Personality and Individual Differences 102 (2016) 204–210

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Measurement equivalence of Schwartz's refined value structure across countries and modes of data collection: New evidence from Estonia, Finland, and Ethiopia

Laur Lilleoja [a,*], Henrik Dobewall [b], Toivo Aavik [c], Micha Strack [d], Markku Verkasalo [b]

[a] School of Governance, Law and Society, Tallinn University, Uus-Sadama 5, 10120 Tallinn, Estonia
[b] Institute of Behavioural Sciences, University of Helsinki, Siltavuorenpenger 1a, 00014 Helsinki, Finland
[c] Georg-Elias-Müller-Institute of Psychology, University of Goettingen, Goßlerstraße 14, 37073 Goettingen, Germany
[d] Department of Psychology, University of Tartu, Department of Psychology, Näituse 2, Tartu 50409, Estonia

ARTICLE INFO

Article history:
Received 12 March 2016
Received in revised form 5 July 2016
Accepted 9 July 2016
Available online xxxx

Keywords:
Values
Measurement invariance
Bi-factor model
PVQ-R

ABSTRACT

Schwartz et al. (J Pers Soc Psychol 103(4): 663–688, 2012) recently proposed a refined theory that describes universal aspects in the content and the structure of human values. We propose a bi-factor model in order to model the common method variance (CMV) imposed by people's response styles, which has been ignored in previous tests for measurement invariance in the revised Portrait Values Questionnaire (PVQ-R).

Four unique samples were used—two from Estonia ($N = 1954$, 82% female, $M_{age=}31.2$ years, $SD_{age} = 13.2$, online mode; $N = 1309$, 69% female, $M_{age=}28.2$ years, $SD_{age} = 9.6$, paper-and-pencil mode), one from Finland ($N = 250$, 80% female, $M_{age=}24.6$ years, $SD_{age} = 6.7$), and one from Ethiopia ($N = 253$, 26% female, $M_{age=}21.8$ years, $SD_{age} = 5.0$)—allowing assessment of the cross-cultural measurement invariance of the value circle, as well as the effect of the increasingly common mixed-mode data collection method.

After taking CMV into account, the refined value structure holds relatively well in the Estonian and Finnish samples, but not in the Ethiopian context. PVQ-R data collected through paper-and-pencil and online modes can be combined, while there exist small limitations for the comparison of relationships and latent means across Estonian and Finnish samples.

1. Introduction

During the last two decades, the theory of basic human values (Schwartz, 1992) has gained popularity among researchers from many different fields. The main attraction of this theory lies in the fact that it has succeeded in describing universal aspects in the content and structure of basic human values, which has made it especially suitable for cross-cultural value research (Schwartz, 1994).

Though the issue is often ignored, meaningful cross-cultural comparisons and mixed mode data collection require that respondents understand and interpret questions in the same manner and use measurement instruments in the same way (Saris & Gallhofer, 2014). Concurrently with the growing popularity of Schwartz's value theory, an increasing amount of research has tested the universality of the proposed value structure and the measurement invariance (i.e., the degree of comparability) of the related value scales across samples, which, in turn, has led to several improvements in the structure and measurement instruments. Recently, a refined value theory was introduced, together with a revised 57-item Portrait Values Questionnaire, the PVQ-R (Schwartz, Cieciuch, Vecchione, Davidov, et al., 2012).

It is well known that value ratings reflect not only values themselves but also people's response styles (He & van de Vijver, 2015; Schwartz, Verkasalo, Antonovsky, & Sagiv, 1997). We argue that inter-individual differences in the extent to which people agree or disagree with presented values items, independent of their content, imposes common method variance (CMV) on the data. CMV is commonly defined as the "variance that is attributable to the measurement method rather than to the constructs the measures represent" (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, p. 879) and this method bias is a potential problem in personality and individual differences research, because it is one of the main sources of measurement error. It has been shown that this bias can be adjusted for by modeling a latent method factor (Welkenhuysen-Gybels, Billiet, & Cambré, 2003), but so far very few values studies have done this (Schwartz et al., 2012; Strack & Dobewall, 2012; Verkasalo, Lönnqvist, Lipsanen, & Helkama, 2009), while none has tested its model for measurement invariance across countries or modes of data collection. To overcome this limitation, we

* Corresponding author.
E-mail addresses: laur@tlu.ee (L. Lilleoja), henrik.dobewall@helsinki.fi (H. Dobewall), toivo.aavik@ut.ee (T. Aavik), mstrack@uni-goettingen.de (M. Strack), markku.verkasalo@helsinki.fi (M. Verkasalo).

http://dx.doi.org/10.1016/j.paid.2016.07.009
0191-8869/

follow Holzinger and Swineford's (1939) advice and apply a latent method factor model, commonly referred to as a bi-factor model, to our data. Bi-factor models have recently been applied to measurements of Big 5 personality traits (Biderman, Nguyen, Cunningham, & Ghorbani, 2011) and self-esteem (Motl & DiStefano, 2002). Furthermore, bi-factor models are known to increase model fit as well as the predictive validity of a given scale (Biderman, 2014). In summary, we expect that Schwartz's refined value structure will become more comparable across countries and modes of data collection when controlling for CMV.

1.1. Schwartz's theory of basic human values

According to Schwartz's value theory, people's value ratings are, in any culture, locatable under a limited number of motivationally-distinct basic value types (Schwartz, 1992). These universal value types have dynamic interrelationships and form a quasi-circular structure, where similar value types (like *hedonism* and *stimulation*) are close to each other and conflicting value types (like *benevolence* and *power*) appear on opposite sides. In recent years, there has been intense discussion about the universality and specification of the value structure (e.g., Fontaine, Poortinga, Delbeke, & Schwartz, 2008; Knoppen & Saris, 2009; Steinmetz, Isidor, & Baeuerle, 2012; Strack & Dobewall, 2012), which has led to a refined 19-factor model (Schwartz et al., 2012). An overview of the 19 value types and their theoretical quasi-circular structure can be found in Fig. 1.

1.2. Measurement equivalence of value structure

A well-established way to test for equivalence of measurement instruments is through Structural Equation Modeling (SEM; Meredith, 1993). Commonly, three levels of equivalence are tested across groups: configural (which maintains an equal factor structure), metric (which sets the loadings of included items as equal), and scalar (which imposes equality constraints on the intercepts of each included item) invariance. Only if the latter two levels of equivalence hold, does the comparison of relationships and means yield valid results (Saris & Gallhofer, 2014). For cases where one or more equality constraints needs to be dropped, the concept of partial invariance was developed (Byrne, Shavelson, & Muthén, 1989). If partial scalar invariance holds (i.e., for a given concept, at least two indicators need to be invariant across groups), then latent means of the scale can still be validly compared.

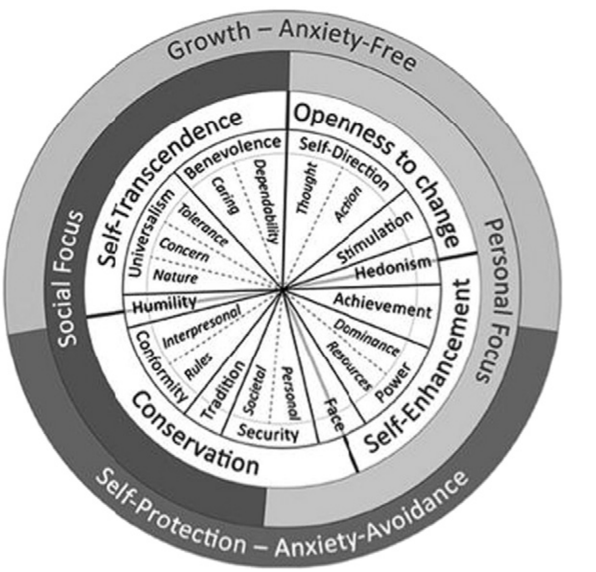

Fig. 1. The circular motivational continuum of 19 values in the refined theory of basic human values.
(Adapted from Schwartz et al., 2012)

Several studies have used Schwartz's instruments to test for cross-cultural measurement invariance for a subset of his value types, but few for the entire value circle at once (e.g., Davidov et al., 2008). Most have reported some form of deviation, sometimes already at the configural level (Saris, Knoppen, & Schwartz, 2013). Cieciuch, Davidov, Vecchione, Beierlein, and Schwartz (2014) tested cross-national invariance of PVQ-R across samples from Finland, Germany, Israel, Italy, New Zealand, Poland, Portugal, and Switzerland. They found full metric invariance for 16 of the 19 values and full or partial scalar invariance for 10 of the 19 values across nearly all countries. It is important to point out, however, that the authors excluded 9 items which did not work in the expected way and they performed equivalence tests for each higher order value separately, not for the entire value circle at once.

In sum, the measurement invariance of the new PVQ-R scale has been tested in only a few studies (e.g., Cieciuch et al., 2014) and in a limited number of countries. These initial results indicate that the new PVQ-R performs better in comparison to earlier Schwartz instruments. However, none of these past studies (e.g., Cieciuch & Davidov, 2012; Lilleoja & Saris, 2014; Schwartz et al., 2012) has tested the measurement invariance of the full Schwartz value structure while including the method factor, which means that CMV among value items is ignored and substance and style are conflated (Strack & Dobewall, 2012; Hinz, Brähler, Schmidt, & Albani, 2005; He & van de Vijver, 2015; Schwartz et al., 1997). In the following, we describe a bi-factor model (Fig. 2) that consists of 19 latent substance factors (i.e., Schwartz's refined value types) which are paired according to a method factor, labelled CMV.

Previous work has shown that response styles vary across countries dependent on their level of socioeconomic development (Smith, 2011; Strack & Dobewall, 2012). This makes it likely that the variance attributable to CVM also varies between the samples studied.

1.3. Mode effects

As in many other fields, mixed-mode data collection is becoming increasingly popular in values research. As each data collection mode has its specificities, these mediate respondents' answers in different ways, and responses to the same questions may not always be comparable (Revilla, 2013). Therefore, it cannot be implicitly assumed that data collected through different modes are necessarily comparable. Although, in the present case, the paper-and-pencil and online questionnaire modes used are both self-administered and have similar formats, we do expect to find measurement equivalence. Two earlier PVQ studies also reported strict invariance across these modes of data collection (Cieciuch, Davidov, Oberski, & Algesheimer, 2015; Davidov & Depner, 2011), but, in this study, we can also test whether the strength of method effects (CMV) differs across paper-and-pencil and online survey modes.

The current study explores measurement equivalence of the new PVQ-R values scale (Schwartz et al., 2012). Unlike previous studies, we also control for the effects of method bias. Applying a bi-factor model has the strength of preserving the quasi-circular structure of Schwartz's human values, and, at the same time, takes into account the CMV among all value items. We use 4 unique datasets for the analyses – two from Estonia (one collected online and the other in paper-and-pencil mode) and one each from Finland and Ethiopia. This enables us to test for both cross-cultural and cross-mode equivalence. These analyses show whether PVQ-R data collected in paper-and-pencil and online modes can be combined and whether relationships and latent means can be compared across these countries.

Our hypotheses are the following:

a) Schwartz's refined values structure is equivalent across countries and modes of data collection when controlling for CMV.

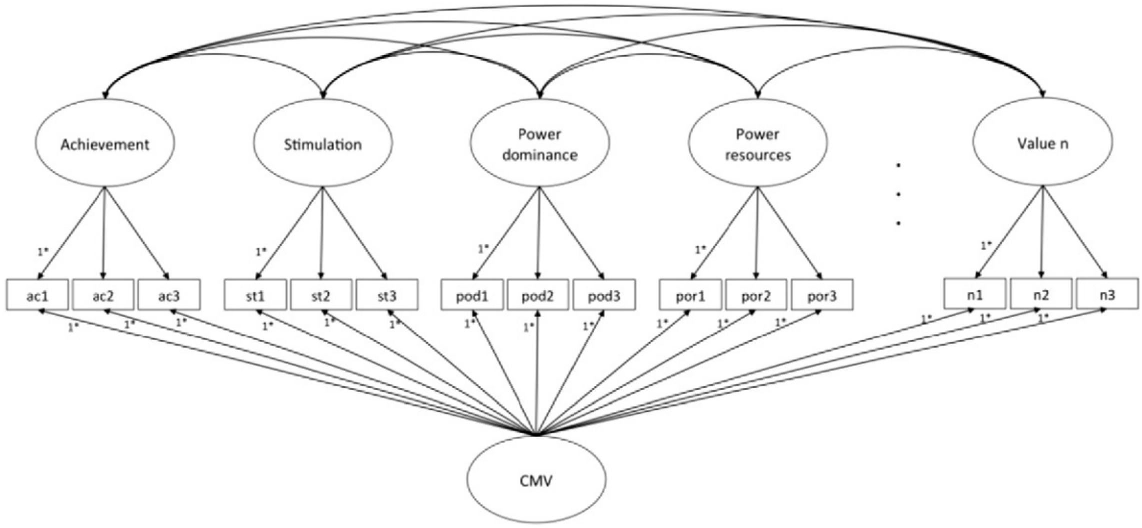

Fig. 2. Bi-factor values model.

b) The variance explained by CMV decreases with the level of socioeconomic development of a country.
c) CMV is similar across self-administered modes of data collection.

2. Methods

2.1. Participants

The dataset used includes 4 sub-samples, with a total of 3766 respondents (Table 1). All data were collected in 2013 and 2014. The respondents were mostly university students. Participation in the study was voluntary and respondents were assured that their data would be kept anonymous.

In Estonia, self-reported data were collected using the Estonian version of PVQ-R[1] online and with paper-and-pencil questionnaires. The existence of two data sets collected using same questionnaire but different modes additionally enables us to test for the mode effect. Finnish data were collected using an online version of the Finnish translation of the PVQ-R. The Finnish sample was used as point of reference as samples from that country usually perform rather well in invariance testing (Schwartz et al., 2012).

The Ethiopian data were collected using a paper-and-pencil version of the PVQ-R questionnaire in English, as Ethiopian higher education is entirely in English. All the respondents were 3rd year undergraduate students from Wolaita Sodo University in South-West Ethiopia. The inclusion of the Ethiopian sample is especially valuable, as, to the best of our knowledge, there currently exists only one PVQ-R study which uses samples from Africa (Tamir, Schwartz, Cieciuch, Riediger, et al., 2015).

2.2. Measurement of basic human values

The 57-item Portrait Values Questionnaire (PVQ-R) includes a larger number of value types than previous instruments. Each of the 19 value types was measured with three items answered along a 6-point scale. This new questionnaire includes only one sentence for each item, not two as in earlier versions of the PVQ. An example of a Self-direction–thought item is: "It is important to her/him to form her/his own opinions and have original ideas". For Estonian and Finnish samples, the reliabilities were moderate to high, and ranged from 0.50 (Humility in both Estonian subsamples) to 0.90 (Conformity–Rules in the Finnish subsample).

2.3. Model evaluation

For the invariance test of the PVQ-R, we conducted a multi-group confirmatory factor analysis (MGCFA). Analyses were conducted in R (version 3.1.2 (2014-10-31)) using the structural equation modeling function of the lavaan package (version 0.5-17; Rosseel, 2012).

As it has been shown that the most commonly used evaluation procedures for structural equation models (global fit indices) are not always reliable (Saris, Satorra, & van der Veld, 2009), we additionally used a program called Jrule (Van der Veld, Saris, & Satorra, 2009; Pornprasertmanit, 2014, for R), which determines whether misspecifications are present in a given model, while taking into account the power of the test (Saris et al., 2009). We aimed to detect standardized deviations of 0.1 in the loadings, and of 0.2 in the intercepts, of the observed variables.

In the first phase of analysis, configural structure was tested separately within each subsample. Based on earlier research (Biderman et al., 2011; He & van de Vijver, 2015; Hinz et al., 2005; Schwartz et al., 1997), we expected potentially strong method effects/CMV. Therefore, in addition to the 19 interrelated latent factors, we included an extra method factor which had equal loading for each item and which was not related with the other factors (Fig. 2). In this bi-factor model, we, further, did not assume that the latent method factor would be invariant across samples (see Smith, 2011).

3. Results

3.1. Factorial structure

In the first step of the data analysis, we tested the structure of the bi-factor values model without imposing any equality constraints across

Table 1
Sample overview.

Sample	Language	Mode	N	Year	M_{age}	SD_{age}	% of female
1	Estonian	Online	1954	2013/2014	31.2	13.2	82
2	Estonian	Paper-and-pencil	1309	2013/2014	28.2	9.6	69
3	Finnish	Online	250	2014	24.6	6.7	80
4	English (Ethiopia)	Paper-and-pencil	253	2013	21.8	5.0	26

[1] We used the PVQ-R3, which has since been replaced with the modified PVQ-RR.

Table 2
Initial configural fit and problematic items.

Sample	Chi^2	DF	RMSEA	CFI	CMV (95% CI)	Problematic items								
						fac3	hum1	sep3	ac1	st2	tr3	unc1	he3	ses1
1	8170	1367	0.050	0.86	0.159 (0.146/. 172)	x	x	x	x	x	x	x	x	
2	5868	1367	0.050	0.86	0.144 (0.128/0.161)	x	x	x	x	x	x	x		
3	2339	1367	0.053	0.87	0.126 (0.092/0.161)	x							x	x
4a	Not identified													
4b	2802	1538	0.057	0.61	0.317 (0.256/0.378)									

95% CI = Confidence interval for variance of CMV factor.

samples. Table 2 presents the fit indices of the configural model for each sample. Besides global fit indices, the table also lists items which, based on Jrule indications, do not empirically fit the model. We, further, considered an item problematic if it had low loadings on its proposed factor, or if it correlated strongly with factors that are not in its immediate neighborhood within the circle (see Fig. 1). Table 2 also includes the variance of the CMV factor, which can be interpreted as the strength of the method effect.

Based on the RMSEA, the fit for samples 1–3 could be considered acceptable, but based on CFI and Jrule, this was not the case. Jrule identified at least 4 clearly misbehaving items within each sample, most of which appeared in at least two samples, and these deviations seemed to be related to wording issues with the questions.

The variance in CMV was higher (15.9%) for the Estonian online sample (sample 1) than for the Estonian paper-and-pencil sample (sample 2; 14.4%) and lower (12.6%) for the Finnish online sample (sample 3) than both Estonian samples, although all these differences were statistically non-significant.

In sample 4, from Ethiopia, the model (4a) did not converge, which indicated severe misfit. To measure the CMV variance in the Ethiopian sample, we reran a method factor model with only two factors, where the first latent factor was a value factor with one loading fixed on 1 for identification and the other items estimated freely, and the second latent factor was a CMV factor, with equally constrained loadings on all items (4b). It became evident that, for the Ethiopian sample, the method effect was more than twice as large as in the other samples.

At least under the given circumstances, the refined value structure did not work in the Ethiopian context and we excluded the Ethiopian sample from subsequent analyses. Table 3 compares fit indices after exclusion of the nine problematic items identified by Jrule.

Based on the RMSEA and CFI, all tested models were acceptable. Jrule indicated several more potential misspecifications, but, as they did not exist systematically across samples, they were treated as a sampling fluctuation. Therefore, it can be concluded, with the exception of a few problematic items, that the theoretical model was empirically confirmed in both Estonian samples and in the Finnish sample.

3.2. Data collection mode on comparability of relationships and latent means

Next, we tested for the mode effect by conducting invariance tests across the two Estonian samples. Problematic items apparent in both Estonian samples (fac3, hum1, sep3, ac1, st2, tr3, unc1, and he3; see Table 2) were excluded from the model. Table 4 displays the results of metric and scalar invariance tests.

All loadings and intercepts were invariant across the samples tested, as none of the equality constraints was detected by Jrule as misspecified. As shown in Table 2, the CMV also did not differ significantly across modes of data collection. Thus, there seems to be no systematic mode effect at play. This indicates that relationships and latent means are comparable across modes, and that Estonian data from online and paper-and-pencil modes can be combined.

3.3. Cross-cultural comparability of relationships and latent means

To perform the invariance test across Finnish and aggregated Estonian samples, a ninth item (ses1), which was problematic in previous analyses (see Table 2), was excluded. Table 5 displays the results for both metric and scalar invariance tests.

Based on fit indices, metric invariance holds, but according to Jrule indications, there were two non-invariant loadings. After allowing the loadings for pod1 and sdt3 to be freely estimated, there were no more misspecified parameters. This indicated that the PVQ-R is partially metric invariant across the two samples, allowing for valid comparisons of relationships between values and other constructs.

The scalar invariance model revealed five non-invariant intercepts, namely, for the items hum3, sdt3, por1, fac2, and bec1. After setting these items free, the global model fit indices suggested an at least acceptable fit, which means that partial scalar invariance holds and that the latent means are comparable across Estonian and Finnish samples.

3.4. Observed means vs. latent means and observed ranks vs. latent ranks

Fig. 3 gives an overview of the observed and latent means and ranks of each value type, which enables us to comprehend differences within and across countries and modes of data collection. While observed means are often reported in substantive research without prior tests for equivalence, the latent means presented correct for both random measurement error and bias due to CMV. Because of the latter, the latent means were always lower than the observed means.

When correlating observed and latent value ranks within each sample, we calculated Spearman's rhos of 0.90 for the Estonian samples and 0.92 for the Finnish sample. To evaluate change across countries, we correlated observed rank differences with latent rank differences and found a correlation of $r = 0.75$ across Estonian online and Finnish samples and $r = 0.80$ across Estonian paper-and-pencil and Finnish samples. Based on these correlations, we can conclude that the random measurement error and CMV have a small effect on value hierarchies within countries and a slightly larger effect across countries, while there is no significant effect across modes of data collection.

Table 3
Configural fit after exclusion of 9 problematic items.

Sample	Chi^2	DF	RMSEA	CFI
1	4829	908	0.047	0.90
2	3300	908	0.045	0.91
3	1430	908	0.048	0.91

Table 4
Global fit indices and Jrule indications for different levels of measurement invariance in data sets collected in paper-and-pencil and online modes (8 items excluded).

Model	Chi^2	DF	RMSEA	CFI	Jrule indications
EE1 vs EE2 metric (full)	8224	1845	0.046	0.90	No misspecifications
EE1 vs EE2 scalar (full)	8339	1873	0.046	0.90	No misspecifications

Table 5
Global fit indices for different levels of measurement invariance between Estonian and Finnish samples (9 items excluded).

Model	Chi^2	DF	RMSEA	CFI	Deviations
EE vs FI metric (partial)	8439	1843	0.045	0.90	–
EE vs FI metric (full)	8473	1845	0.045	0.90	pod = ~pod1; sdt = ~sdt3
EE vs FI scalar (partial)	8814	1867	0.046	0.90	–
EE vs FI scalar (full)	9123	1872	0.047	0.89	hum3 ~ 1; sdt3 ~ 1; por1 ~ 1; fac2 ~ 1; bec1 ~ 1

4. Discussion

The general aim of this paper was to test the measurement invariance of Schwartz's refined value structure across three countries and two modes of data collection. We argue that common method variance (CMV) is, at least partially, imposed by people's response styles (He & van de Vijver, 2015; Schwartz et al., 1997), as this method bias expresses itself as a positive correlation among all value items (Hinz et al., 2005). Value research usually corrects for this response style bias by centering answers within-persons (Fischer, 2004; Schwartz et al., 2012). For this purpose, the average rating a person gives to all value items (a variable often termed MRAT) is subtracted from the rating given for each single value item. After this transformation, the linkages between values and attitudes, behavior, and background variables become consistently more logical and theoretically meaningful than the associations among non-centered ratings (Borg & Bardi, 2016). Centered data, however, cannot be analyzed with SEM approaches that test for measurement equivalence (Fischer, 2004). Moreover, response biases have been found to vary across cultures (Smith, 2011; Strack & Dobewall, 2012; He et al., 2014), making it more difficult to separate substance and style in values measurement. To overcome this problem, we applied a bi-factor model (Holzinger & Swineford, 1939), which enabled us to control for CMV while testing for measurement equivalence of the entire value circle at once. The main strength of

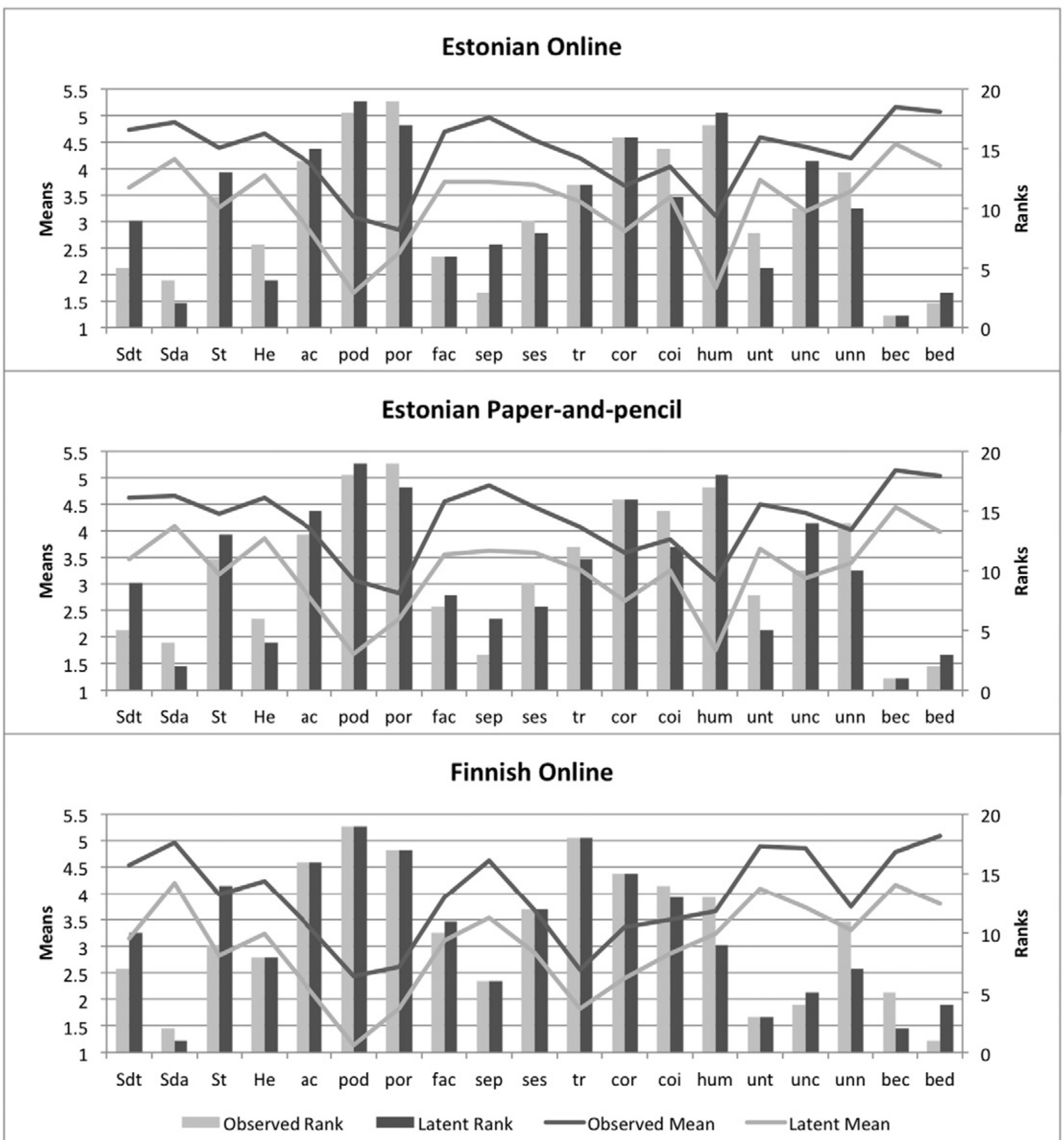

Fig. 3. Observed and latent means and ranks of value types across three samples.

the current study, therefore, is that a bi-factor model has been employed on the full Schwartz values structure, while previous work has tested the claim of universality only for a subset of Schwartz's value types. It should be noted that Verkasalo et al. (2009) modeled the response tendency for PVQ-21 using a principal component factor analysis, and, recently, Schwartz et al. (2012) also used a method factor, but neither study tested for measurement equivalence. The current work closes this gap.

Our first hypothesis was only partially supported, as the refined value structure held relatively well in the Estonian and Finnish, but not Ethiopian, contexts. This raises questions about the universality of the Ethiopian data. The internal consistency of all factors was very low for Ethiopian respondents,[2] which seems to indicate that the participants in this sample conceptualized value items systematically differently to the structure stated by Schwartz's refined theory of individual values.

Our second hypothesis was confirmed, as Ethiopian respondents seemed to be much more affected by method (CMV) than their Estonian and Finnish counterparts, which could be either a cultural phenomenon or related to fewer experiences in participating in surveys (He et al., 2014; Smith, 2011). This outcome also supports earlier findings of compatibility and conflict among values fit, with a more consistent theoretical value structure in samples drawn from socioeconomically more developed countries (Fontaine et al., 2008; Strack & Dobewall, 2012). Nevertheless, in a recent PVQ-R study by Tamir et al. (2015), the African sample from Ghana seemed to work well, and thus we cannot rule out that language issues (as the survey was answered in English) rather than culture accounted for the reported non-invariance in this sample.

Similar to the findings of Cieciuch et al. (2014), we identified 9 items that functioned poorly for Finnish and/or Estonian respondents and we therefore suggest that the wording of these questions be revised in subsequent PVQ-R versions. After the exclusion of these 9 items, our analyses supported partial scalar equivalence (with 2 non-invariant loadings and 5 non-invariant intercepts) across Estonian and Finnish samples. This shows that the PVQ-R works better than earlier Schwartz value instruments (e.g., SVS, PVQ-40, and PVQ-21), for which only a few studies have found support for higher levels of measurement invariance (e.g., Cieciuch & Davidov, 2012). In further support of hypothesis 1, we found full scalar equivalence across datasets for the paper-and-pencil and online modes, which is line with earlier findings (Cieciuch et al., 2015; Davidov & Depner, 2011) and suggests that values data from these modes can be safely combined.

Our third hypothesis was empirically supported, as we did not find significant differences in the variance attributable to CMV across online and paper-and-pencil modes. Neither were there significant differences in the observed or latent means of the 19 value types across these two modes.

Finally, the analyses indicate that CMV, along with other types of measurement error (see Lilleoja & Saris, 2015), has a significant impact on value structure and its comparability across countries. Comparison of observed and latent ranks revealed several changes in value hierarchies, which highlights the necessity of correcting for CMV in Schwartz's refined values structure. This is especially notable, as most previous tests for measurement equivalence in Schwartz's values have been conducted without separating substance from style. The comparison of the observed and latent means further revealed that random measurement error and CMV affected ranks within countries less than across countries. This is an important finding, as the order in which values are emphasized within a society defines the cultural norms prevalent in that society (Schwartz, 1994).

[2] Average reliability was 0.41, ranging from 0.30 (Humility) to 0.55 (tradition).

4.1. Conclusions

According to our study, the PVQ-R scale seems to be quite a suitable instrument for assessing human values across countries as well as modes of data collection. After accounting for method bias (CMV), the refined value structure holds relatively well in the Estonian and Finnish, but not Ethiopian, contexts. Data collected in paper-and-pencil and online modes are combinable, as full scalar equivalence holds. Across Estonian and Finnish samples, partial metric and scalar equivalence was found, which means that relationships and latent means can be compared, with some minor limitations.

References

Biderman, M. D. (2014). Maximizing validity of personality questionnaires. *Paper presented at the 10th annual river cities industrial-organizational psychology conference*. Chattanooga: TN.

Biderman, M. D., Nguyen, N. T., Cunningham, C. J., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the big five. *Journal of Research in Personality*, *45*, 417–429.

Borg, I., & Bardi, A. (2016). Should ratings of the importance of personal values be centered? *Journal of Research in Personality*, *63*, 95–101.

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

Cieciuch, J., & Davidov, E. (2012). A comparison of the invariance properties of the PVQ-40 and the PVQ-21 to measure human values across German and polish samples. *Survey research methods. 6.* (pp. 37–48).

Cieciuch, J., Davidov, E., Vecchione, M., Beierlein, C., & Schwartz, S. H. (2014). The cross-national invariance properties of a new scale to measure 19 basic human values a test across eight countries. *Journal of Cross-Cultural Psychology*, *45*, 764–776.

Cieciuch, J., Davidov, E., Oberski, D. L., & Algesheimer, R. (2015). Testing for measurement invariance by detecting local misspecification and an illustration across online and paper-and-pencil samples. *European political science. 14(4).* (pp. 521–538).

Davidov, E., & Depner, F. (2011). Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Quality and Quantity*, *45*, 375–390.

Fischer, R. (2004). Standardization to account for cross-cultural response bias a classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology*, *35*(3), 263–282.

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

He, J., & van de Vijver, F. J. R. (2015). Self-presentation styles in self-reports: Linking the general factors of response styles, personality traits, and values in a longitudinal study. *Personality and Individual Differences*, *81*, 129–134.

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

Holzinger, K., & Swineford, F. (1939). *A study in factor analysis: The stability of a bifactor solution. Supplementary educational monograph, no.48.* Chicago: University of Chicago Press.

Knoppen, D., & Saris, W. E. (2009). Do we have to combine values in the Schwartz' human values scale? A comment on the Davidov studies. *Survey research methods. 3.* (pp. 91–103).

Lilleoja, L., & Saris, W. E. (2014). Testing a new operationalization of the basic values on Estonian- and Russian-speaking subpopulations in Estonia. *Social Indicators Research*, *116*, 153–172.

Lilleoja, L., & Saris, W. E. (2015). Does correction for measurement error have an effect on the structure of basic human values? *Survey research methods. 9.* (pp. 169–187).

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525–543.

Motl, R. W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*, *9*, 562–578.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Pornprasertmanit, S. (2014). Modification indices and their power approach for model fit evaluation. (Retrieved from) http://rpackages.ianhowson.com/cran/semTools/man/miPowerFit.html

Revilla, M. (2013). Measurement invariance and quality of composite scores in a face-to-face and a web survey. *Survey research methods. 7.* (pp. 17–28).

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research* (2nd ed.). Hoboken, NJ: Wiley.

Saris, W. E., Knoppen, D., & Schwartz, S. H. (2013). Operationalizing the theory of human values: Balancing homogeneity of reflective items and theoretical coverage. *Survey research methods. 7.* (pp. 29–44).

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

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*, 19–45.

Schwartz, S. H., Verkasalo, M., Antonovsky, A., & Sagiv, L. (1997). Social desirability and value priorities: Much substance, some style. *British Journal of Social Psychology*, *36*, 3–18.

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., ... Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, *103*, 663–688.

Smith, P. B. (2011). Communication styles as dimensions of national culture. *Journal of Cross-Cultural Psychology*, *42*, 216–233.

Steinmetz, H., Isidor, R., & Baeuerle, N. (2012). Testing the circular structure of human values: A meta-analytical structural equation modelling approach. *Survey research methods. 6.* (pp. 61–75).

Strack, M., & Dobewall, H. (2012). The value structure in socioeconomically less developed European countries still remains an ellipse. *Europe's Journal of Psychology*, *8*, 587–602.

Tamir, M., Schwartz, S. H., Cieciuch, J., Riediger, M., ... Vishkin, A. (2015). Desired emotions across cultures: A value-based account. *Journal of Personality and Social Psychology*. http://dx.doi.org/10.1037/pspp0000072 (Advance online publication).

Van der Veld, W., Saris, W. E., & Satorra, A. (2009). *Jrule 2.0, judgement aid rule for structural equation models: User manual*.

Verkasalo, M., Lönnqvist, J. E., Lipsanen, J., & Helkama, K. (2009). European norms and equations for a two dimensional presentation of values as measured with Schwartz's 21-item portrait values questionnaire. *European Journal of Social Psychology*, *39*, 780–792.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology*, *34*(6), 702–722.