Delta CFI \geq .010$, $\Delta RMSEA \geq .015$, $\Delta SRMR \geq .030$ when comparing a metric to a configural invariance model suggest loading non-invariance. When comparing a scalar to a metric invariance model, differences of $\Delta CFI \geq .010$, $\Delta RMSEA \geq .015$, $\Delta SRMR \geq .010$ suggest intercept non-invariance. If full metric but not full scalar invariance held, we fitted partial scalar invariance models by releasing the equality constraints for the item intercept with the highest modification index; in this manner, we tested whether a partially scalar invariant model with at least two invariant items, which would still allow for mean-level comparisons, would fit.

Table 6 summarizes the measurement invariance results. Table S3 in SOM 3 provides detailed model fit information. As Table 6 shows, based on Chen's (2007) criteria, full metric invariance held for all values and grouping variables; even full scalar invariance was reached across age groups for Conservation, Openness to Change, and Self-Enhancement as well as across educational groups for Conservation. For those values and grouping variables for which full scalar did not hold, partial scalar invariance held after releasing one, two or–in one case–three item intercepts. This implies that comparisons of latent means across all groups are possible for all higher-order values.

External validity

We computed linear correlations between each value dimension and each of the correlates or criteria to assess the nomological net and criterion validity of the four scales. Figure 4 shows these correlations. Table S5 in SOM 3 shows the same coefficients in a tabular view, with 95% confidence intervals. As is recommended and common practice in this research tradition (e.g., Rudnev, 2021; Schwartz et al., 1997), we used within-person centered mean scores for each value dimension. For the interested reader, we also present correlations based on uncentered scores in Table S6 in SOM 3.

Based on meta-analytic effect size guidelines for individual difference research by Gignac and Szodorai (2016), we consider correlations of .10, .20, and .30 as "relatively small," "typical" and "relatively large," corresponding roughly to the 20th, 50th, and 75th percentile of correlation sizes observed in individual differences research. The following summary mainly discusses those 30 out of 76 associations with effect sizes of $|r| \geq .10$. These associations were all statistically significant at $p < .05$ or lower with this sample size (see Table S5).

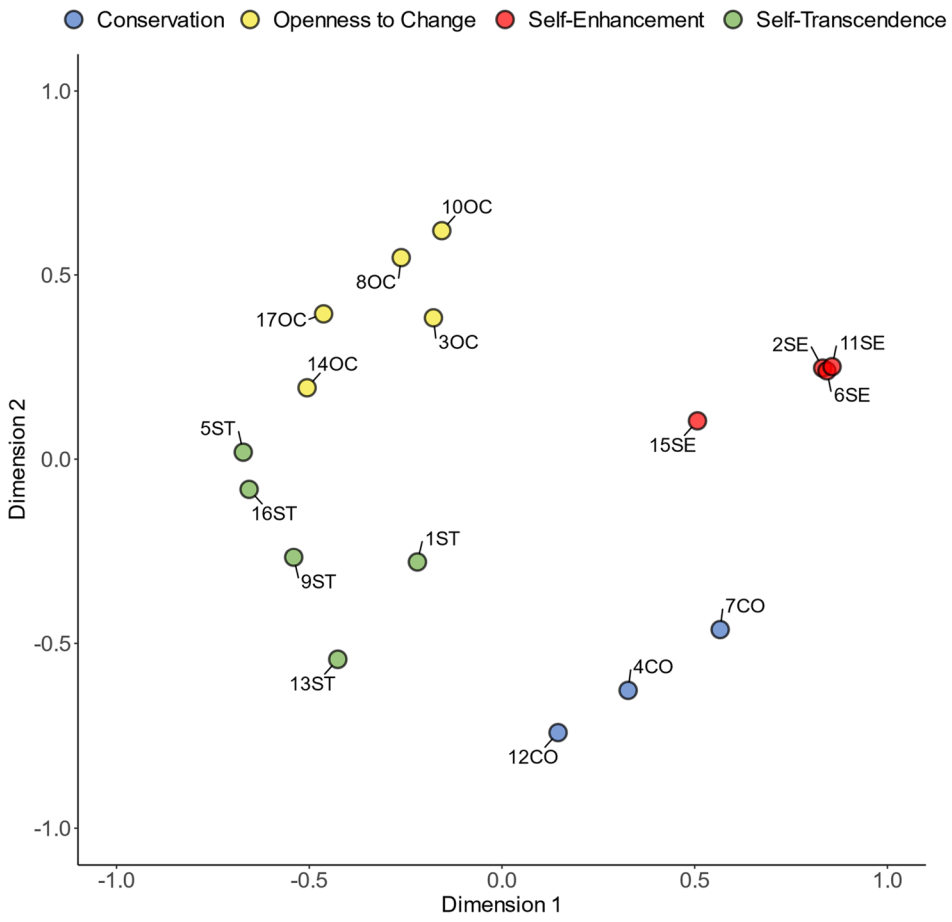

Figure 3. Location of the 17 items in a two-dimensional MDS configuration in wave "bd" (2014; $n = 4{,}035$).

Table 6. Highest level of invariance achieved according for each higher-order value dimension and grouping variable.

Dimension	Group	Invariance level reached with full invariance	Partial invariance
Conservation	Age group	Scalar	(Full scalar)
Conservation	Education	Scalar	(Full scalar)
Conservation	Gender	Metric	Partial scalar
Self-Enhancement	Age group	Metric	Partial scalar
Self-Enhancement	Education	Scalar	(Full scalar)
Self-Enhancement	Gender	Metric	Partial scalar
Openness to Change	Age group	Scalar	(Full scalar)
Openness to Change	Education	Metric	Partial scalar
Openness to Change	Gender	Metric	Partial scalar
Self-Transcendence	Age group	Metric	Partial scalar
Self-Transcendence	Education	Metric	Partial scalar
Self-Transcendence	Gender	Metric	Partial scalar

Notes. Level of invariance achieved according to Chen's (2007) criteria. Partial invariance refers to models with full metric and partial scalar invariance (only relevant for cases in which full scalar invariance did not hold).

There were several value differences across socio-demographic variables. Compared with men, women attached greater importance to Self-Transcendence but less importance to Self-Enhancement. Those with higher education (i.e., general university entrance qualification) attached greater importance to Openness to Change and less importance to Conservation than those with less education. Older people attached more importance to Conservation and less importance to Self-Enhancement than younger people. These effect sizes were relatively small to typical (Gignac & Szodorai, 2016).

There were also several relevant associations between values and the Big Five. Self-Enhancement was related to lower Open-Mindedness, Conscientiousness, and (as hypothesized) Agreeableness. Self-Transcendence was related to higher Open-Mindedness, Conscientiousness, and (as hypothesized) especially Agreeableness. Openness to Change was related to higher Open-Mindedness (as hypothesized) and Emotional Stability. Finally, Conservation was related to higher Conscientiousness but lower Open-Mindedness. These effect sizes were mostly relatively small to typical.

Regarding the relative importance attached to different life areas, Self-Enhancement correlated with viewing work and finances as more important but neighbors as less important. Self-Transcendence correlated with viewing finances as less important but friends and neighbors as more important. Conservation correlated with viewing neighbors as more important but leisure and, unexpectedly, friends as less important. Openness to Change correlated with viewing friends as more important. Most of these correlations were small, though a few were typical.

Associations with political orientation were typical in size. In line with preregistered expectations, respondents

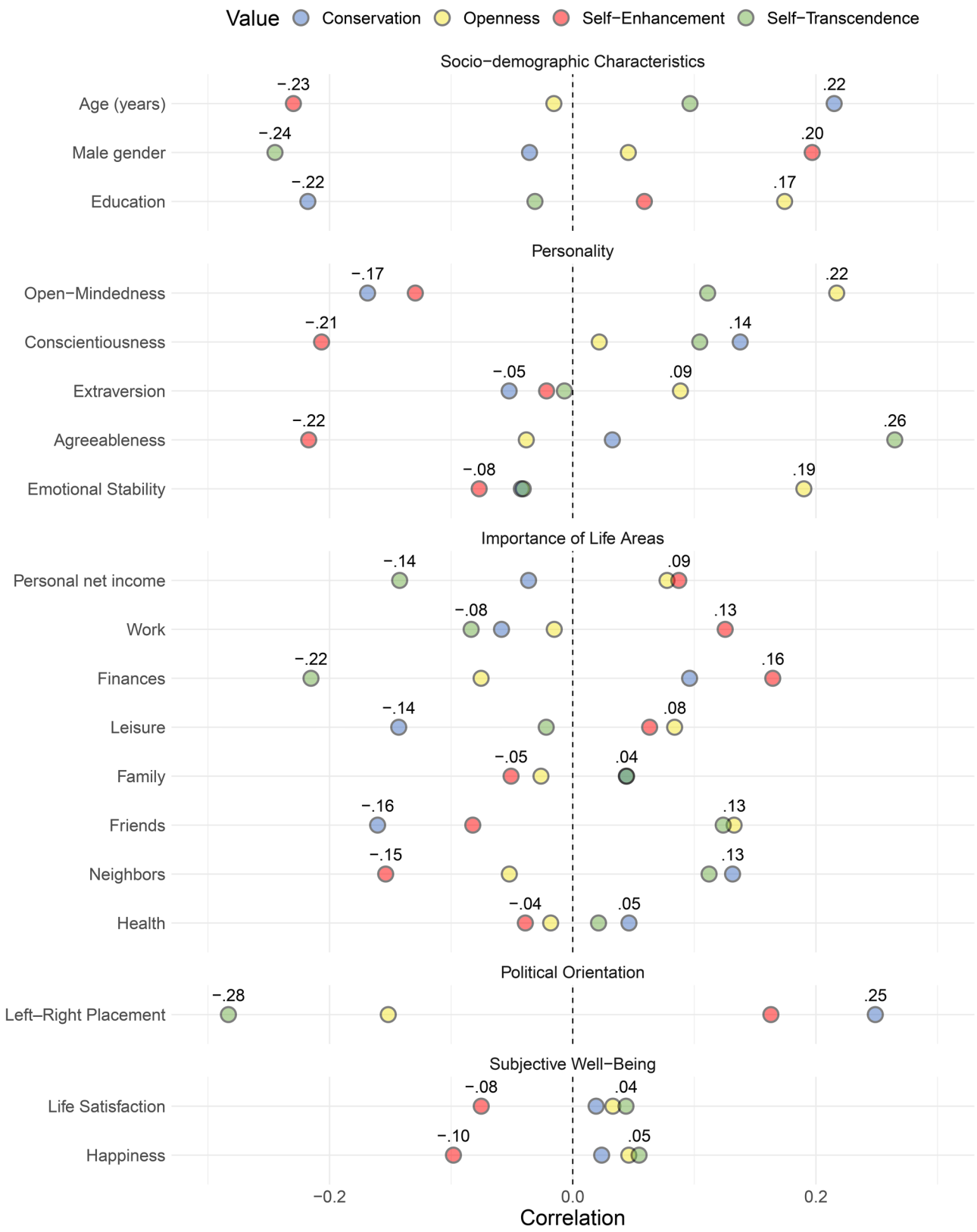

Figure 4. Correlations between the higher-order values and external correlates/criteria. Numeric values of the strongest positive and negative correlation are shown.

who scored higher on Self-Transcendence and Openness to Change located themselves toward the political left and those who scored higher on Self-Enhancement and Conservation located themselves toward the right.

Associations with subjective well-being were very small, though systematic, and in line with preregistered expectations. The more importance respondents ascribed to Self-Enhancement values, the lower their life satisfaction and happiness.

Discussion

This paper evaluated the psychometric properties of HOVS17, a short inventory to assess the four higher-order values in Schwartz's (1992) model of basic human values. Originally developed for the GESIS Panel, this inventory is tailored to the needs of multi-thematic surveys in which questionnaire space is limited. Moreover, the scale is suitable for any research whose focus is on the four higher-order values, rather than the 10 basic human values.

Our results suggest that HOVS17 has satisfactory psychometric properties. All four subscales were clearly unidimensional. Internal consistencies ranged from .55 for Conservation to .81 for Openness to Change. Given (1) the heterogeneity of the value scales, (2) the small number of items, and (3) the purpose of the scale (i.e., research rather than individual diagnostics), we consider these reliabilities to be acceptable (Self-Enhancement) and good (Self-Transcendence and Openness to Change) except for Conservation.

One-year test–retest stabilities ranged from 0.65 to 0.69. Because test–retest stabilities reflect not only unreliability but also true change and state fluctuations, we consider these stabilities acceptable. Of note, they are similar to those for the global Big Five dimensions measured with the ultra-short Big Five Inventory-10 (Rammstedt & John, 2007). For Conservation, test–retest reliability was higher than internal reliability, whereas it was similar or lower for the other dimensions; this suggests that the lower internal consistency of this subscale likely was due to scale length (three items). Using latent-variable models that account for unreliability by explicitly modeling measurement error is an effective way of dealing with limits to reliability. That said, some research applications may benefit from higher reliabilities, (e.g., when values are independent variables and when observed scores are used). This would require adding additional items, especially for Conservation. We provide some suggestions below.

Attesting to the good structural properties of the inventory, the subscales showed good fit when modeled as reflective latent variables in single-factor CFAs. Moreover, the 17 items were located in the two-dimensional MDS space as hypothesized, forming four regions clearly identifiable as corresponding to the four higher-order values. The CFA findings are noteworthy because the higher-order values represent heterogeneous higher-level constructs that were not originally conceived as latent variables through factor-analytic means. Rather, they emerged as segments of a quasi-circumplex model (Schwartz, 1992; Schwartz et al., 2012) and are typically treated as formative indices. These results suggest that the attempt to create short scales that conform to the psychometric standards of unidimensionality, reliability, and factorial validity, when modeled *via* CFA, was successful.

The levels of measurement invariance across gender, age groups, and educational strata varied across the higher-order value dimensions. Full metric invariance held for all dimensions and grouping variables. This implies that the meanings of the latent variables are equivalent across these sociodemographic variables. In a few instances, even full scalar invariance held. The lack of full scalar invariance for some scales and groups implies that the latent means are not fully comparable. Researchers interested in unbiased mean-level comparisons across sociodemographic strata may model partially invariant models in which some intercepts are free to vary across groups; or resort to Bayesian approximate invariance or alignment (e.g., Cieciuch et al., 2014). As additional analyses showed, partial scalar invariance holds for all the values and groups. Alternatively, researchers may prioritize design-based over model-based inference (Robitzsch & Lüdtke, 2022), and therefore choose to use the fully invariant model despite its poorer fit. This choice should be made transparent in any research using HOVS17. We hasten to emphasize, however, that measurement invariance is only relevant for researchers who opt to treat higher-order values as reflective constructs.

The higher-order values exhibited small to typical associations with the selected correlates and criteria (all $r < |.30|$). These associations were generally weaker than those reported in prior research for the 10 basic values, such as the links of basic human values to well-being (Sortheix & Schwartz, 2017) or to personality (Parks-Leduc et al., 2015). They supported only some of the preregistered hypotheses (see https://osf.io/whqfv and SOM 2). Note, however, that these correlations were based on observed scores without correction for attenuation and broad higher-order values than more narrowly focused basic values. Moreover, according to meta-analytic, effect size guidelines (Gignac & Szodorai, 2016), a correlation as small as $|r| = .10$ corresponds to the 20th percentile of correlations in individual-difference research and a correlation of $|r| = .19$ corresponds to the 50th percentile. With this in mind, the results can be seen as providing support for the validity of the higher-order value scales of HOVS17.

Numerous group differences in values and value associations with criterion variables replicated findings with basic or higher order values in the literature. In particular, the scales were sensitive to socio-demographic group differences in values, including age, gender, educational differences and income. The higher-order values were associated with some Big Five traits (Open-Mindedness and Agreeableness), as hypothesized and in line with prior meta-analyses (Parks-Leduc et al., 2015). At the same time, relations were small enough to demonstrate discriminant validity from all personality traits as hypothesized. As hypothesized, the higher-order values were also related to the importance attached to different life domains. The strongest value-related differences emerged in the importance attached to finances, leisure, friends, and neighbors and work but less so to family and health. Moreover, the values were linked to political orientation (left–right placement) meaningfully as in prior research (e.g., Piurko et al., 2011). Associations with well-being were small; the largest associations was a negative correlation between Self-enhancement and happiness ($r = -.10$).

Recommendations for using the scale

We envision that most users of HOVS17 will analyze summative index scores, following the dominant practice in value research. Doing so is compatible with a view of higher-order

values as formative as well as reflective constructs. For conceptual (i.e., converting ratings to value priorities) and psychometric (i.e., eliminating individual differences in the usage of the response scale) reasons, we recommend using scores centered at the within-person mean. This is especially important when values–outcomes associations are of interest. Rudnev (2021) discusses this in-depth and recommends when and when not to center values.

When using summative indices, it should be noted that these are point estimates, not true scores, so they contain measurement error. If the main interest is in the reliable variance shared among the indicators of a higher-order values, rather than the total variance, researchers may prefer to model each higher-order value with CFA (e.g., Grosz et al., 2021). Alternatively, one may model each summative index score as a single-indicator latent variable, using test–retest stability as a reliability estimate.

Researchers might also wish to improve the relatively low internal consistency of the Conservation. It is the shortest and most heterogeneous of the four subscales, having only three items to capture three distinct basic values. Low internal consistency is not a problem if the value is conceived as a formative construct, yet it is if the value is construed from a reflective measurement perspective. That the Conservation test–retest stability was as high as that of the other scales, suggests its lower internal consistency is largely a consequence of the short scale length. Hence, those interested in reflective measurement might want to add one or more items to the Conservation subscale to increase internal consistency and/or emphasize the other components of Conservation such as conformity and security more. Replacing one of the two similarly worded items that violated independence with one or more alternative items is another option. Readers can find two candidate items on the last page of SOM 1, which also introduce new content (Security—Personal and Conformity—Interpersonal).[3] Note, however, that adding items would entail using items that were not as clearly positioned under Conservation in the MDS space. Value items often have relatively high mean endorsement because values refer to *desirable* end states. Users might wish to increase item difficulties by slightly adapting item wording.

In line with recent trends in personality psychology (e.g., Mõttus, 2016), researchers may occasionally be interested in analyzing all 17 single items, rather than aggregate value scores. We caution, however, that the item selection of HOVS17 is neither an exhaustive nor a balanced representation of the basic human values in Schwartz's theory. The purpose of HOVS17 is solely to measure the four higher-order values. One potentially justifiable exception is using single items to maximize predictive power when values are a predictor in multiple regression. However, additional analyses (not shown) suggested that using the 17 individual items as predictors of the criteria selected for the present study increased (adjusted) R^2 only slightly compared to using the four higher-order values. Measurement error in the single items is among the likely reasons for this. We therefore generally advise against analyzing the individual items of HOVS17 in most cases and recommend using the four scale scores.

[3]Readers are invited to contact the second author for guidance in selecting or adjusting items for the other three subscales.

Limitations and future directions

Our study has two main limitations deserving note. First, although we reported correlations with a range of constructs, these are only a small share of potential correlates and criteria. Given the nature of GESIS Panel as a multi-theme survey, most constructs were measured with single items or ultra-short scales that cannot capture their full breadth. This may lead to validity correlations that are biased downward, for example because of attenuation; or biased upward, for example because of missing content (e.g., Chapman & Elliot, 2019; Weiss & Costa, 2014; but see Rammstedt et al., 2021). Moreover, we considered only linear correlations. Future research might utilize statistical approaches that allow for non-linearity (e.g., Lee et al., 2022). It is also desirable to investigate a broader set of outcomes including some that do not depend on self-report.

The Schwartz Values are firmly rooted in cross-cultural research. Against this backdrop, another limitation is that we validated the scale only in one language, German, for which the GESIS Panel provides a uniquely rich data source. It would be desirable to validate the scale in multiple languages to widen its scope of application. The existence of the pool of PVQ-R items (of which HOVS17 is a subset) in many languages should facilitate such an effort.

Conclusion

HOVS17 assesses the four higher-order Schwartz values with 3 to 5 items each (17 items in total) drawn from the PVQ-57. It serves as the annual measure of values in the GESIS Panel. Our analyses demonstrate that HOVS17 has satisfactory psychometric properties. We conclude that the HOVS17 provides a good basis for studying the life-span development, predictors, and consequences of values when focusing on the broad higher-order values, rather than the more fine-grained basic values. We also suggest possible areas of improvement to the scale that users might explore.

Acknowledgments

The authors thank Jan Cieciuch for providing material on the scale construction process. CL would like to thank the GESIS Panel team for helpful insights on the data collection. ED would like to thank the University of Zurich Research Priority Program "Social Networks" for their help during work on this manuscript.

Author contributions

Clemens M. Lechner: Conceptualization, Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing; Constanze Beierlein: Conceptualization, Writing - Review & Editing; Eldad Davidov: Conceptualization, Writing - Review & Editing; Shalom H. Schwartz: Conceptualization, Writing - Review & Editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Open Scholarship

This article has earned the Center for Open Science badge for Preregistered. The materials are accessible with account registration at https://osf.io/mxjw9/.

ORCID

Clemens M. Lechner http://orcid.org/0000-0003-3053-8701
Constanze Beierlein http://orcid.org/0009-0008-4272-0598
Eldad Davidov http://orcid.org/0000-0002-3396-969X

References

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. https://doi.org/10.1037/met0000200

Beierlein, C., Davidov, E., Cieciuch, J., Rammstedt, B., & Schwartz, S. H. (2014, July 14). *Schwartz Values Short Scale (SVSS-4) – An economical measurement tool for cross-cultural research [presentation]* [Paper presentation]. International Congress of the International Association of Cross-Cultural Psychology (IACCP), Reims, France. https://osf.io/fdg3z/files/osfstorage/65608426f0b8e21387a9d7b3

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2018). Establishing an open probability-based mixed-mode panel of the general population in Germany: The GESIS panel. *Social Science Computer Review*, *36*(1), 103–115. https://doi.org/10.1177/0894439317697949

Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Chapman, B. P., & Elliot, A. J. (2019). Brief report: How short is too short? An ultra-brief measure of the big-five personality domains implicates "agreeableness" as a risk for all-cause mortality. *Journal of Health Psychology*, *24*(11), 1568–1573. https://doi.org/10.1177/1359105317720819

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5*, 982. https://doi.org/10.3389/fpsyg.2014.00982

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. https://doi.org/10.1093/poq/nfn035

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gouveia, V. V., Vione, K. C., Milfont, T. L., & Fischer, R. (2015). Patterns of value change during the life span: Some evidence from a functional approach to values. *Personality & Social Psychology Bulletin*, *41*(9), 1276–1290. https://doi.org/10.1177/0146167215594189

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Grosz, M. P., Schwartz, S. H., & Lechner, C. M. (2021). The longitudinal interplay between personal values and subjective well-being: A registered report. *European Journal of Personality*, *35*(6), 881–897. https://doi.org/10.1177/08902070211012923

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(2), 139–164. https://doi.org/10.1177/014662168500900204

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Lee, J. A., Bardi, A., Gerrans, P., Sneddon, J., van Herk, H., Evers, U., & Schwartz, S. (2022). Are value–behavior relations stronger than previously thought? It depends on value importance. *European Journal of Personality*, *36*(2), 133–148. https://doi.org/10.1177/08902070211002965

Mair, P. (2018). *Modern psychometrics with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-93177-7

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*(2), 97–112. https://doi.org/10.1177/1088868314541857

Milfont, T. L., Milojev, P., & Sibley, C. G. (2016). Values stability and change in adulthood: A 3-year longitudinal study of rank-order stability and mean-level differences. *Personality & Social Psychology Bulletin*, *42*(5), 572–588. https://doi.org/10.1177/0146167216639245

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, *30*(4), 292–303. https://doi.org/10.1002/per.2041

Parks-Leduc, L., Feldman, G., & Bardi, A. (2015). Personality traits and personal values: A meta-analysis. *Personality and Social Psychology Review*, *19*(1), 3–29. https://doi.org/10.1177/1088868314538548

Piurko, Y., Schwartz, S. H., & Davidov, E. (2011). Basic personal values and the meaning of left-right political orientations in 20 countries. *Political Psychology*, *32*(4), 537–561. https://doi.org/10.1111/j.1467-9221.2011.00828.x

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rammstedt, B., Lechner, C. M., & Danner, D. (2021). Short forms do not fall short: A comparison of three (extra-)short forms of the Big Five. *European Journal of Psychological Assessment*, *37*(1), 23–32. https://doi.org/10.1027/1015-5759/a000574

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Robitzsch, A., & Lüdtke, O. (2022). Some thoughts on analytical choices in the scaling model for test scores in international large-scale assessment studies. *Measurement Instruments for the Social Sciences*, *4*(1), 1–20. https://doi.org/10.1186/s42409-022-00039-w

Rudnev, M. (2021). Caveats of non-ipsatization of basic values: A review of issues and a simulation study. *Journal of Research in Personality*, *93*, 104118. https://doi.org/10.1016/j.jrp.2021.104118

Sagiv, L., & Schwartz, S. H. (2022). Personal values across cultures. *Annual Review of Psychology*, *73*(1), 517–546. https://doi.org/10.1146/annurev-psych-020821-125100

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*(4), 19–45. https://doi.org/10.1111/j.1540-4560.1994.tb01196.x

Schwartz, S. H. (2003). A proposal for measuring value orientations across nations. In European Social Survey (Ed.), *Questionnaire development package of the European Social Survey* (pp. 259–319). https://www.europeansocialsurvey.org/sites/default/files/2023-06/ESS_core_questionnaire_human_values.pdf

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. https://doi.org/10.1016/S0092-6566(03)00069-2

Schwartz, S. H., Caprara, G. V., & Vecchione, M. (2010). Basic personal values, core political values, and voting: A longitudinal analysis. *Political Psychology*, *31*(3), 421–452. https://doi.org/10.1111/j.1467-9221.2010.00764.x

Schwartz, S. H., & Cieciuch, J. (2022). Measuring the refined theory of individual values in 49 cultural groups: Psychometrics of the Revised Portrait Value Questionnaire. *Assessment*, *29*(5), 1005–1019. https://doi.org/10.1177/1073191121998760

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*(4), 663–688. https://doi.org/10.1037/a0029393

Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Value priorities and social desirability: Much substance, some style. *British Journal of Social Psychology*, *36*(1), 3–18. https://doi.org/10.1111/j.2044-8309.1997.tb01115.x

Sortheix, F. M., Parker, P. D., Lechner, C. M., & Schwartz, S. H. (2019). Changes in young Europeans' values during the global financial crisis. *Social Psychological and Personality Science*, *10*(1), 15–25. https://doi.org/10.1177/1948550617732610

Sortheix, F. M., & Schwartz, S. H. (2017). Values that underlie and undermine well–being: Variability across countries. *European Journal of Personality*, *31*(2), 187–201. https://doi.org/10.1002/per.2096

Steenkamp, J.-B E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–107. https://doi.org/10.1086/209528

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). Springer US. https://doi.org/10.1007/978-1-4615-4397-8_3

Weiss, A., & Costa, P. T. (2014). Personality and all-cause mortality: Individual-participant meta-analysis of 3,947 deaths in 76,150 adults. *American Journal of Epidemiology*, *179*(6), 791–792. https://doi.org/10.1093/aje/kwt300

Widaman, K. F., & Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*(2), 788–806. https://doi.org/10.3758/s13428-022-01849-w

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and loopholes. *European Journal of Psychological Assessment*, *31*(4), 231–237. https://doi.org/10.1027/1015-5759/a000309

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement*, *30*(2), 121–144. https://doi.org/10.1177/0146621605278814