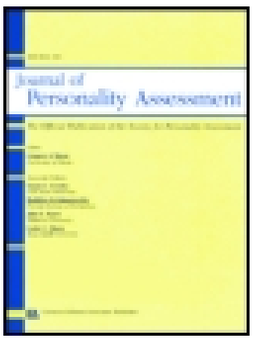

Journal of Personality Assessment

ISSN: 0022-3891 (Print) 1532-7752 (Online) Journal homepage: http://www.tandfonline.com/loi/hjpa20

The Development and Validation of Brief and Ultrabrief Measures of Values

Carson J. Sandy, Samuel D. Gosling, Shalom H. Schwartz & Tim Koelkebeck

To cite this article: Carson J. Sandy, Samuel D. Gosling, Shalom H. Schwartz & Tim Koelkebeck (2016): The Development and Validation of Brief and Ultrabrief Measures of Values, Journal of Personality Assessment, DOI: 10.1080/00223891.2016.1231115

To link to this article: http://dx.doi.org/10.1080/00223891.2016.1231115

View supplementary material

Published online: 21 Oct 2016.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=hjpa20

Download by: [Carson Sandy] **Date:** 21 October 2016, At: 15:42

JOURNAL OF PERSONALITY ASSESSMENT
http://dx.doi.org/10.1080/00223891.2016.1231115

Routledge
Taylor & Francis Group

The Development and Validation of Brief and Ultrabrief Measures of Values

Carson J. Sandy,[1] Samuel D. Gosling,[1,2] Shalom H. Schwartz,[3,4] and Tim Koelkebeck[5]

[1]Department of Psychology, The University of Texas at Austin; [2]Music, Mind and Wellbeing, The University of Melbourne, Parkville, Victoria, Australia; [3]Department of Psychology, The Hebrew University of Jerusalem, Israel; [4]Expert Institute and International Research and Teaching Laboratory for Socio-Cultural Research, National Research University Higher School of Economics, Moscow, Russia; [5]MyType, Inc., San Francisco, California

ABSTRACT

Values are a central personality construct and the importance of studying them has been well established. To encourage researchers to integrate measures of values into their studies, brief and ultrabrief instruments were developed to recapture the 10 values measured by the 40-item Portrait Values Questionnaire (PVQ; Schwartz, 2003). Rigorous psychometric procedures based on separate derivation ($N = 38{,}049$) and evaluation ($N = 29{,}143$) samples yielded 10- and 20-item measures of values, which proved to be successful at capturing the patterns and magnitude of correlations associated with the original PVQ. These instruments should be useful to researchers who would like to incorporate a values scale into their study but do not have the space to administer a longer measure.

ARTICLE HISTORY
Received 30 November 2015
Revised 29 July 2016

There is now strong evidence for a near-universal structure of basic human values (e.g., Bilsky, Janik, & Schwartz, 2011; Schwartz, 1992; Schwartz & Bardi, 2001). For example, the structure of Schwartz's 10 basic values has been supported in 210 samples from 67 countries (Schwartz & Boehnke, 2004). These values (universalism, self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, and benevolence) relate to one another in a quasi-circumplex structure such that values that are adjacent to one another are more positively correlated than values more distant on the circular structure. In addition, the relative ranking of the importance of the 10 values shows substantial consistency across different samples and geographical contexts. This consistency suggests that a pancultural hierarchy fosters successful societal functioning (Schwartz & Bardi, 2001).

The basic values have numerous established external correlates (e.g., age, gender, education level, political orientation, religiosity; Schwartz, 2006) and predict a broad range of meaningful decisions and behaviors (e.g., Bardi & Schwartz, 2003; Sagiv & Schwartz, 2004). In previous research, values have predicted such diverse behaviors as alcohol consumption (Dollinger & Kobayashi, 2003), worrying (Schwartz, Sagiv, & Boehnke, 2000), social attraction (Boer et al., 2011), proenvironmental behavior (Grunert & Juhl, 1995; Schultz & Zelezny, 1998), and political attitudes and voting behavior (e.g., Schwartz, Caprara, & Vecchione, 2010).

Despite such promising findings, the quantity of values-related research has lagged behind that of other individual-difference constructs, such as traits. A keyword search of articles published in the top (based on Impact Factor from the 2012 Journal Citation Reports® Social Sciences Edition [Thomas Reuters, 2013]) personality journals (*Personality and Social Psychology Review, Journal of Personality and Social Psychology, Journal of Personality, European Journal of Personality, Personality and Social Psychology Bulletin, Journal of Research in Personality, Journal of Personality Assessment, Personality and Individual Differences*) over the past 20 years yielded 2,413 articles using the keyword *traits* compared to 411 articles using the keyword *values*.[1]

Some researchers have essentially equated traits with personality (Buss, 1989; Hofstee, 1984) but theorists have been quick to point out that traits alone offer only a limited understanding of personality (McAdams, 1995). In attempts to push more integrative models of personality, researchers have proposed values as a candidate for offering a more comprehensive understanding of the person (e.g., McAdams & Pals, 2006; Parks-Leduc, Feldman, & Bardi, 2015). The incremental contribution of values over traits is supported by their differing patterns and strengths of correlations with outcomes and constructs, such as religiosity (Roccas, Sagiv, Schwartz, and Knafo, 2002), subjective well-being (e.g., Haslam, Whelan, & Bastian, 2009), and voting preferences (e.g., Caprara, Vecchione, & Schwartz, 2009).

A recent meta-analysis of 60 studies reporting correlations between values and the Big Five personality traits suggested that the two types of constructs are distinct (Parks-Leduc et al., 2015); such findings provide an impetus for researchers to measure both values and traits when studying the effects of individual differences. This more comprehensive approach will provide scholars with a more holistic view of the person and

CONTACT Carson J. Sandy cjosandy@gmail.com MotiveMetrics, 410 Sherman Avenue, Palo Alto, CA 94306.
Carson J. Sandy is now at MotiveMetrics in Palo Alto, CA.

Supplemental data for this article can be accessed on the publisher's website.

[1]The search result for *values* overestimates the number of studies examining basic psychological values because the search term also captures additional constructs such as *social value orientation, truth value*, and *mate value*.

© 2016 Taylor & Francis

also contribute to a better understanding of how traits and values differentially affect and predict behavior.

One of the main reasons for the relative scarcity of values research (in comparison to research on personality traits) could be the costs in time and effort of measuring values, especially for investigators for whom values are not their primary research interest. Even the already vibrant field of personality trait research saw an upsurge in studies measuring personality with the availability of very brief, psychometrically sound personality trait measures (e.g., Gosling, Rentfrow, & Swann, 2003). Here, we aim to develop a short measure of values with the hope that it will similarly facilitate a wider integration of values measures into ongoing research. Naturally, researchers whose central interests are in values will continue to use the existing longer instruments.

Background of the empirical study of values

Building on the seminal work of Rokeach (1973), Schwartz and Bilsky (1987, 1990) sought to develop a theory of basic human values that applies across cultural contexts and is grounded in human social nature. Early efforts specified three facets of every value: goal type (terminal vs. instrumental), interests served (collective vs. individual), and motivational domain (e.g., hedonism). Subsequently, Schwartz reduced this to two facets, interests served (social vs. personal) and motivational domain. For example, the value excitement serves a personal interest in the stimulation motivational domain. The development of these facets resulted in a systemic theory of the content and organization of the value systems of individuals that has been validated empirically (Schwartz, 1992; Smith & Schwartz, 1997).

Schwartz (1992) settled on 10 basic human values that are organized on a circular motivational continuum such that adjacent values (values that are closer together in the circle) are conceptually more closely related. Values include universalism, self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, and benevolence values. The circular structure has been supported in adolescent, student, adult, and representative samples from more than 80 countries (Schwartz, 2006, 2011). In addition to replicating the circular structure with self-reported values, Bardi and Schwartz (2003) found that behaviors chosen as likely to express each value (e.g., "obey my parents" [conformity], "Use environmentally friendly products" [universalism]) also exhibited a circular structure of relations that corresponded to the circular motivational structure of values. Thus, value-expressive behaviors validate the motivational structure, too.

Measurement of the Schwartz values model

The 57-item Schwartz Value Survey (SVS; Schwartz, 1992) was created to evaluate these 10 basic human values. Participants endorse the appropriate level of personal importance (from −1 to 7) of each value item. For example, to measure power, a respondent indicates how important "social power—*control over others, dominance*" is as a guiding principle in his or her life. The SVS has strong psychometric validity and remains one of the most pervasive measures of values.

One issue with the SVS was that the values circumplex[2] failed to replicate in about 5% of cross-cultural samples, mainly in Africa (Schwartz et al., 2001). A major reason for this failure is that the SVS requires respondents to have an abstract conception of the values they endorse. The SVS provides no information to contextualize the values in a real-world setting. Some participants, especially in societies with a non-Western education system, might have had difficulty interpreting the abstract items of the SVS.

To circumvent the problem of using abstract value concepts, Schwartz developed the Portrait Values Questionnaire (PVQ–40; Schwartz, 2003; Schwartz et al., 2001). This 40-item instrument presents participants with short portraits of gender-matched individuals. Respondents rate how much the person described in the portrait is like himself or herself. For example, a PVQ power item states, "It is important to him to be in charge and tell others what to do. He wants people to do what he says." Respondents' own values are inferred from their self-reports of how similar they are to the individuals described in the portraits. With its greater concreteness, the PVQ was more successful in confirming the value circumplex in countries where the SVS had not (Schwartz, 2006). The goal of this study is to develop brief and ultrabrief scales, based on the PVQ–40, thereby providing researchers with more options to incorporate values measures into their research.

Why is a shorter instrument needed?

Researchers in the field of personality assessment have identified numerous contexts where short scales are advantageous (e.g., Gosling et al., 2003; Paulhus & Bruce, 1992; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002). Some of these contexts include longitudinal studies (tracking individuals on many constructs over time), personality-rating studies (rating individuals on multiple personality variables), large-scale Internet studies (where participants might lack the patience to take long questionnaires), and prescreening (researchers desire to identify a number of traits before moving forward with a full study).

In response to the demand for short measures, a wide range of very brief scales have been developed and validated. These address such topics as relationships (Wei, Russell, Mallinckrodt, Vogel, 2007), personality traits (Gosling et al., 2003), intelligence (Minshew, Turner, & Goldstein, 2005), and self-esteem (Robins, Hendin, & Trzesniewski, 2001). Typically, these scales demonstrate convergence with the full scale and correlate with external variables as predicted. The short scales tend to have lower reliabilities (e.g., internal consistency, test–retest) than the full scales but remain satisfactory (Ziegler, Kemper, & Kruyen, 2014).

Abbreviating the PVQ–40

In abbreviating the PVQ–40, we had two objectives. First, we wished to create brief (two items per value) and ultrabrief (one item per value) questionnaires to allow researchers to incorporate values into studies that cannot accommodate the length of

[2]Strictly speaking, the circular motivational structure of values is a quasi-circumplex (Schwartz & Boehnke, 2004). For simplicity, we refer to it as a circumplex.

the standard measures. Second, to promote incremental research, we wanted to construct the scales in a way that would readily facilitate comparisons across the scales of different lengths. Third, we wished to provide a record of all relevant selection and validation psychometrics.

A brief 21-item measure of values, derived from the PVQ, already exists, the PVQ–21 (Schwartz, 2003). Schwartz developed this abbreviated scale for use in the European Social Survey (ESS; Schwartz, 2003). A number of issues associated with the development and publication of the PVQ–21 prompted us to create a new scale. First, several of its items were modified versions of PVQ–40 items, making direct psychometric comparisons between the 21-item and 40-item versions problematic. Second, few details were reported regarding the sample of 444 individuals used for the item-derivation sample other than that they originated from the United Kingdom and Netherlands. Third, no external variables were measured in the scale-development process; evaluating the pattern of relationships with values and other constructs is a key step in establishing the validity of a new measure (e.g., Gosling et al. 2003). Fourth, convergent validity scores are not reported. An abbreviated scale's correlation with the original scale is an important step in demonstrating that it can effectively recapture the psychometric properties of the long-form version. Finally, no test–retest data are available; this form of reliability is particularly important because, as we discuss later, internal consistency indexes are vulnerable to underestimating the reliability of short scales.

The current scale-development efforts sought to address these drawbacks of the PVQ–21. First, to facilitate comparability across scales, we drew on PVQ–40 items to construct both the 10- and 20-item scales, using identical methods and equivalent samples. This procedure guaranteed that the items of the 10-item scale were a subset of the items on the 20-item scale. Because the longer instruments perfectly subsume the shorter instruments, researchers can directly compare findings across studies that used either of the two new scales or the existing PVQ–40. We also compared the performance of our new 10- and 20-item scales with that of the existing 21-item PVQ. Second, we present detailed demographic information on participants and documentation of the data collection procedures. Third, a number of external variables were collected, allowing us to examine their expected relationships with the values scales, furnishing further evidence regarding the validity of the new scales. Fourth, convergent validity scores with respect to the original PVQ–40 were estimated and used in the item-selection and validation phases. Finally, we estimated test–retest reliability data.

Overview of this research

Phase 1 of the study involved the selection of items for the new scales. To derive the best performing 10-item and 20-item measures, the following psychometric criteria were used: convergence with the full scale, internal consistency, and patterns of predicted external correlates. Phase 2 then evaluated the abbreviated scales in a new sample. Additionally, two follow-up studies were performed. The first study aimed to identify the time savings a researcher might gain from using the new shorter scales The second study estimated test–retest reliability.

A number of automated methods, such as those based on item response theory and genetic algorithms, can also be used to select the subsets of items that best represent the longer scales from which the items are drawn. Elsewhere we have empirically compared the results of the traditional and automated methods and found the results to be virtually identical (Sandy, Gosling, & Koelkebeck, 2014). Here we use the traditional method, which is more widely understood.

Method

Participants

The procedures required a large sample to detect the subtle psychometric effects of measurement differences, and one that was reasonably diverse to establish the generalizability of the findings. It was also necessary to obtain responses to a number of other variables to evaluate patterns of external correlates. Collecting the data online facilitated meeting these criteria (Gosling & Mason, 2015). Specifically, two samples of volunteers responded to items hosted on an application (MyType) running on Facebook. Respondents received feedback on their scores that they could post to their walls if they wished. Sample 1 (derivation sample) consisted of 38,049 individuals (63% female; ages 18–94, $M = 26.42$, $SD = 10.01$) who responded between December 2009 and November 2010. Sample 2 (evaluation sample) consisted of 29,143 individuals (62% female; ages 18–92, $M = 27.45$, $SD = 9.52$) who responded between August 2010 and February 2011 while Sample 1 was being analyzed. Sample 1 and Sample 2 were both pulled from a larger database of participants and are completely nonoverlapping (meaning that no participants appear in both data sets). The nationality of the participants was as follows: United States (72%), Singapore (8%), Canada (3%), Australia (3%), and Great Britain (3%). Twelve percent of participants were from various other countries and 11% did not report a location.

Measures

The data were collected via a series of questionnaires hosted by MyType.com, an application developed for Facebook. A primary feature of this application is to provide users the opportunity to respond to and get feedback on psychometrically valid personality scales. Participants arrived at MyType.com through embedded links in other Web sites, online search engines, and word of mouth. Two broad classes of measures were collected. First were the values measures from which the short scales were derived. Second were variables that could be used to compare the patterns of external correlates of the old and new instruments. These were demographics, Big Five, political orientation, and religiosity, which MyType happened to be collecting.

Portrait Values Questionnaire 40-item

One survey available through MyType was the PVQ (PVQ–40; Schwartz, 2003; Schwartz et al., 2001), a 40-item questionnaire that includes descriptions of people who endorse certain values. Respondents use a 6-point scale ranging from 1 (*not like me at all*) to 6 (*very much like me*) to describe the extent to which the

individuals portrayed in 40 brief descriptions are similar to themselves (see online supplemental Appendix A for sample items). Respondents' own values are inferred from these self-reported similarity estimates. Reliability reached acceptable levels (typically greater than .7; Nunnally, 1978) in both samples with one exception (tradition). Cronbach's alphas for the selection and evaluation samples were, respectively, conformity = .73/.73, tradition = .51/.51, benevolence = .85/.86, universalism = .84/.85, self-direction = .88/.89, stimulation = .77/.78, hedonism = .80/.81, achievement = .84/.84, power = .69/.69, and security = .65/.64.

Portrait Values Questionnaire 21-item

Schwartz created a 21-item version of the PVQ (PVQ–21; Schwartz, 2003) for inclusion in the ESS (Schwartz, 2003). To evaluate whether the new 10- and 20-item instruments offer any improvement over this version, we also calculated scores for a close proxy of the PVQ–21. It was necessary to use a proxy for the PVQ–21 because eight of its items were slightly modified from the PVQ–40 items. Our data set only included items from the PVQ–40, so we substituted the items from the PVQ–40 that matched the modified PVQ–21 items most closely for our proxy. The differences between the original PVQ–40 and modified PVQ–21 items are quite minor. For example, "Enjoying life's pleasures is important to him. He likes to 'spoil' himself" was modified to "Having a good time is important to him. He likes to 'spoil' himself" in the PVQ–21.

Nonetheless, to evaluate the extent to which these minor edits had an impact on responses to the items, we conducted a series of tests that examined the equivalences between the items. Data were collected via Amazon's Mechanical Turk (see Buhrmester, Kwang, & Gosling, 2011, for a description of Mechanical Turk and support for the reliability of data collected on that platform). With a sample of 307 participants, we evaluated correlations between the PVQ–21 items and their PVQ–40 equivalents. The correlations ranged from .43 to .77 (M = .64). Combining these estimates across studies and correcting them for attenuation (using the test–retest correlations derived from perfectly identical items) yielded item-equivalence estimates that ranged from .53 to .91 (M = .74, SD = .10). Based on these estimates, we concluded that the item-equivalences were sufficiently strong to proceed with our proxy measure of the PVQ–21. However, readers should keep in mind that our proxy is not identical to the PVQ–21. Cronbach's alphas for the selection and evaluation samples for the PVQ–21 were, respectively, conformity = .67/.68, tradition = .18/.16, benevolence = .83/.85, universalism = .73/.75, self-direction = .77/.79, stimulation = .70/.70, hedonism = .67/.67, achievement = .73/.73, power = .55/.55, and security = .40/.39.

The Big Five Inventory

The Big Five Inventory (BFI; John, Donahue, & Kentle, 1991) is designed to measure the factors of the Big Five (Extraversion [E], Agreeableness [A], Conscientiousness [C], Neuroticism [N], Openness [O]). It consists of 44 short items that respondents rate on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Cronbach's alpha reached acceptable levels in both the selection and validation samples, respectively (E = .77/.77, A = .75/.70, C = .78/.77, N = .81/.77, O = .71/.67).

Demographic and attitudinal variables

These variables were age (measured in years), gender, income, education level, political conservatism, and religiosity. Income was self-reported household income (seven intervals ranging from less than $25,000 to $200,000 or more). Political conservatism was measured with a single-item scale ranging from 1 (*extremely liberal*) to 5 (*extremely conservative*). Religiosity was measured with a single-item scale ranging from 1 (*not religious at all*) to 5 (*devoutly religious*).

Phase 1: Item selection

We based item selection on three factors—reliability, patterns of predicted external correlations, and mean external correlations. We measured reliability via Cronbach's alpha, which estimates internal consistency, or the degree to which the items tap the same construct. For scales that are designed to estimate unitary constructs, as the value scales are, items that yield higher alpha levels are preferred. In the case of the 10-item measure, internal consistency could not be evaluated because only one item measured each value. It is important to note that although we used Cronbach's alpha as one of the criteria for selecting items, this index can underestimate reliability, especially when applied to very short measures of broad constructs. This bias results from the fact that alpha is a function of the number of items on a scale and average correlation among the items. Typically the goal of a brief measure is to capture the full breadth of a construct (i.e., content validity) with as few items as possible. To capture the breadth of the construct, maximally different items could be chosen, thereby reducing the mean interitem correlations. To keep the scale short, few items are selected. Together both these factors (low interitem correlations, few items) will tend to diminish the alpha coefficients.

If the new value instruments are to serve as useful alternatives to the PVQ–40, the values they measure should show patterns of correlation with external variables that are similar to those found with the PVQ–40. To evaluate the degree to which this is the case, we performed a three-step analysis. First, we centered each participant's value scores around his or her own scale mean. This is an important step when evaluating external correlations because it controls for response biases such as acquiescence or social desirability (e.g., Schwartz, 2004). Next, we computed the correlations of the values measured with the PVQ–40 with 11 external variables (e.g., personality traits, sociodemographic variables). Finally, we computed the correlations of the values measured with the new instruments with the same 11 external variables. For this analysis, we are concerned with the similarity between the patterns of value correlations of the PVQ–40 and the abbreviated scales. Recapturing the pattern of correlations supports the construct validity of a measure—or its expected relationships with other constructs (whether they be related or unrelated to the construct of interest; Cronbach & Meehl, 1955).

Ideally, the value correlations of the new instruments should demonstrate both convergent and discriminant validity. So, in the second step, we computed column-vector correlations between the sets of correlations (i.e., between the correlations of the PVQ–40 and of the abbreviated scales with the external variables). We first transformed the correlations to a linear

scale using Fisher's *r*-to-*z* formula. Strong vector correlations would indicate that an abbreviated scale captures external predictions in a manner similar to the full PVQ–40.

In addition to reproducing the original pattern of correlations, it is also important that the correlations with the new instrument are of an approximately similar magnitude as the correlations with the original instrument; of course, some reduction in magnitude usually results as a consequence of the somewhat lower reliability associated with short scales. To examine whether the new scales recaptured the magnitude of correlations found with the original scale, the mean of the absolute correlations with external variables was computed for the original and new values scales.

The 10-item and 20-item scales that optimized these three criteria (reliability, patterns of external correlates, magnitude of external correlations) were selected for further evaluation (in Stage 2). The 10-item and 20-item scale are henceforth referred to as the Ten Item Value Inventory (TIVI) and the Twenty Item Value Inventory (TwIVI), respectively.

Results

Table 1 presents the scale reliabilities for the TwIVI, the PVQ–40, and the PVQ–21. Scale reliabilities could not be estimated for the TIVI because it consists of only one item per scale. Reliabilities for the TwIVI, with two items per value, ranged from .33 to .91 (M = .71). Reliabilities for the PVQ–21 proxy ranged from .18 to .83 (M = .62). Reliabilities for the full 40-item PVQ ranged from .51 to .88 (M = .76). Two values (security and tradition) had consistently low reliabilities, even in the full PVQ–40. Not surprisingly, with three to six items per scale, alpha levels for the PVQ–40 were higher on average than for any of the short scales.

The vector correlations revealed that the values measured with the new, abbreviated scales had patterns of correlation with the 11 external variables that were highly similar to the patterns for the PVQ–40. Vector correlations ranged from .78 to .98 (M = .91) for the 10-item TIVI, from .85 to .99 (M = .93) for the 20-item TwIVI, and from .65 to .99 (M = .91) for the PVQ–21 proxy (for full results, see online supplemental Table S.1).

Table 1. Scale alpha reliabilities in the selection and validation samples.

	TwIVI		PVQ–40		PVQ–21.	
Value	Selection	Validation	Selection	Validation	Selection	Validation
Conformity	.61	.60	.73	.73	.67	.68
Tradition	.50	.50	.51	.51	.18	.16
Benevolence	.91	.91	.85	.86	.83	.85
Universalism	.76	.77	.84	.85	.73	.75
Self-direction	.81	.82	.88	.89	.77	.79
Stimulation	.70	.70	.77	.78	.70	.70
Hedonism	.85	.85	.80	.81	.67	.67
Achievement	.79	.79	.84	.84	.73	.73
Power	.80	.80	.69	.69	.55	.55
Security	.33	.33	.65	.64	.40	.39
M	.71	.71	.76	.76	.62	.63

Note. TwIVI = Twenty Item Values Inventory; PVQ–40 = 40-item Portrait Values Questionnaire; PVQ–21 = 21-item Portrait Values Questionnaire. The Ten Item Values Inventory (TIVI) is not represented in this table because internal consistency cannot be estimated for a scale that only uses one item per dimension. Reliabilities from both the selection phase (Sample 1) and validation phase (Sample 2) are presented side by side.

Mean external correlates ranged from .09 to .14 (M = .11) for the one-item scales of the TIVI, from .09 to .14 (M = .12) for the two-item scales of the TwIVI, and from .08 to .13 (M = .11) for the PVQ–21 (for full results, see online supplemental Table S.2). The abbreviated scales almost fully reproduced the magnitude of external correlates seen in the scales of the full PVQ–40 scale, which ranged from .10 to .15 (M = .13). A full summary of external correlates can be found in Table 2. The compositions of the different scales, including the degree to which the derived items overlap with the PVQ–21, are shown in Table 3.

Phase 2: Validation

To ensure that the psychometric criteria derived in the original sample are not capitalizing on correlated error, it is necessary to compute internal consistency reliability, vector correlations, and mean correlations in a new, independent sample. Thus, the goal of Phase 2 was to undertake an evaluation of the scales in a new sample. To evaluate the degree to which each new scale recaptured the full original scale, we also assessed the convergent and discriminant validity between the new short scales and the scales of the original 40-item instrument. Additionally, we tested the ability of the abbreviated scales to recapture the value hierarchy (or the mean ranking of the 10 values) of the PVQ–40.

Measures

The same variables were measured in Phase 2 as in Phase 1. These variables were also measured in the same manner (i.e., the same questionnaires were used).

Validation criteria

It is important to establish that the measures are reliable in an independent sample to ensure that they are sufficiently precise for use in research. We evaluated reliability and external correlations in the same manner as in Phase 1. Predicting the pattern of external correlation (both the magnitude and pattern) is an important step in establishing a nomological network, thereby supporting the construct validity of the instruments (Cronbach & Meehl, 1955).

In addition to examining the pattern and magnitude of the external correlates, we also tested the predicted relationships that the values would have with external variables based on previous literature. For each value dimension, we formed predictions about the directionality of the relationship the dimension would have with the external variables. These predictions were made without reference to the findings in this data set. Informed predictions were made only for external variables with sufficient background research to support the hypothesis. Hypotheses for values and the Big Five are based on meta-analyses performed by Fischer and Boer (2015), Parks (2007), Parks-Leduc et al. (2015), and Roccas et al. (2002). Religiosity hypotheses are based on meta-analyses by Saroglou, Delpierre, and Dernelle (2004) and Schwartz and Huismans (1995). The predictions for sociodemographic variables (e.g., age, gender) are based on Schwartz and Rubel (2005) and on representative

Table 2. External correlations for selection sample.

	CO	TR	BE	UN	SD	ST	HE	AC	PO	SE
PVQ–40										
Extraversion	−.10**	−.12**	.19**	.05**	.20**	.39**	.29**	.30**	.43**	.06**
Agreeableness	.26**	.24**	.57**	.35**	−.07**	.02*	−.02**	−.24**	−.39**	−.01
Conscientiousness	.29**	.22**	.06**	−.02*	−.02*	−.14**	−.15**	.16**	.17**	.45**
Emotional Stability	.01	.04**	.11**	.04**	.11**	.15**	.01	−.04**	−.02**	−.02**
Openness	−.18**	−.14**	.14**	.26**	.38**	.23**	.02**	.06**	.01	−.11**
Conservatism	.23**	.32**	.01	−.24**	−.09**	−.08**	−.07**	−.03*	.01	.18**
Religiosity	.28**	.47**	.23**	.05**	−.07**	.01	−.08**	−.06**	−.04**	.13**
Income	−.01	−.07**	−.04**	−.06**	.01	−.02*	−.01	.08**	.13**	.05**
Education	.02**	−.01	−.01	.01	.01	−.03**	−.04**	−.02**	.02**	.04**
Gender	.03**	.03**	.03**	.05**	−.01*	.01	.01	−.02**	−.05**	.03**
Age	.01	−.01	.05**	.08**	.06**	−.09**	−.07**	−.16**	−.05**	.12**
PVQ–21										
Extraversion	−.11**	−.23**	.18**	.06**	.20**	.37**	.28**	.25**	.30**	−.04**
Agreeableness	.14**	.19**	.49**	.34**	−.10**	−.01	−.05**	−.20**	−.39**	.00
Conscientiousness	.29**	.23**	.08**	−.02*	−.10**	−.11**	−.15**	.07**	.12**	.25**
Emotional Stability	−.04**	−.02	.01	.06**	.06**	.17**	.00	−.10**	−.07**	−.12**
Openness	−.20**	−.17**	.11**	.27**	.36**	.26**	.01	.03**	−.06**	−.16**
Conservatism	.25**	.22**	.00	−.21**	−.09**	−.08**	−.08**	−.03*	.02	.20**
Religiosity	.24**	.23**	.17**	.03**	−.05**	.00	−.10**	−.05**	−.07**	.11**
Income level	.01	−.04**	−.04**	−.06**	.00	−.01	.00	.07**	.12**	.03**
Education level	.02*	−.02**	−.02*	.02*	−.01*	−.02*	−.02**	−.02*	.01	.01
Gender	.02**	.01*	.03**	.05**	−.01**	−.01*	.00	−.00	−.05**	.05**
Age	−.01	.00	.03**	.08**	.05**	−.07**	−.04**	−.14**	−.07**	.08**
TwIVI										
Extraversion	−.06**	.03**	.19**	.05**	.18**	.37**	.31**	.24**	.49**	.08**
Agreeableness	.21**	.13**	.56**	.33**	.01	−.01	.04**	−.27**	−.34**	−.03**
Conscientiousness	.27**	.22**	.06**	−.04**	−.10**	−.11**	−.13**	.17**	.20**	.50**
Emotional Stability	−.00	.00	.01	.00	.07**	.17**	.05**	−.04**	.02*	.00
Openness	−.18**	−.13**	.13**	.18**	.45**	.26**	.03**	−.01	.06**	−.07**
Conservatism	.23**	.38**	−.02	−.26**	−.10**	−.08**	−.06**	−.01	.02	.15**
Religiosity	.28**	.56**	.19**	.03**	−.02*	.00	−.05**	−.07**	.01	.11**
Income level	.00	−.02*	−.03**	−.06**	−.01	−.01	−.03**	.08**	.11**	.05**
Education level	.01*	.02**	.00	−.01	.00	−.02*	−.06**	−.03**	.03**	.05**
Gender	.03**	.05**	.05**	.05**	−.02**	−.01*	.00	−.04**	−.03**	.02**
Age	−.01	.03**	.07**	.05**	.04**	−.07**	−.09**	−.20**	−.01	.11**
TIVI										
Extraversion	.04**	.06**	.19**	.04**	.13**	.36**	.31**	.22**	.39**	.06**
Agreeableness	.20**	.16**	.52**	.29**	.04**	−.06**	.02**	−.21**	−.36**	−.03**
Conscientiousness	.24**	.13**	.05**	−.03**	−.03**	−.10**	−.14**	.10**	.17**	.58**
Emotional Stability	.06**	.04**	.02*	−.00	.09**	.17**	.05**	−.07**	−.04**	.00
Openness	−.12**	−.06**	.12**	.13**	.34**	.20**	.03**	−.03**	−.00	−.06**
Conservatism	.22**	.37**	−.03*	−.19**	−.08**	−.04**	−.05**	−.01	.04**	.10**
Religiosity	.28**	.61**	.19**	.02**	−.02**	.01	−.05**	−.06**	.01	.09**
Income level	−.00	−.03**	−.03**	−.06**	.00	−.01	−.03**	.07**	.09**	.03**
Education level	.02*	.03**	−.00	−.01	.03**	−.04**	−.06**	−.03**	.02**	.06**
Gender	.03**	.04**	.05**	.05**	−.01*	−.04**	.00	−.02**	−.02**	.05**
Age	.02**	.02**	.05**	.04**	.05**	−.09**	−.08**	−.20**	−.03**	.12**

Note. CO = Conformity; TR = Tradition; BE = Benevolence; UN = Universalism; SD = Self-direction; ST = Stimulation; HE = Hedonism; AC = Achievement; PO = Power; SE = Security; PVQ–40 = 40-item Portrait Values Questionnaire; PVQ–21 = 21-item Portrait Values Questionnaire; TwIVI = Twenty Item Values Inventory; TIVI = Ten Item Values Inventory.
*$p < .01$. **$p < .001$.

national data from 27 countries in rounds one and two of the ESS. Hypotheses for values and political conservatism were formed based on the work of Piurko, Schwartz, and Davidov (2011) and Schwartz et al. (2010). Table 4 displays the full matrix of predictions.

In addition, we evaluated the convergent and discriminant validity of the abbreviated scales by correlating the 10 values measured by each scale with the 10 values measured by the PVQ–40. High correlations on the diagonal of the intercorrelation matrices would support convergent validity. Traditionally, near-zero correlations on the off-diagonals support discriminant validity. However, the values form a circumplex, so a different test of discriminant validity is appropriate. Values adjacent in the circumplex (e.g., power and achievement) should correlate positively, opposing values (e.g., power and benevolence) should correlate negatively, and only orthogonal values (e.g., power and stimulation) should correlate near zero. This pattern is found with the PVQ–40 values.

Finally, we evaluated the ability of the new scales to recapture the value hierarchy of the PVQ–40. The value hierarchy, or ranking

Table 3. Overlap of items across measures.

Item	TIVI	TwIVI	PVQ–21
1		SD	SD
2			PO
3	UN	UN	UN
4			AC
5			SE
6		ST	ST
7			CO
8			UN
9			
10	HE	HE	HE
11			SD
12	BE	BE	BE
13	AC	AC	AC
14			SE
15	ST	ST	ST
16		CO	CO
17	PO	PO	PO
18			BE
19			UN
20	TR	TR	
21	SE	SE	
22	SD	SD	
23		UN	
24			
25		TR	TR
26			HE
27		BE	
28	CO	CO	
29			
30			
31			
32			
33		AC	
34			
35		SE	
36			
37		ST	
38			TR
39		PO	
40			

Note. TIVI = Ten Item Values Inventory; TwIVI = Twenty Item Values Inventory; PVQ–21 = 21-item Portrait Values Questionnaire. The letter abbreviations indicate which of the 10 values the item was used to measure: CO = Conformity; TR = Tradition; BE = Benevolence; UN = Universalism; SD = Self-direction; ST = Stimulation; HE = Hedonism; AC = Achievement; PO = Power; SE = Security.

of values within a sample (as established through the sample means), describes the relative value priorities of a given population. There is considerable convergence regarding the relative importance (and unimportance) of the values across countries. Benevolence, self-direction, and universalism typically emerge as the most important values, and power, tradition, and stimulation as the least important (Schwartz & Bardi, 2001). However, the value priorities of different populations tend to vary around this pancultural baseline. Given the importance of the value hierarchy for understanding particular populations, it is important for any short scale of values to be able to effectively capture this hierarchy. We therefore compared the mean value ranking obtained with each shorter scale to that obtained with the PVQ–40.

Table 4. External correlate predictions for validation sample.

Value	E	A	C	N	O	Gender	Age	Religiosity	Conservatism
Conformity	x	+	+	+	−	/	/	+	+
Tradition	−	+	+	+	−	/	/	+	+
Benevolence	+	+	+	x	+	+	+	/	/
Universalism	x	+	+	x	+	/	+	/	−
Self-direction	+	x	+	−	+	/	/	−	−
Stimulation	+	x	−	−	+	/	−	−	/
Hedonism	+	x	−	−	+	/	−	−	−
Achievement	+	−	+	x	x	/	/	/	/
Power	+	−	x	−	−	−	/	/	/
Security	x	x	+	+	−	/	+	/	+

Note. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness; − = negative correlation; + = positive correlation; / = insufficient literature to make a prediction; x = no predicted relationship.

Results

Reliability

The reliability estimates in the validation sample were generally similar to those in the selection sample. Table 1 reports coefficient alphas between .33 and .91 (M = .71) for the values measured by the TwIVI. For the PVQ–21, reliabilities ranged from .16 to .85 (M = .63) and for the PVQ–40 from .51 to .89 (M = .76). As previously noted, reliability could not be estimated for the TIVI because the scales consisted of only one item each.

Correlations with external variables

Each computed scale (PVQ–40, PVQ–21, TwIVI, and TIVI) was correlated with a set of external variables (e.g., Big Five, gender). We compared these correlations with the predicted relationships that we posited based on past research (predictions can be seen in Table 4). As shown in Table 5, the relationships between the full PVQ–40 and the external variables closely matched the predicted relationships in direction and significance. Overall, 82% of the 90 predictions matched both in the predicted direction and significance level. Results for the external correlations of the PVQ–21, TwIVI, and TIVI are also found in Table 5. The results revealed that the TwIVI matched 85% of the initial predictions. The PVQ–21 and the TIVI both matched 83% of predictions. As can be seen in Table 5, many of the predictions that failed were in the proper predicted direction but only met marginal significance.

The full pattern of external correlations was also similar to those in the selection sample. Online supplemental Table S.1 reports vector correlations for the TwIVI of .95 to .99 (M = .97). Vector correlations for the TIVI ranged from .82 to .99 (M = .91). Vector correlations for the PVQ–21 ranged from .79 to .99 (M = .93).

The correlations of values with external variables were generally stronger for the longer, more reliable measures, as psychometric principles would predict. Online supplemental Table S.2 reports absolute mean correlations for the PVQ–40 of .09 to .13 (M = .11), for the 20-item TwIVI of .09 to .13 (M = .11), for the 10-item TIVI scale of .07 to .12 (M = .10), and for the 21-item PVQ–21 of .08 to .12 (M = .10). Table 5 provides full details of the correlations with external variables.

Convergence across measures

The correlations along the diagonals of Table 6 reveal generally strong convergence between the values measured by the abbreviated measures and by the PVQ–40. As expected, the highest convergence emerged for the two-item measures of the TwIVI (M r = .91). Convergent validity for the PVQ–21 values was also strong (M r = .91). The TIVI also performed well with a mean r of .81.

For the reasons discussed earlier, we assessed the discriminant validity of the scales by computing the intervalue correlations of each abbreviated scale with the intervalue correlations

Table 5. External correlations validation sample.

	CO	TR	BE	UN	SD	ST	HE	AC	PO	SE
PVQ–40										
Extraversion	−.08**	−.08**	.19**	.05**	.16**	.36**	.28**	.25**	.37**	.07**
Agreeableness	.26**	.25**	.47**	.33**	−.04**	.01	−.02	−.24**	−.36**	.00
Conscientiousness	.26**	.17**	.05**	−.02	.02	−.11**	−.11**	.17**	.17**	.40**
Emotional Stability	.02	.03*	.08**	.01	.07**	.12**	.03*	−.03*	−.01	−.01
Openness	−.14**	−.11**	.13**	.22**	.31**	.20**	.02	.04**	.00	−.09**
Conservatism	.22**	.30**	.02*	−.24**	−.10**	−.07**	−.06**	−.02*	.01	.17**
Religiosity	.27**	.46**	.20**	.04**	−.06**	.00	−.09**	−.06**	−.04**	.13**
Income level	.00	−.06**	−.03**	−.05**	.02*	−.02	−.01	.07**	.12**	.05**
Education level	.02*	−.01	−.03**	−.02*	−.01	−.05**	−.05**	−.03**	.03**	.03**
Gender	.03**	.03**	.03**	.04**	−.01	.01	.01	−.01*	−.04**	.03**
Age	.00	−.01	.05**	.08**	.06**	−.09**	−.07**	−.15**	−.04**	.11**
PVQ–21										
Extraversion	−.10**	−.19**	.17**	.05**	.16**	.34**	.27**	.20**	.25**	−.03*
Agreeableness	.14**	.22**	.39**	.31**	−.03*	−.02	−.05**	−.21**	−.36**	.01
Conscientiousness	.26**	.20**	.07**	−.02	−.05**	−.08**	−.12**	.08**	.12**	.20**
Emotional Stability	−.04**	−.02	−.01	.03*	.04**	.13**	.01	−.07**	−.05**	−.10**
Openness	−.17**	−.14**	.10**	.23**	.30**	.23**	.01	.01	−.06**	−.14**
Conservatism	.22**	.21**	.01	−.22**	−.10**	−.07**	−.06**	−.03**	.02*	.19**
Religiosity	.23**	.22**	.16**	.02	−.05**	−.01	−.10**	−.06**	−.06**	.10**
Income level	.01	−.03**	−.03**	−.05**	.01	.00	.00	.06**	.11**	.03**
Education level	.02*	−.02*	−.04**	−.01	−.03**	−.03**	−.03**	−.03**	.02*	.00
Gender	.02**	.01	.03**	.04**	−.01	−.01	.01	.00	−.05**	.05**
Age	−.01	−.01	.03**	.07**	.05**	−.07**	−.05**	−.13**	−.07**	.07**
TwIVI										
Extraversion	−.06**	.07**	.19**	.03**	.15**	.34**	.23**	.19**	.44**	.09**
Agreeableness	.22**	.12**	.46**	.30**	.03*	−.02	.03**	−.26**	−.32**	−.02
Conscientiousness	.23**	.17**	.05**	−.03*	−.06**	−.08**	−.10**	.17**	.19**	.45**
Emotional Stability	.01	.01	−.01	−.02	.05**	.13**	.06**	−.02	.03*	.02
Openness	−.14**	−.11**	.11**	.15**	.38**	.23**	.03**	−.01	.06**	−.05**
Conservatism	.22**	.36**	−.01	−.26**	−.11**	−.07**	−.05**	.00	.02	.13**
Religiosity	.27**	.53**	.17**	.02*	−.02*	−.01	−.06**	−.08**	.01	.11**
Income level	.00	−.02*	−.02	−.06**	−.01	.00	−.02*	.08**	.10**	.05**
Education level	.01	.02*	−.02**	−.03**	−.02*	−.03**	−.08**	−.03**	.03**	.05**
Gender	.02**	.04**	.04**	.05**	−.01*	−.01	.01	−.03**	−.03**	.01*
Age	−.02**	.03**	.06**	.05**	.04**	−.07**	−.09**	−.18**	−.00	.11**
TIVI										
Extraversion	.04**	.10**	.18**	.03*	.12**	.33**	.30**	.17**	.34**	.06**
Agreeableness	.21**	.13**	.43**	.26**	.04**	−.06**	.02	−.23**	−.34**	−.01
Conscientiousness	.21**	.09**	.05**	−.03*	.00	−.07**	−.11**	.10**	.16**	.52**
Emotional Stability	.06**	.04**	.00	−.02	.05**	.14**	.05**	−.05**	−.02	−.00
Openness	−.09**	−.05**	.11**	.11**	.30**	.18**	.03*	−.02	.00	−.06**
Conservatism	.22**	.33**	−.01	−.21**	−.09**	−.04**	−.04**	−.01	.03**	.10**
Religiosity	.27**	.59**	.17**	.01	−.02*	−.01	−.06**	−.06**	.01	.10**
Income level	−.00	−.03**	−.03**	−.05**	.01	−.01	−.03**	.08**	.08**	.03**
Education level	.01	.03**	−.03**	−.03**	.01	−.05**	−.07**	−.04**	.02**	.05**
Gender	.03**	.04**	.04**	.04**	−.01	−.03**	.00	−.02*	−.02**	.04**
Age	.01	.03**	.05**	.04**	.04**	−.09**	−.08**	−.18**	−.03**	.11**

Note. CO = Conformity; TR = Tradition; BE = Benevolence; UN = Universalism; SD = Self-direction; ST = Stimulation; HE = Hedonism; AC = Achievement; PO = Power; SE = Security; PVQ–40 = 40-item Portrait Values Questionnaire; PVQ–21 = 21-item Portrait Values Questionnaire; TwIVI = Twenty Item Values Inventory; TIVI = Ten Item Values Inventory. Terms shown in bold italics indicate a conflict with predicted correlations. Terms shown in bold indicate consistency with predicted relationships. Nonbolded numbers indicate that no predictions were made about those correlations.

*$p < .01$. **$p < .001$.

of the PVQ–40 scales. The patterns of correlations for the short scales very closely matched those of the criterion correlation matrix of the PVQ–40. Correlations were .92 (TwIVI), .93 (TIVI), and .92 (PVQ–21).

Value hierarchy

The ranking of the mean score for each abbreviated scale was compared to the ranking of the scores for the PVQ–40 scales. The TwIVI came the closest to recapturing the value hierarchy of the full scale, with only 4 (out of 50 possible) deviations. The PVQ–21 and the TIVI each deviated 8 ranks. Rank-order correlations with the PVQ–40 were .98 for the TwIVI, .94 for the PVQ–21, and .91 for the TIVI. A full report of the ranks and mean differences for each of the abbreviated scales can be seen in online supplemental Tables S.3, S.4, and S.5.

Test–retest reliability

Test–retest reliability statistics are particularly important metrics to gather when evaluating short measures because, unlike Cronbach's alpha, they are not deflated by the low number of items (e.g., Gosling et al., 2003). We therefore conducted a

Table 6. Convergent and discriminant validity with the full PVQ–40 in validation sample.

Scales	CO	TR	BE	UN	SD	ST	HE	AC	PO	SE
PVQ–40										
Conformity	**1**									
Tradition	.62**	**1**								
Benevolence	.36**	.27**	**1**							
Universalism	.24**	.15**	.72**	**1**						
Self–direction	.07**	−.05**	.69**	.67**	**1**					
Stimulation	−.05**	−.07**	.45**	.45**	.59**	**1**				
Hedonism	.05**	−.09**	.47**	.42**	.57**	.61**	**1**			
Achievement	.15**	−.09**	.30**	.26**	.49**	.41**	.45**	**1**		
Power	−.06**	−.17**	−.26**	−.25**	−.09**	.07**	.09**	.44**	**1**	
Security	.49**	.34**	.30**	.28**	.25**	.08**	.22**	.36**	.17**	**1**
TwIVI										
Conformity	**.93****	.59**	.34**	.21**	.07**	−.02**	.07**	.17**	−.02**	.47**
Tradition	.49**	**.82****	.06**	−.10**	−.23**	−.14**	−.16**	−.05**	.06**	.28**
Benevolence	.33**	.23**	**.95****	.70**	.66**	.43**	.46**	.30**	−.23**	.28**
Universalism	.20**	.10**	.65**	**.91****	.60**	.38**	.40**	.24**	−.22**	.23**
Self–direction	.06**	−.04**	.67**	.67**	**.94****	.58**	.52**	.43**	−.13**	.19**
Stimulation	−.08**	−.08**	.40**	.41**	.56**	**.95****	.53**	.38**	.07**	.05**
Hedonism	.09**	−.02**	.55**	.49**	.62**	.65**	**.95****	.43**	−.01	.22**
Achievement	.12**	−.10**	.12**	.09**	.30**	.29**	.34**	**.92****	.51**	.33**
Power	−.03**	−.09**	−.13**	−.14**	.01	.10**	.07**	.41**	**.91****	.16**
Security	.40**	.29**	.14**	.16**	.11**	.01	.07**	.24**	.20**	**.84****
TIVI										
Conformity	**.76****	.54**	.43**	.28**	.18**	.11**	.17**	.17**	−.07**	.42**
Tradition	.38**	**.71****	.13**	−.04**	−.13**	−.06**	−.15**	−.07**	−.01*	.19**
Benevolence	.32**	.24**	**.90****	.67**	.62**	.41**	.42**	.28**	−.22**	.27**
Universalism	.19**	.08**	.66**	**.83****	.62**	.37**	.40**	.26**	−.23**	.20**
Self–direction	.12**	.00	.68**	.67**	**.89****	.53**	.51**	.43**	−.15**	.23**
Stimulation	−.16**	−.09**	.18**	.19**	.31**	**.81****	.39**	.26**	.16**	−.05**
Hedonism	.03**	−.04**	.45**	.39**	.51**	.60**	**.89****	.36**	.02**	.16**
Achievement	.13**	−.10**	.12**	.08**	.25**	.26**	.33**	**.85****	.47**	.27**
Power	−.01*	−.07**	−.25**	−.26**	−.15**	−.02*	−.03**	.31**	**.86****	.12**
Security	.31**	.23**	.12**	.10**	.09**	−.01	.02**	.18**	.15**	**.60****
PVQ–21										
Conformity	**.86****	.53**	.06**	−.03**	−.19**	−.23**	−.15**	.05**	.07**	.39**
Tradition	.59**	**.81****	.26**	.19**	.04**	−.07**	.01	.00	−.14**	.38**
Benevolence	.33**	.23**	**.95****	.70**	.71**	.45**	.49**	.34**	−.21**	.30**
Universalism	.21**	.11**	.74**	**.95****	.71**	.46**	.44**	.28**	−.26**	.25**
Self–direction	.03**	−.06**	.64**	.62**	**.95****	.57**	.56**	.45**	−.07**	.20**
Stimulation	−.08**	−.08**	.40**	.41**	.56**	**.95****	.53**	.38**	.07**	.05**
Hedonism	−.01	−.12**	.34**	.31**	.44**	.54**	**.96****	.40**	.15**	.17**
Achievement	.14**	−.11**	.29**	.23**	.44**	.35**	.44**	**.92****	.38**	.29**
Power	−.05**	−.17**	−.36**	−.34**	−.22**	−.04**	.04**	.36**	**.94****	.14**
Security–21	.45**	.30**	.19**	.15**	.11**	−.06**	.16**	.26**	.14**	**.83****

Note. CO = Conformity; TR = Tradition; BE = Benevolence; UN = Universalism; SD = Self-direction; ST = Stimulation; HE = Hedonism; AC = Achievement; PO = Power; SE = Security; PVQ–40 = 40-item Portrait Values Questionnaire; TwIVI = Twenty Item Values Inventory; TIVI = Ten Item Values Inventory; PVQ–21 = 21-item Portrait Values Questionnaire. The numbers along the diagonal reflect convergent validity—or the correlations between the PVQ–40 and each other scale. The numbers on the upper diagonal reflect the correlations between the PVQ–40 values and the values computed by each other scale. The lower diagonal reflects discriminant validity—or the intercorrelations within each scale. Only the lower diagonal for the PVQ–40 correlations is reported because the numbers mirror the upper diagonal.
*$p < .01$. **$p < .001$.

study to establish the test–retest reliability of the two new proposed measures (the TIVI and the TwIVI).

Participants and method

Two samples (one for the TIVI and one for the TwIVI) of participants were assessed twice, with a 2-week interval between assessments. Data were again collected via Amazon's Mechanical Turk. One hundred and fifteen participants took the TIVI at Time 1 and 76 of them were available to retake it at Time 2. One hundred and twenty-one participants took the TwIVI at Time 1 and 46 of them were available to retake it at Time 2. Demographic variables were not collected for this sample.

Table 7. Test–retest correlations for TwIVI and TIVI.

Value	TwIVI	TIVI
Conformity	.65	.70
Tradition	.58	.79
Benevolence	.69	.73
Universalism	.81	.60
Self-direction	.77	.60
Stimulation	.71	.64
Hedonism	.76	.60
Achievement	.51	.68
Power	.66	.62
Security	.53	.67
M	.67	.66

Note. TwIVI = Twenty Item Values Inventory; TIVI = Ten Item Values Inventory. All correlations were significant at $p < .001$.

Results

Test–retest correlations for both the TwIVI and TIVI can be seen in Table 7. Correlations for the TwIVI ranged from .51 to .81 ($M = .67$). Correlations for the TIVI ranged from .60 to .79 ($M = .66$). These results are in line with similarly abbreviated personality measures (e.g., Gosling et al., 2003). Additionally, in the case of the PVQ–40, Schwartz et al. (2001) reported 2-week test–retest correlations ranging from .66 to .88 ($M = .81$) and Bardi, Lee, Hofmann-Towfigh, and Soutar (2009) reported 9-month test–retest correlations ranging from .58 to .68 ($M = .63$).

Given the importance of value priorities in predicting behavior, profile stability coefficients—or the relative stability of the value priorities from Time 1 to Time 2—were computed (Schwartz & Bardi, 2001). These coefficients were calculated by averaging the intraindividual correlations for the TwIVI and TIVI. The stability coefficient for the TwIVI was .86 ($SD = .15$) and slightly lower at .77 ($SD = .31$) for the TIVI.

Discussion

Our goal was to develop psychometrically sound brief (20-item) and ultrabrief (10-item) measures of values. We evaluated the instruments' convergent and discriminant validities, reliabilities, and external correlations. The TwIVI outperformed the PVQ–21 in almost every area of psychometric evaluation. As expected, the longer measures outperformed the 10-item measure (TIVI). All of the new measures met a satisfactory standard of reliability and validity. Additionally, all of the measures were able to almost completely reproduce the pattern of predicted relationships with external variables.

We wanted to develop measures that made direct cross-instrument comparisons possible, with the ultrabrief measure being a subset of the brief measure, and the brief measure being a subset of the PVQ–40. Our psychometric analyses allowed us to achieve this goal, with the TIVI being completely subsumed by the TwIVI. Specifically, in four cases (at the item-selection phase), there were two candidate items with virtually identical psychometric properties, so we chose the ones that would ensure cross-instrument overlap. The online supplemental Appendices present the TIVI and the TwIVI along with their scoring instructions. Normative data (calculated from the Mytype.com data) for the two new measures can be found in online supplemental Tables S.6 and S.7.

Of the two measures, we recommend using the TwIVI over the TIVI for six reasons. First, the TwIVI has slightly superior psychometric properties (see Table 1 and online supplemental Tables S.1 and S.2). Second, the TwIVI more successfully recaptures predicted relationships between values and external variables. Third, the TwIVI comes closer than the TIVI to duplicating the value hierarchy of the PVQ–40. Fourth, the two items on each scale allow researchers to undertake rudimentary checks for random responding and other rating nuisances and to compute indexes of internal consistency (although researchers are warned against relying too heavily on such measures; see McCrae, Kurtz, Yamagata, & Terracciano, 2011). Fifth, the TwIVI offers greater content validity because there are two items measuring each dimension instead of one. The TIVI remains a viable option for researchers extremely short on time, questionnaire space, or both. Our results showed that, even with a lower internal consistency, the TIVI maintains equally high test–retest reliabilities as the TwIVI.

Limitations and future directions

A number of limitations with this study should be noted. First, our test–retest sample was relatively small and, therefore, generalizability of these findings is limited. Second, our samples were heavily U.S.-based and collected via a Facebook application; future research should examine these scales in a broader sample of countries and using a broader array of assessment methods. Additionally, we were unable to collect nationality information on participants in the test–retest study. This limits the generalizability of those findings. Third, the emphasis on a data-driven approach resulted in the selection of some items that suffer from low face validity. For example, the single item assessing tradition in the TIVI was, "Religious belief is important to him/her. S/he tries hard to do what his/her religion requires," which would appear to miss nonreligious people high on tradition. Nonetheless, we retained such items because the analyses clearly showed them to have the strongest convergent validity with the PVQ–40 scales; had we abandoned our data-driven decisions for these items, we would be compelled to do so for all the scales, unhitching us from a key strength of this work.

Finally, the TwIVI sometimes suffered from low internal consistency. As previously noted, however, we view this more as an apparent limitation than a genuine limitation, because researchers should not rely too heavily on measures of internal consistency when it comes to evaluating the efficacy of abbreviated measures. To attenuate for the bias in alpha reliabilities, they should put more weight on other forms of reliability (e.g., test–retest) and validity (e.g., convergent, external correlates).

Funding

The work of Shalom H. Schwartz on this article was done within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) that was supported by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

References

Bardi, A., Lee, J. A., Hofmann-Towfigh, N., & Soutar, G. (2009). The structure of intraindividual value change. *Journal of Personality and Social Psychology*, *97*, 913–929.

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

Buss, A. H. (1989). Personality as traits. *American Psychologist*, *44*, 1378–1388.

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

Boer, D., Fischer, R., Strack, M., Bond, M. H., Lo, E., & Lam, J. (2011). How shared preferences in music create bonds between people: Values as the missing link. *Personality & Social Psychology Bulletin*, *37*, 1159–1171.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5.

Caprara, G. V., Vecchione, M., & Schwartz, S. H. (2009). Mediational role of values in linking personality traits to political orientation. *Asian Journal of Social Psychology*, *12*, 82–94.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.
Dollinger, S. J., & Kobayashi, R. (2003). Value correlates of collegiate alcohol abuse. *Psychological Reports*, *93*, 848–850.
Fischer, R., & Boer, D. (2015). Motivational basis of personality traits: A meta-analysis of value-personality correlations. *Journal of Personality*, *83*, 491–510.
Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*, 877–902.
Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504–528.
Grunert, S., & Juhl, H. (1995). Values, environmental attitudes, and buying of organic foods. *Journal of Economic Psychology*, *16*, 39–62.
Haslam, N., Whelan, J., & Bastian, B. (2009). Big Five traits mediate associations between values and subjective well-being. *Personality and Individual Differences*, *46*, 40–42.
Hofstee, W. K. B. (1984). What's in a trait: Reflections about the inevitability of traits, their measurement, and taxonomy. *Personality Psychology in Europe*, *1*, 75–81.
John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory—Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality*, *63*, 365–396.
McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, *61*, 204–217.
McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*, 28–50.
Minshew, N. J., Turner, C. A., & Goldstein, G. (2005). The application of short forms of the Wechsler Intelligence Scales in adults and children with high functioning autism. *Journal of Autism and Developmental Disorders*, *35*, 45–52.
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
Parks, L. (2007, April). *Personality and values: A meta-analysis*. Paper presented at the Society of Industrial and Organizational Psychology annual conference, New York, NY.
Parks-Leduc, L., Feldman, G., & Bardi, A. (2015). Personality traits and personal values: A meta-analysis. *Personality and Social Psychology Review*, *19*, 3–29.
Paulhus, D. L., & Bruce, M. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology*, *63*, 816–824.
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*, 537–561.
Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*, 151–161.
Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging*, *17*, 423–434.
Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin*, *28*, 789–801.
Rokeach, M. (1973). *The nature of human values*. New York, NY: The Free Press.
Sagiv, L., & Schwartz, S. H. (2004). Values, intelligence and client behavior in career counseling: A field study. *European Journal of Psychology of Education*, *19*, 237–255.
Sandy, C. J., Gosling, S. D., & Koelkebeck, T. (2014). Psychometric comparison of automated versus rational methods of scale abbreviation: An illustration using a brief measure of values. *Journal of Individual Differences*, *35*, 221–235.
Saroglou, V., Delpierre, V., & Dernelle, R. (2004). Values and religiosity: A meta-analysis of studies using Schwartz's model. *Personality & Individual Differences*, *37*, 721–734.
Schultz, P., & Zelezny, L. C. (1998). Values and proenvironmental behavior: A five-country survey. *Journal of Cross-Cultural Psychology*, *29*, 540–558.
Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advantages and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental and social psychology* (Vol. 25, pp. 1–65). New York, NY: Academic Press.
Schwartz, S. H. (2003). A proposal for measuring value orientations across nations. In *Questionnaire development report of the European Social Survey* (pp. 259–319). Retrieved from http://www.europeansocialsurvey.org/index.php?option=com_docman&task=doc_view&gid=126&Itemid=80
Schwartz, S. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, *38*, 230–255.
Schwartz, S. H. (2006). Basic human values: An overview. Retrieved from http://www.fmag.unict.it/Allegati/convegno%207-8-10-05/Schwartzpaper.pdf
Schwartz, S. H. (2010). Basic values: How they motivate and inhibit prosocial behavior. In M. Mikulincer, & P. R. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 221–241). Washington, DC: American Psychological Association.
Schwartz, S. H. (2011). Values: Cultural and individual. In F. J. R. van de Vijver, A. Chasiotis, & S. M. Breugelmans (Eds.), *Fundamental questions in cross-cultural psychology* (pp. 463–493). Cambridge, UK: Cambridge University Press.
Schwartz, S. H., & Bardi, A. (2001). Value hierarchies across cultures: Taking a similarities perspective. *Journal of Cross-Cultural Psychology*, *32*, 268–290.
Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Social Psychology*, *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., Caprara, G. V., & Vecchione, M. (2010). Basic personal values, core political values, and voting: A longitudinal study. *Political Psychology*, *31*, 421–452.
Schwartz, S. H., & Huismans, S. (1995). Value priorities and religiosity in four Western religions. *Social Psychology Quarterly*, *58*, 88–107.
Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, *32*, 519–542.
Schwartz, S. H., & Rubel, T. (2005). Sex differences in value priorities: Cross-cultural and multi-method studies. *Journal of Personality and Social Psychology*, *89*, 1010–1028.
Schwartz, S. H., Sagiv, L., & Boehnke, K. (2000). Worries and values. *Journal of Personality*, *68*, 309–346.
Smith, P. B., & Schwartz, S. H. (1997). Values. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology* (pp. 77–118). Needham Heights, MA: Allyn & Bacon.
Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The Experiences in Close Relationship Scale (ECR) short form: Reliability, validity, and factor structure. *Journal of Personality Assessment*, *88*, 187–204.
Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales—Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, *35*, 185–189.