RECSM
Research and Expertise Centre for Survey Methodology

# Does Correction for Measurement Error Have an Effect on the Structure and Comparability of Basic Human Values?

**Laur Lilleoja and Willem E. Saris**

RECSM Working Paper Number 45
October 2015

http://www.upf.edu/survey/_pdf/RECSM_wp045.pdf

Does Correction for Measurement Error Have an Effect on the Structure and Comparability of Basic Human Values?

Laur Lilleoja[1]; Willem E. Saris[2]

Abstract

It is a well-known fact that survey data always contains measurement errors, which can in different ways bias the results of an analysis (Alwin, 2007). In quantitative research, this problem has usually been ignored due to lack of quality estimates and the complexity of correction procedures. However, this argument is not valid anymore, as there exists a new approach to determining the quality of any specific question based on a prediction of quality using the program *SQP2 (Survey Quality Predictor 2.1)*, which in turn allows for the correction of measurement error by following a relatively simple procedure as described by Saris and Gallhofer (2014) and illustrated by DeCastellarnau and Saris (2014).
During recent years, there has been an increasing number of studies researching the structure of basic human values (Schwartz, 1992), which has resulted in improvements to the theory and led to a new 19-factor value model (Schwartz et al., 2012; Saris, Knoppen, & Schwartz, 2013). However, none of the recent studies have concurrently taken into account random and systematic errors, which can potentially have an effect on the structure. Therefore in this paper we will try to overcome this shortcoming and will be analysing the effect of the correction for both types of measurement error on the structure of basic human values.
We use the same representative data from Estonia that was used in a study by Lilleoja and Saris (2014), which enabled us to compare value structures before and after the correction for measurement error. Due to the existence of a large ethnic minority in Estonian society, the sample additionally allowed testing the equivalence of the measurements in the two subpopulations after correcting for measurement error.
This study shows that the correction for measurement error provides additional support for the validity of a new value structure and it also sheds some new light on cross-cultural equivalence.

Keywords: PVQ-40; Correction for measurement error; SQP2; Metric invariance;

## 1 Introduction

The popularity of empirical values research is continually increasing, and is setting high expectations for the methodology of value research. Currently, the most dominant theoretical approach to value studies is the theory of basic human values by Shalom Schwartz (1992). This theory is attractive, as it has succeeded in describing the structure that links together the universe of basic values.

Still, due to its ambitious scope, it has received various criticisms, leading to several improvements. The initial 10-factor model has been expanded to a new 19-factor model and the value questionnaires have also undergone several revisions.

But there has been much less research on the measurement part; for example the effect of measurement error on value structures has gotten very little attention so far. The disregard for measurement error is a more general problem in quantitative research and it definitely needs more attention, as it can strongly bias the conclusions. Alwin (2007) has suggested that even as much as 50% of the variance of the observed variables in survey research is error. So if it turns out that the relationships between individual values are biased by the measurement error,

[1] School of Governance, Law and Society, Tallinn University; email: laur@tlu.ee
[2] The Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra, Barcelona
[3] Besides JRule for LISREL (Van der Veld, Saris, & Satorra, 2009), there [illegible]
[2] The Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra, Barcelona

the overall value structure might also be affected. The same risk applies for the relations between values and other variables.

Many earlier studies (e.g. Davidov & Schmidt, 2007; Davidov, Schmidt, & Schwartz, 2008; Davidov, 2008; Knoppen & Saris 2009a; Knoppen & Saris, 2009b; Saris, Knoppen, & Schwartz, 2013) have in fact used multiple indicators and structural equation modelling (SEM) while analysing the value structure, which has made it possible to control for random errors. However, this does not control for systematic errors (including method effects), which have potentially a strong effect on results as suggested by Campbell and Fiske (1959). There have also been SEM-based analyses (Beierlein et al., 2012; Schwartz et al., 2012) that include separate method factors to control for response style (acquiescence), but it is not clear how large portion of systematic errors these approaches remove; definitely not all of them however.

This research will try to fill in the gaps in our knowledge about the effects of measurement error on the structure of basic human values and on the relationship between values and other variables, by using an innovative approach of program-based predictions for the qualities of each value item, which allows for the correction of all method effects. The structure of basic values will be re-analysed after correcting for measurement error and the results will be compared with evidence found in an earlier study (Lilleoja & Saris, 2014). A representative national data set from Estonia is used, which also allows testing for the equivalence of the measurement across two ethnolinguistic groups.

### 2.1 The structure of basic human values

Schwartz (1992) has defined values as desirable, trans-situational goals, varying in importance, that serve as guiding principles in people's lives. According to his Value Theory, every individual value in any culture is locatable under 10 universal, motivationally-distinct basic values - *hedonism, stimulation, self-direction, security, universalism, benevolence, conformity, tradition, power and achievement*.

There also exists a universal structure of dynamic relations among those basic values: pursuing one type of value will always create a conflict with other types of values. Based on these kinds of relations, Schwartz has outlined the circular structure of basic values, where more similar value types (like *hedonism* and *stimulation*) are close to each other and conflicting value types (like *benevolence* and *power*) appear on opposite sides. Based on this opposition, value types also form two bi-polar contrasting higher-order dimensions: *self-enhancement* vs. *self-transcendence* and *openness to change* vs. *conservation* value types (Schwartz, 1992; Schwartz, 1994).

Initially for the data collection a 57-item questionnaire called the *Schwartz Value Survey* (SVS) with abstract value labels was used (Schwartz 1992), which was later replaced by the 40-item *Portrait Value Questionnaire* (PVQ), which includes 40 short verbal portraits of different people (Schwartz 2007).

Schwartz originally used the *Smallest Space Analysis* (SSA, a type of multidimensional scaling [MDS]) to map the circular continuum of 10 value types in a two-dimensional space and, according to him, the validation of this theoretical structure has been confirmed in more than 67 different cultures (Schwartz, 1992, 2005).

Still, there has been a lot of critical discussion questioning the legitimacy of this structure (e.g. Mohler & Wohn, 2005; De Clercq & Fontaine; 2006; Perrinjaquet et al., 2007; Davidov & Schmidt, 2007; Davidov, Schmidt, & Schwartz, 2008; Davidov, 2008; Knoppen & Saris 2009a; Fischer et al., 2010; Knoppen & Saris, 2009b). While Schwartz argued that it is possible that some samples might not fit 100% into the structure (which, in general, meant that some of the basic values had to be unified), cited papers referred to more systematic deviations.

SSA can be very useful in the first stage of this kind of analysis, but it also has strong disadvantages, such as the absence of a clear criterion for determining exact boundaries between value types in a two-dimensional MDS space (Knoppen & Saris, 2009a). In order to overcome this limitation, the Confirmatory Factor Analysis (CFA) has been used.

The CFA approach has revealed several substantive deviations from the theory-based structure (Davidov & Schmidt, 2007; Davidov et al., 2008; Davidov, 2008). It has been suggested that certain values may consist of weakly-related sub-values (Schwartz & Boehnke, 2004), while between others there might not be enough discriminant validity (Perrinjaquet et al., 2007; Schmidt et al., 2007), leading to only 7 independent basic values (Davidov & Schmidt, 2007; Davidov et al., 2008; Davidov, 2008).

Unfortunately, the CFA also has its weaknesses, the central one being the determination of model fit. As all widely-used criteria (like the *chi* 2 test, the RMSEA and the CFI) ignore the power of the test, they can only detect misspecifications for which the test is sensitive (Saris et. al., 2009). That in turn can lead to model rejection due to very small misspecifications to which the test is very sensitive, and acceptance of the wrong model due to the lack of power of the test.

Schwartz however has criticised the CFA approach because it contradicts the view of values as arrayed on a continuum, as it seeks to confirm relatively pure factors, and each item ideally loads on only one factor (Schwartz, 2011). The latter statement is nonetheless not true because cross-loadings are in principle allowed in the CFA but, in that case, they have to be specified in the model. If they are ignored this represents a misspecification which leads to improper estimates (like correlations larger than 1.0).

In response to previous studies, Knoppen and Saris (2009b) proposed an alternative approach to overcome the usual CFA shortcomings by analysing the value structure in smaller parts and using the program *Judgment Aid Rule* (JRule[3]), which provides different model fit evaluation criteria (Van der Veld, Saris, & Satorra, 2009). While applying this approach on the PVQ-40 questionnaire Saris and his colleges  (Knoppen & Saris, 2009c; Saris et al., 2013) arrived at a new 19-value type structure, whereby several original values were split into independent sub-values. Several other studies have also provided similar findings (e.g. Beierlein et al., 2012; Cieciuch & Schwartz, 2012).

The main limitation of the study by Saris et al. (2013) was that it was based on student samples. To confirm the wider validity of the proposed structure, it was retested on a representative sample of the Estonian population (Lilleoja & Saris, 2014). That paper showed that the alternative value-structure, described by Saris and colleges (2013), had a relatively good fit with the representative Estonian data, and while two pairs of sub-values had to be reunified, the final model included 17 value types[4]: *humility , societal security, conformity, preserving nature, health, maintain traditions, personal security, equality, benevolence, tolerance, autonomy of action, autonomy of thought, power, wealth, hedonism, achievement* and *stimulation* (Lilleoja & Saris, 2014). In this paper, we will use the same data set and retest the value structure after correcting for measurement error to find out if there exist any differences in the structure[5].

## 2.2 Measurement error and correction for the measurement error

It is well known that measurement error can considerably attenuate the relationships between variables (Andrews, 1984; Alwin & Krosnick, 1991; Költringer, 1995; Scherpenzeel, 1995; Saris & Gallhofer, 2014), while Alwin (2007) has even stated that up to 50% of the variance of the observed variables in survey research is error variance. Therefore the variables that we

---

[3] Besides JRule for LISREL (Van der Veld, Saris, & Satorra, 2009), there exists a Jrule for Mplus (Oberski, 2010) and a corresponding function in R (Pornprasertmanit, 2014).

[4] The list of PVQ items and factor structures can be found in Appendix 1.

[5] Based on new evidence, Schwartz et al. (2012) have recently developed a new 57-item value scale, PVQ-R, which corresponds better to the larger number of value types and has also enabled the mapping of a new refined value structure.

are measuring can (due to random errors and peculiarities in question formulations) considerably differ from the variables that we would like to measure (Saris & Revilla, 2015), which in turn can significantly bias our conclusions.

Due to the wide usage of scales with multiple indicators, it has become relatively common to control for random error by using the CFA. However, only a small number of researchers apply complete correction for measurement error (taking into account both random and systematic errors) in their research and therefore it is not surprising that the given procedure can bring new insights even to intensively studied topics, like the relationship between income and satisfaction (Saris 2001), and so it can potentially have effects in the context of value research as well. On the one hand, it can hinder finding empirical confirmation for the theoretical value structure and, on the other hand it can make data incomparable across countries.

The main arguments for not (completely) correcting for the measurement error have been the lack of quality estimates for the indicators to be used in the analysis, and the complexity of the correction procedure. It is assumed that quality determination needs more complex research designs, like the test-retest design (Lord, Novick, & Birnbaum, 1968), Quasi-simplex models (Heise, 1969; Wiley & Wiley, 1970) or the Multitrait-multimethod (MTMM) design (Campbell & Fiske, 1959), which means a large quantity of extra questions. However the given arguments are not valid anymore, as there exists a new approach to determining the quality of specific questions, based on a prediction of the quality using the program Survey Quality Predictor 2.1 (SQP2)[6] (Oberski et al., 2011), which makes it rather easy to correct for the measurement error.

SQP2 is a free open-source tool, which allows researchers to obtain a quality prediction based on a meta-analysis of MTMM experiments done on more than 3,700 questions (Saris & Gallhofer, 2014). To get predictions, the researcher has only to code the characteristics of the questions. The program can also be used to design better survey items, as it can also make suggestions for improvements of the questions if their quality is not good enough (Survey Quality Predictor, 2015). It is important to keep in mind, however, that SQP2 provides predictions, not exact measures[7]. In addition, the quality estimates enable correcting for measurement error, not for possible errors produced by nonresponse, coverage problems, processing deviations, etc.

To correct for measurement error in the measurement part of the value structure, we use the measurement model described on Figure 1.

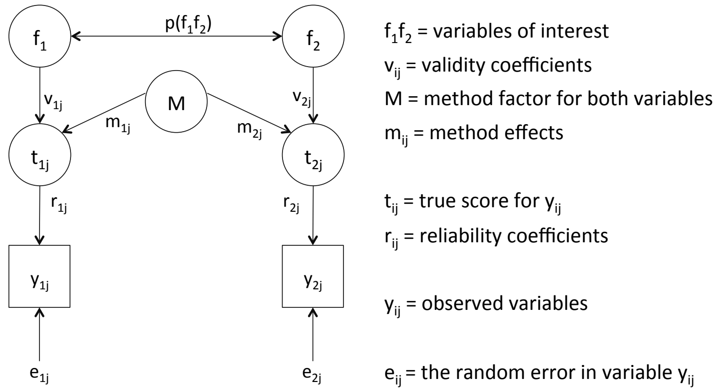


*Figure 1. The measurement model for two traits measured with the same method. Taken over with permission from Saris & Gallhofer, 2014.*

6 More information about this program can be found on *sqp.upf.edu*.

7 SQP2 quality predictions are based on the Random Forest approach (Breiman 2011); the current overall explained variance ($R^2$) for SQP2 reliability predictions is 0.65 and for validity predictions is 0.84 (Oberski et al 2011).

Reliability is the degree to which an assessment tool produces stable and consistent results, while validity refers to how well an indicator measures what it is purposed to measure (Saris & Gallhofer, 2014). Based on the given model (Figure 1), the correlation between observed variables $\rho(y_{1j},y_{2j})$ is equal to the joint effect of the variables that we want to measure ($f_1$ and $f_2$) plus the spurious correlation due to the method factor, which can be described as follows (Formula 1):

$$\rho(y_{1j}, y_{2j}) = r_{1j}v_{1j}\rho(f_1, f_2)v_{2j}r_{2j} + r_{1j}m_{1j}m_{2j}r_{2j} \quad (1)$$

From this it follows that the real correlation between variables of interest is equal to the observed correlation minus the spurious relationships due to the common method, divided by the quality of both variables, as described by the next formula (2):

$$\rho(f_1, f_2) = \frac{\rho(y_{1j}y_{2j}) - (r_{1j}m_{1j}m_{2j}r_{2j})}{(r_{1j}v_{1j}v_{2j}r_{2j})} \quad (2)$$

This formula can also be expressed as a linear function (3), where the correlation between the variables of interest is equal to 1 divided with the product of quality coefficients of both variables times the observed correlation minus the product of the method effects of both variables divided by their validity coefficients:

$$\rho(f_1, f_2) = \frac{1}{r_{1j}v_{1j}v_{2j}r_{2j}}\rho(y_{1j}y_{2j}) - \frac{m_{1j}m_{2j}}{v_{1j}v_{2j}} \quad (3)$$

To obtain validity and reliability coefficients and method effects, we use SQP2. Because the correlations between the observed variables are known and the reliability, validity and method effects can be obtained from SQP2, the correlations can be corrected for measurement error using Formula 2 and then the value model can be easily estimated on the basis of correlations[8] corrected for measurement error using SEM programs like STATA, LISREL or R (DeCastellarnau & Saris, 2014; Saris & Revilla, 2015). The difference with the standard procedure is that this procedure also corrects for the method effects or as it is also called "common method variance" (cmv).

## 2.3 Invariance testing

In addition to structural changes, we also took an interest in testing for measurement invariance across two linguistic subsamples (Estonian-speakers and Russian-speakers) after correction for measurement error.

A typical approach for invariance testing (Meredith, 1993) seeks functional equivalence without distinguishing between the cognitive and the measurement processes. This means that normally it is assumed that both the way in which the respondents understand the question (how they cognitively react to the stimuli), and the way in which they respond to the question, have to be the same across groups. Saris and Gallhofer (2014) have shown that for meaningful comparisons it is enough to have only a cognitive equivalence (the interpretation of questions has to be same), as the measurement part can be corrected, which in turn can lead to different results compared with the standard test.

[8] The described procedure can also be used to correct the covariance matrix for measurement error. For simplicity we use correlations throughout our analyses, although we did not find any differences in the results when using corrected covariance matrixes instead. For correction for measurement errors and analysis of covariance matrices we refer to DeCastellarnau & Saris (2014).

Therefore the latter might be a reason why several value studies haven't found (full) invariance in cross-cultural comparisons (Davidov et al., 2008; Davidov, 2008; Vecchione et al., 2015). While the opposite scenario is also possible – it might be that in other studies (Davidov & Schmidt, 2007; Davidov, 2010; Davidov & Depner, 2011; Cieciuch & Davidov, 2012) the invariance found is hidden by the measurement error and in reality cognitive equivalence does not exist.

When cognitive and measurement processes are separated, equivalence should only be tested for the cognitive part, as the response part can be corrected using SQP2 predictions (Saris & Gallhofer, 2014). Figure 2 describes the difference between these two approaches, where the left-side model stands for the standard metric invariance test and the right-side model for the new approach. The standard approach includes observed responses (y), error terms (ζ) and loadings (λ). The new approach includes two levels where observed responses (y) and reactions to the questions (f) are separated. As coefficients (q) for the measurement part can be obtained with SQP2, only the upper part needs to be estimated and the equivalence can be tested by fixing the loadings (c) of the cognitive part to be same across groups. After correcting the measurement part, instead of errors (e) also the unique components (u) are estimated.

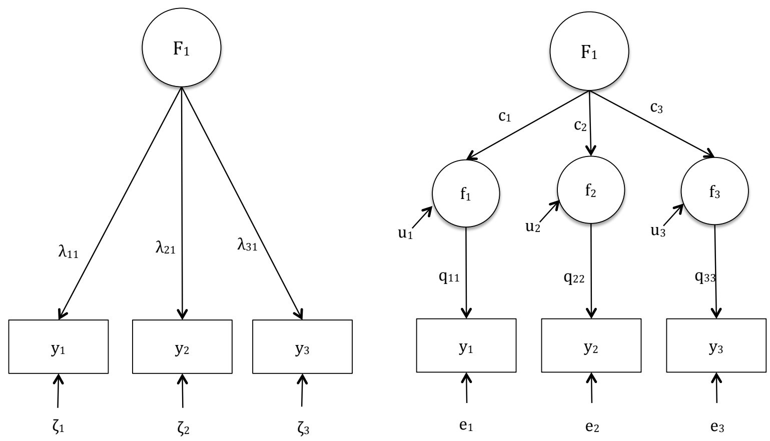


*Figure 2. Standard test for metric invariance versus a new approach where the cognitive (upper part) and the response processes (lower part) are distinguished.*

**3 Methods**

3.1 Data

To study the effect of measurement error on value structure, we use the same representative data set from the Estonian population, which was used in the previously-mentioned paper by Lilleoja and Saris (2014). The given sample allows comparison of changes in the value model before and after the correction for measurement error and it also allows testing for cognitive level equivalence between two different cultural groups – ethnic Estonians and ethnic Russians living in Estonia.

The dataset used in the analysis consists of a probability sample of the Estonian population with 1,240 respondents, of whom 776 (63%) were Estonian-speakers and 464 (37%) Russian-speakers. The survey response rate was over 50%. The sample structure matches rather well with population distributions with respect to background variables (Lilleoja & Saris, 2014). The data was collected in the autumn of 2008 using translated versions of the original PVQ 40 questionnaire (Sõmer, 2011).

### 3.2 Correction for measurement error

The very first step of the study is obtaining validities, reliabilities and method effects for each indicator, which are needed to correct for measurement error. Every question was therefore separately coded using SQP2. As the scales were the same and the questions were rather similar, the quality coefficients were also expected to be similar – for the Estonian-speaking population, the average reliability was .8 (min .79; max .82), the validity .93 (min .92; max .94) and the method effect .37 (min .35; max .39). For Russian-speakers the average reliability was .81 (min .80; max .84), the validity .94 (min .93; max .94) and the method effect .35 (min .33; max .38). Detailed information about the reliability and validity coefficients and method effects can be found in Appendix 2.

The next step is the correction of the initial correlation matrix – the same one that was used in the earlier study by Lilleoja and Saris (2014). The procedure for correction of measurement error was conducted separately for Estonian-speakers and Russian-speakers, using the measurement model described in Figure 1 and Equation 2. The covariance matrices of the value indicators before and after correction for measurement error are available at http://www.upf.edu/survey/working/extrafiles.html.

As all the predicted qualities were very similar for all the indicators, all the correlations between the indicators changed systematically after the correction for the measurement error. The relationship between correlations before and after the correction can be derived directly from formula 3 by using average coefficients. Equation 4 describes the function for Estonian-speakers and equation 5 for Russian-speakers, where *r* stands for observed correlations and *ρ* for correlations corrected for measurement error:

$$\rho = \frac{1}{.80{*}.93{*}.93{*}.80} r - \frac{.37{*}.37}{.93{*}.93} = 1.807r - 0.1583 \quad (4)$$

$$\rho = \frac{1}{.81{*}.94{*}.94{*}.81} r - \frac{.35{*}.35}{.94{*}.94} = 1.7249r - 0.1386 \quad (5)$$

We see that in this case the correction for lack of quality increases the positive correlations and decreases negative correlations while the cmv always decreases the correlations. The fit of these equations to the data was in both cases very good ($R^2$=.99) thanks to the similarity of the predictions of the quality coefficients for the different questions.

Given these relationships, the correlations that were stronger than 0.2 turned out to be underestimated, and they became more positive, while correlations less than 0.2 were overestimated and they became more negative. The correlations close to 0 changed less because the subtracted cmv levelled the positive effect of quality measures. The original correlations at both extremes (especially the negative ones due to the subtraction of the method effect) transformed substantially. For example, within the Estonian-speaking sample the correlation between achievement items *ac4* and *ac13* increased from .57 to .86, between *sd1* and *un3* it did not change significantly (.18 and .15) and between item *ac24* and *tr38* it changed from -.28 to -.68.

### 3.3 Estimation of factor structure

Because the analysis of a complete model with 19 factors and approximately 3 indicators for each factor leads in general to problems, like it was also the case in the previous analysis[9], we

[9] In the earlier study it turned out that several items had to be omitted in order to get a reasonable solution. We expected this also in this case. By analyzing sub-models of the big model one can better see which items create the problems and if these are the same as in the earlier analysis.

have decided to follow the procedure suggested by Knoppen and Saris (2009a, 2009b), who recommend before continuing with larger models, to first analyse the structure of smaller parts of the model that include only two theoretically adjacent value types. We went even further, by testing models for all possible combinations between 17 value types. This technique allows, on the one hand, more precise detection of potential deviations from the theoretical structure and, on the other hand, it helps to keep control over the model so that it does not get too complex. In this way, we also solved the problem of input matrix being not positive definite and avoided any non-convergence issues. In the next stage, each of 4 higher-level value dimensions was tested, taking into account the modifications done within small models.

The potential shortcoming of this approach is that we can end up with a model that has a slightly different structure compared to the reference model described by Lilleoja and Saris (2014), which in turn does not allow an immediate comparison of model fits before and after the correction for measurement error. Although as Saris et al. (2009) have shown, the standard fit tests that are based on the *chi*-square can, in the case of a high test power, lead to false model rejections, and as it is expected that after correction for measurement error the loadings will get stronger, which in turn will increase the *chi*-square and amplify the sensitivity of standard tests, the comparisons of the model fits will be questionable anyway.

Therefore, for evaluating the fit of the models we will additionally use the JRule program (Van der Veld et al., 2009), which takes into account the power of the test while detecting potential misspecifications. For every model we will first detect and correct the misspecifications (which make theoretical sense, and have standardised expected parameter change (epc) over .1) that are apparent in both samples and then continue with misspecifications that are apparent only in one sample.

For the analyses, we used the program Lavaan in R package (version 0.4-14, Rosseel, 2012). As we presume that the relationships between concepts are comparable across Estonian- and Russian-speakers, we restricted loadings to be the same across groups in all models, which enabled us to simultaneously test for metric invariance.

## 4 Results

### 4.1 Factor structure

Next we will present the empirical results concerning factor structure after correction for measurement error. Figure 3 gives a visual example of the procedure we used based on the Estonian-speakers' sample, using two theoretically adjacent values – *benevolence* and *equality* (the abbreviations for value items can be found in Appendix 1).

Model 1 describes the structure before correction for measurement error, with *benevolence* having four indicators and *equality* three indicators. According to the standard fit measures (CFI=.99; RMSEA=.033 (Bentler, 1990; Steiger, 1990)) the model is acceptable, although based on JRule, there appear to exist 4 possible misspecifications.

Earlier analyses (Saris et al. 2013; Lilleoja & Saris, 2014) have shown that one of the *equality* items (*un23*) should be excluded, as it does not fit empirically into the model. Indeed, after exclusion of *un23*, as described in Model 2, JRule does not highlight any other potential misspecifications.

Model 3 describes the structure of the *benevolence* and *equality* factors after correction for measurement error. As theoretically there are no more errors, the error terms are now treated as unique components. As we were also interested in testing whether the exclusion of item *un23* in the previous analyses could have been caused by measurement error, we included it as well. As expected, after correction for measurement error, the loadings systematically got higher, which due to the increased power of the test might be one reason for the worse model fit (CFI=.93; RMSEA=.126). It might be, however, that there exists a structural misfit as well, as JRule identifies 12 possible misspecifications. Based on JRule suggestions, it seems to be necessary to introduce an extra correlation between items *un3* and *un23*, which makes sense, as the wording of these items is very similar (Appendix 1), while *un29* is quite different.

Model 4 describes the final structure including extra correlation between *un3* and *un23*. Now the model fit has improved again (CFI=.98, RMSEA=.091) and more importantly the JRule identifies only 3 misspecifications[10]. We can see that in this case the correction for measurement error and differences in the model modifications have not had too much effect on correlations between *benevolence* and *equality* (.74 vs .67).

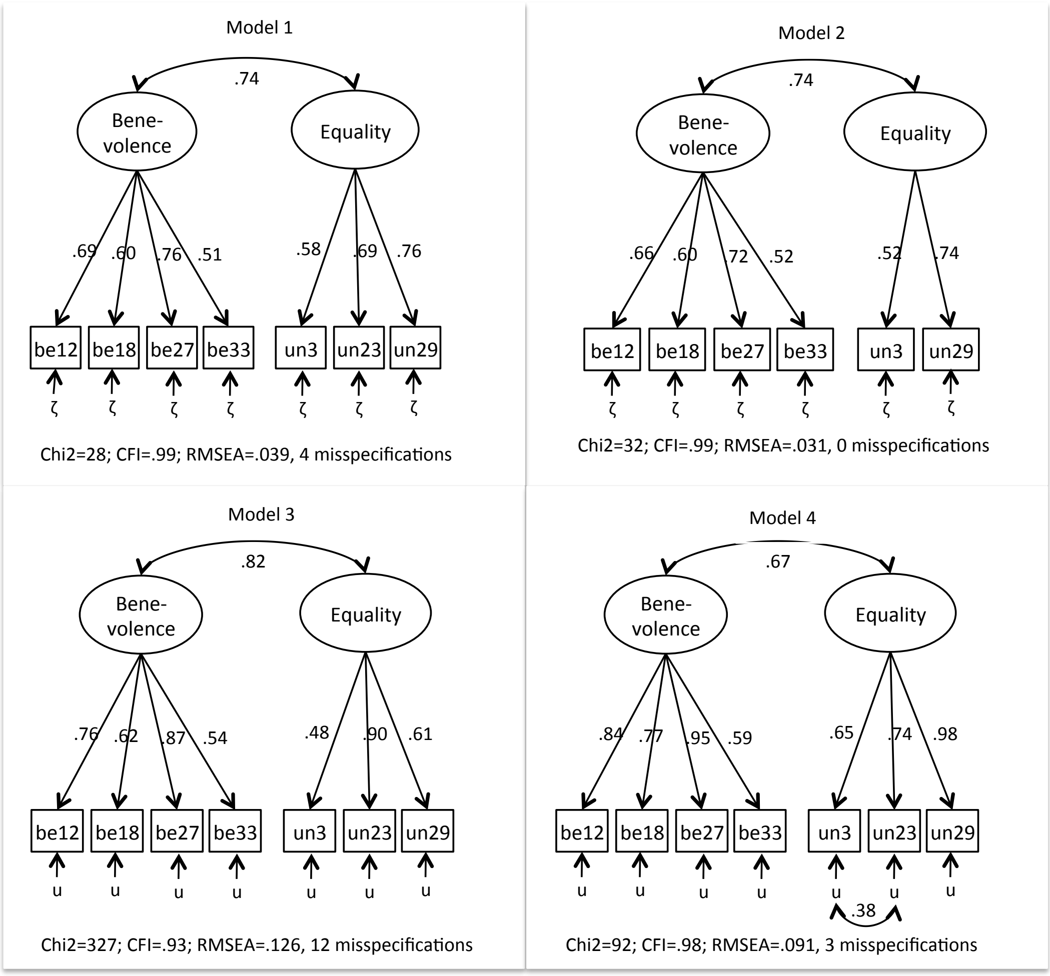


*Figure 3. Factor structure and parameter estimates for values for benevolence and equality before and after the correction for measurement error.*

After applying the same approach on all small models, we arrived at some more changes in the selection of value indicators. In addition to *be27*, items *co36, st15, po39, ac24* and *he10* also had to be excluded from the analyses. But at the same time, besides *un23*, two more indicators – *st6* and t*r20*, which were excluded in previous analyses, were again usable.

[10] At this point of the analysis, we did not see other theoretically justifiable modifications, which we could introduce. Although the later analyses suggested that the item *be27* should be removed, as it tend to correlate strongly with items of other factors (like *he33* and *st30*), which in turn caused convergence problems. The exclusion of *be27* resolved also all the remaining misspecifications of model 4.

As *co36, be27, ac24* and *he10* belonged to factors that had at least 3 indicators, their exclusion will probably have little effect on the relationship between the factors. However we assumed that there will be larger changes in the relationships between the *power*, *stimulation* and *tradition* factors because for *power* now only one indicator was left; for *stimulation, st15* was substituted with *st6*; and for *tradition*, which before had one indicator, there were now three indicators.

Next we tested each higher-order value dimension separately and then combined them with each other while keeping the modifications (correlations between unique components and cross-loadings) from earlier stages. By doing so, we estimated all the potential relations between the factors. The overview of the fit measures for the final higher-order value models can be found in Appendix 3. After correction for measurement error, all the models included more misspecification than before the correction procedure, which is probably the result of increased correlations. Although when following JRule decisions for introducing additional modifications, ultimately all the models were substantively acceptable[11] and the final list of value types matched with the structure described by Lilleoja and Saris (2014)[12].

## 4.2 Invariance

As mentioned earlier, due to the existence of two ethnolinguistic sub-samples, we can also test for cross-cultural equivalence. And as we have corrected for the differences in the response or measurement part, our invariance test will strictly compare the cognitive processes, to find out if the interpretation of the questions is the same across Estonian-speakers and Russian-speakers.

Based on all tested models, there were 7 non-invariant loadings between Estonian-speaking and Russian-speaking samples: *he37; st30; ac32; tr20; co28; un29;* and *un40*. This is considerably different from the results of the metric invariance test before correction for measurement error, where no non-invariant loadings were identified (Lilleoja & Saris 2014)[13]. These results indicate that after correcting the measurement part, the cognitive differences become much more evident and it becomes clear that Estonian-speakers and Russian-speakers understand and interpret several questions differently.

When analysing these items separately, we saw that some of the differences can be related to translations – for example the Estonian translation for item *he37* emphasises the importance of enjoying every moment of life, while the Russian translation talks about the importance of having a good life in general. The same holds for item *st30* – in the Estonian language the description is longer, as it involves an extra part (*he looks for new activities*), which is neither included in the original English version (Appendix 1), nor in the Russian translation.

It is also possible that some non-invariances appear due to cultural differences. For example, the item *tr20* is about the importance of religion, which definitely has different meanings across cultural groups, as less than 1/3 of Estonians identify themselves as religious or inclining to belief, while for Russian-speakers the proportion is more than 2/3 and the vast majority identify themselves as Russian Orthodox (Ringvee, 2014). Also the item *co28*, which is about the importance of respecting parents and the elderly, has supposedly different implications across these cultural groups, as in the Slavic culture respect towards the elderly is more central than among Estonians.

---

[11] The number of shared misspecifications indicated by JRule was in all cases less then 5% of all potential cross-loadings and correlations between unique components.

[12] The overview of item affiliations before and after correction for measurement error can be found in Appendix 1.

[13] Both tests were based on the analyses of correlation matrixes.

## 4.3 Correlations between the values

As we found comparable value structures before and after correction for measurement error, we could now look for changes in the correlations between the latent values. We start with the comparison of inter-value relationships before and after the correction for measurement error.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AutoTh | 0.37 | 0.82 | 0.64 | 0.69 | 0.42 | 0.73 | 0.32 | 0.11 | 0.15 | 0.26 | 0.24 | -0.16 | 0.39 | 0.33 | 0.27 | 0.23 |
| 0.31 | AutoAct | 0.51 | 0.41 | 0.34 | 0.19 | 0.43 | 0.27 | 0.14 | 0.2 | 0.03 | 0.14 | -0.08 | 0.26 | 0.26 | 0.3 | 0.24 |
| 0.8 | 0.36 | Stimul | 0.91 | 0.74 | 0.44 | 0.74 | 0.03 | 0.08 | -0.02 | 0.03 | -0.03 | -0.22 | 0.3 | 0.16 | 0.11 | 0.27 |
| 0.49 | 0.44 | 0.72 | Hedonism | 0.66 | 0.46 | 0.62 | 0.09 | 0.12 | 0.13 | 0.11 | 0.12 | -0.07 | 0.32 | 0.18 | 0.13 | 0.21 |
| 0.44 | 0.14 | 0.78 | 0.47 | Achiev | 0.76 | 0.94 | 0.17 | 0.15 | 0.11 | 0.1 | 0.06 | -0.39 | 0.13 | 0.06 | 0.01 | 0.04 |
| 0.23 | 0.04 | 0.52 | 0.32 | 0.76 | Wealth | 0.65 | 0.07 | 0.14 | 0.07 | -0.07 | -0.07 | -0.33 | -0.2 | -0.1 | -0.17 | -0.24 |
| 0.34 | -0.05 | 0.49 | 0.17 | 0.75 | 0.57 | Power | 0.09 | -0.01 | 0.02 | 0.09 | -0.04 | -0.41 | 0.12 | 0 | 0.02 | 0.04 |
| 0.17 | 0.1 | -0.08 | -0.15 | -0.03 | -0.15 | -0.1 | SocialSec | 0.56 | 0.49 | 0.42 | 0.57 | 0.15 | 0.47 | 0.57 | 0.6 | 0.3 |
| -0.04 | -0.01 | -0.22 | -0.01 | -0.08 | -0.01 | -0.11 | 0.52 | PersSec | 0.37 | 0.32 | 0.52 | 0.45 | 0.33 | 0.38 | 0.47 | 0.17 |
| -0.03 | 0.03 | -0.1 | 0.04 | -0.08 | -0.07 | -0.07 | 0.41 | 0.29 | Health | 0.35 | 0.62 | 0.43 | 0.4 | 0.49 | 0.56 | 0.26 |
| 0.03 | -0.32 | -0.11 | -0.04 | -0.14 | -0.3 | -0.07 | -0.1 | 0.04 | 0.11 | Tradition | 0.56 | 0.28 | 0.46 | 0.45 | 0.56 | 0.28 |
| 0 | -0.12 | -0.14 | -0.07 | -0.28 | -0.37 | -0.26 | 0.32 | 0.41 | 0.46 | 0.27 | Conform | 0.73 | 0.63 | 0.69 | 0.64 | 0.42 |
| -0.59 | -0.39 | -0.53 | -0.31 | -0.68 | -0.66 | -0.45 | -0.04 | 0.35 | 0.33 | 0.17 | 0.64 | Humility | 0.39 | 0.47 | 0.51 | 0.34 |
| 0.28 | 0.07 | 0.14 | 0.23 | -0.05 | -0.47 | -0.21 | 0.39 | 0.27 | 0.31 | 0.38 | 0.39 | 0.22 | Benev | 0.74 | 0.73 | 0.63 |
| 0.1 | 0.06 | 0.03 | 0.11 | -0.04 | -0.3 | -0.23 | 0.55 | 0.33 | 0.47 | 0.35 | 0.56 | 0.38 | 0.56 | Equality | 0.69 | 0.64 |
| 0.04 | 0.11 | -0.09 | 0.11 | -0.06 | -0.34 | -0.05 | 0.5 | 0.4 | 0.5 | 0.46 | 0.44 | 0.36 | 0.68 | 0.71 | Nature | 0.45 |
| 0.11 | 0.12 | 0 | 0.16 | -0.08 | -0.46 | -0.13 | 0.18 | 0.06 | 0.15 | 0.16 | 0.22 | 0.22 | 0.53 | 0.41 | 0.34 | Tolerance |

*Figure 4. Correlations between values before (upper-triangular) and after (lower-triangular) the correction for measurement error for Estonian-speakers.*

Figures 4 and 5 are comparing correlations between 17 value types before and after the correction for measurement error within Estonian-speaking and Russian-speaking samples. As most of the relationships were tested repeatedly – first in small models in combinations of single value types, then in dimension-based models and finally in models combining dimensions, we had the possibility of combining these correlations, which increases the robustness of given measures[14].

It becomes apparent that for the Estonian-speakers almost all the correlations have decreased after correcting for measurement error and only the relations between *wealth* and *stimulation, achievement* and *stimulation*, *autonomy of action* and *hedonism*, and *equality* and *preserving nature* have slightly increased, which is probably related to the differences in selection of indicators.

[14] As we restricted model modifications between items of the same value dimension to be the same in the models of each dimension and in models combining the dimensions, the variation of size of the same correlations in different models was in most cases very small, on average .028.

Interestingly there is a clear pattern in the correlation transformations – the change is moderate (mostly less than .1) between collectivistic values (*humility, tradition, conformity, social security, personal security, health, equality*, *preserving nature, tolerance, benevolence*), but it is significantly higher between individualistic values and between individualistic and collectivistic values (mostly .2 or more). The latter is in line with the theory (Schwartz, 1992) that these values are internally conflicting. The largest systematic change is related to *humility*, whose correlation with *autonomy of thought* changed from -.16 to -.59. The former change between individualistic values is due to the fact that these correlations were mostly very high before the correction for measurement error.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AutoTh | 0.89 | 0.78 | 0.73 | 0.91 | 0.48 | 0.78 | 0.36 | 0.1 | 0.17 | 0.15 | 0.3 | -0.15 | 0.53 | 0.59 | 0.5 | 0.56 |
| 0.72 | AutoAct | 0.52 | 0.47 | 0.57 | 0.31 | 0.51 | 0.36 | 0.24 | 0.15 | 0.04 | 0.29 | 0.01 | 0.3 | 0.46 | 0.37 | 0.42 |
| 0.88 | 0.36 | Stimul | 0.9 | 0.75 | 0.46 | 0.82 | 0.04 | 0.29 | -0.07 | 0.12 | -0.1 | -0.33 | 0.17 | 0.04 | 0.08 | 0.19 |
| 0.71 | 0.54 | 0.85 | Hedonism | 0.73 | 0.56 | 0.68 | 0.1 | 0.14 | 0.09 | 0.09 | -0.05 | -0.25 | 0.22 | 0.08 | 0.1 | 0.05 |
| 0.83 | 0.49 | 0.71 | 0.75 | Achiev | 0.67 | 0.87 | 0.3 | 0.19 | 0.25 | 0.1 | 0.08 | -0.19 | 0.23 | 0.28 | 0.19 | 0.12 |
| 0.41 | 0.22 | 0.4 | 0.61 | 0.7 | Wealth | 0.57 | 0.13 | 0.08 | 0.11 | 0.03 | -0.19 | -0.37 | -0.2 | -0.06 | -0.13 | -0.23 |
| 0.43 | 0.31 | 0.35 | 0.46 | 0.6 | 0.45 | Power | 0.13 | 0.16 | -0.03 | 0.05 | -0.07 | -0.36 | 0.06 | 0.01 | 0.16 | 0.03 |
| 0.21 | 0.15 | -0.06 | -0.07 | 0.11 | -0.06 | -0.07 | SocialSec | 0.55 | 0.4 | 0.44 | 0.69 | 0.3 | 0.6 | 0.69 | 0.6 | 0.28 |
| 0.01 | 0.16 | -0.08 | 0.09 | 0.12 | -0.01 | 0.03 | 0.46 | PersSec | 0.33 | 0.4 | 0.48 | 0.13 | 0.34 | 0.39 | 0.36 | 0.23 |
| -0.07 | -0.05 | -0.16 | -0.07 | 0.03 | -0.02 | -0.22 | 0.29 | 0.26 | Health | 0.23 | 0.46 | 0.31 | 0.25 | 0.51 | 0.25 | 0.11 |
| -0.05 | -0.18 | -0.12 | -0.19 | -0.09 | -0.12 | -0.09 | 0.44 | 0.41 | 0.16 | Tradition | 0.55 | 0.33 | 0.48 | 0.38 | 0.5 | 0.34 |
| -0.47 | -0.16 | -0.42 | -0.64 | -0.54 | -0.57 | -0.4 | 0.39 | 0.49 | 0.14 | 0.49 | Conform | 0.77 | 0.87 | 0.87 | 0.61 | 0.62 |
| -0.66 | -0.33 | -0.52 | -0.66 | -0.57 | -0.7 | -0.56 | 0.12 | -0.03 | 0.17 | 0.21 | 0.63 | Humility | 0.52 | 0.55 | 0.52 | 0.48 |
| 0.14 | 0.06 | 0.16 | 0 | -0.05 | -0.52 | -0.38 | 0.59 | 0.32 | 0.14 | 0.49 | 0.53 | 0.43 | Benev | 0.96 | 0.72 | 0.7 |
| 0.16 | 0.19 | 0.07 | -0.18 | 0.18 | -0.13 | -0.43 | 0.52 | 0.42 | 0.44 | 0.43 | 0.58 | 0.4 | 0.75 | Equality | 0.73 | 0.59 |
| 0.09 | 0.18 | 0.03 | -0.11 | 0.11 | -0.29 | 0.08 | 0.52 | 0.3 | 0.11 | 0.47 | 0.42 | 0.36 | 0.65 | 0.66 | Nature | 0.45 |
| 0.24 | 0.31 | 0.21 | -0.15 | -0.02 | -0.44 | -0.23 | 0.18 | 0.17 | -0.03 | 0.28 | 0.52 | 0.32 | 0.63 | 0.58 | 0.36 | Tolerance |

*Figure 5. Correlations between values before (upper-triangular) and after (lower-triangular) the correction for measurement error for Russian-speakers.*

Despite the fact that the inter-value correlations differ quite a lot between Estonian-speakers and Russian-speakers, the effects of the correction for measurement error were, as expected, very similar, as the quality measures (Appendix 2) were also similar. For Russian-speakers, the correlations also changed more (by becoming more negative) between conflicting values (Table 3). In addition to *humility*, *conformity* also become much more contrasting with different individualistic values (for example the correlation between *conformity* and *autonomy of thought* changed from .3 to -.47).

When comparing correlations between Estonian-speakers and Russian-speakers after correction for measurement error, there are four relationships that differ more than .4. Two *self-direction* values (*autonomy of thought* and *autonomy of action*) have for Estonian-speakers a correlation of .31 while for Russian-speakers it is .72. This means that for Estonian-speakers there exists a conceptual difference between the given factors, while the Russian-speakers perceive them as more overlapping. Other large differences appear in the associations between *hedonism* and *conformity,* and *autonomy of thought* and *conformity*, which for Estonian-speakers are both non-significantly correlated (-.07 and 0), but for Russian-speakers they have a strongly negative relationship (-.64 and -.47), which fits more with the Schwartz theoretical structure. As discussed earlier these differences are probably related to the non-invariant conformity item *co28*, which has different cultural implications and significance for Estonian-speakers and Russian-speakers.

Finally, there is also a difference in correlations between *social security* and *traditions*, which for Estonian speakers have a slightly negative relation (-.1), but for Russian-speakers a strongly positive one (.44). As from a theoretical perspective these two values have similar motivational goals, the deviation appears among Estonian-speakers and probably it is again related with the *tradition* item *tr20*. As the majority of ethnic Estonians are opposed to religion, the *maintain tradition* value has abnormally low scores, which also affect the relationships.

After correction for measurement error, there are no more correlations higher than .9. For Estonian-speakers the correlation between *power* and *achievement* changes from .94 to .75 and between *hedonism* and *stimulation* from .91 to .73; for Russian-speakers the correlation of .96 between *benevolence* and *equality* decreased to .75 and the .91 between *autonomy of thought* and *achievement* decreased to .83. The decrease of correlations is on the one hand related to changes in the structure – slightly different selections of indicators and different modifications. Although in some extent it can also be related with subtraction of method effects (appendix 2).

## 4.4 Changes in the circular structure

Next we were interested in seeing whether the correction for measurement error had any effect on the ordering of value types in the value circle. We used multidimensional scaling (MDS) to map value types in the two-dimensional space for initial and corrected correlation matrices. Figures 6 and 7 present MDS outputs for Estonian-speakers and Russian-speakers, which are produced using the standard R function *cmdscale* (R Development Core Team 2009).

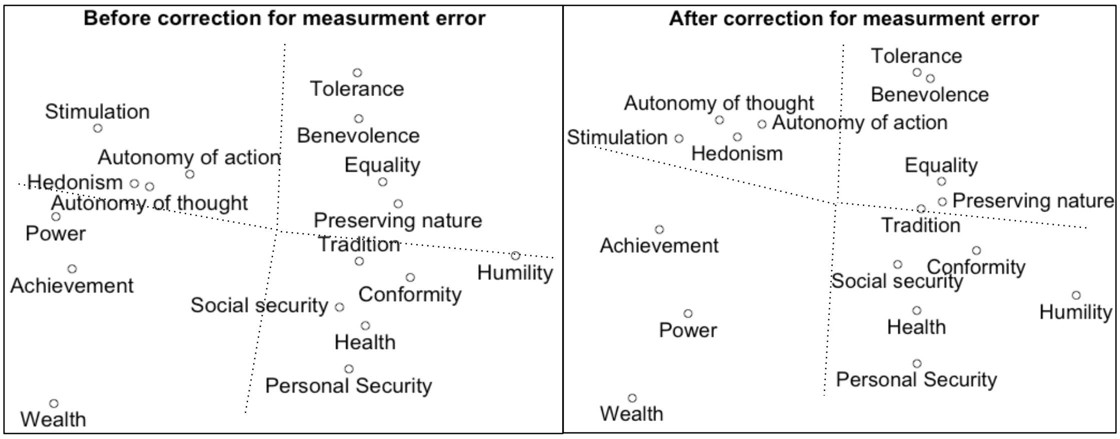


*Figure 6. MDS outputs before and after correction for measurement error for Estonian-speakers*

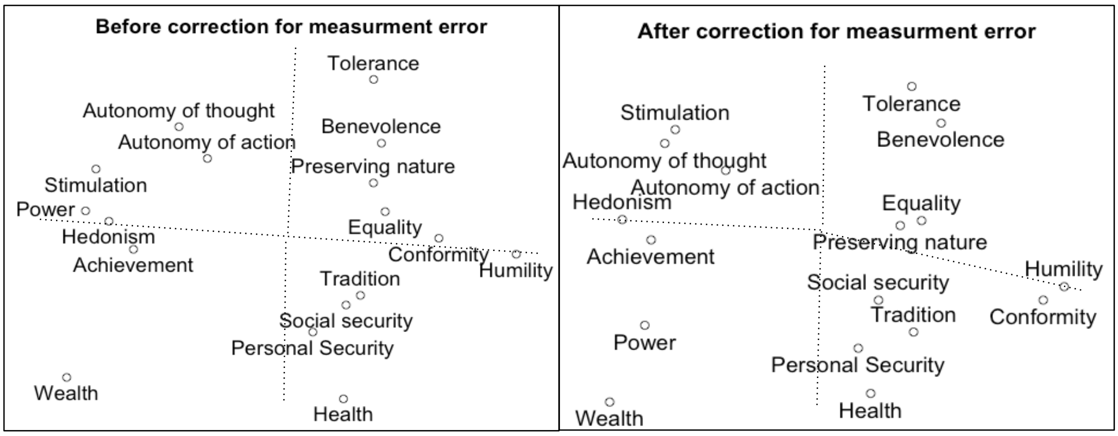


*Figure 7. MDS outputs before and after correction for measurement error for Russian-speakers*

For both samples, the MDS-based circles differ from the structure proposed by Schwartz and colleagues (2012), although they become more similar after correction for measurement error.

For Estonian-speakers the ordering of individualistic values before correction of measurement error is: *stimulation, autonomy of action, autonomy of thought, hedonism, power, achievement* and *wealth*, while after correction for measurement error, it changes to: *autonomy of action, autonomy of thought, hedonism, stimulation, achievement, power and wealth*, which fits with the theoretical structure. The ordering of collectivistic values does not change as a result of the correction process and the main difference with the theoretical structure is that *benevolence* has been located between universalistic values, and *tradition* between *conformity* and *nature*.

For Russian-speakers, *stimulation* is misplaced after correction for measurement error, but otherwise the individualistic side of the value continuum fits with the theoretical ordering. Among collectivistic values, *benevolence* is again very near to *tolerance*, while *traditionalism* has been located between the security values. Therefore it seems that in addition to the clearer contrast between opposing values, the ordering also fits more closely with the theoretical structure after correction for measurement error.

### 4.5 Relationships with external variables

Finally, we tested if the correction for measurement error has an effect on correlations between values and external variables. To do so, we chose 2 variables – *income* (7 categories) and *left-right self-placement* (10-point scale) and corrected them for measurement error[15] while using the procedure described earlier (figure 1).

Schwartz (2006) has seen higher *income* as a mean to choose one's life-style while worrying less about security threats. He has additionally argued that economic well-being reduces a need to restrict one's impulses and to maintain supportive, traditional ties. Based on these assumptions, we except that higher income would promote importance of values like *achievement, stimulation, hedonism, autonomy of thought* and *autonomy of action*, and would reduce importance of *personal and societal security, conformity, humility* and *tradition* values.

Based on earlier studies (Caprara et al. 2006, Caprara et al. 2008, Thorisdottir et al. 2007, Schwartz et al. 2010) we can except that in the Eastern European context the *traditionalism, conformity, humility* along with *power* and *achievement* should predict right wing placement while *benevolence* and universalistic values, like *equality*, *preserving nature, tolerance* should

[15] Based on SQP2 predictions for *income* the reliability coefficient was .74 and validity coefficient .84, and for *left-right self-placement* the same coefficients were respectively .83 and .88.

predict left-wing placement. For openness values (*hedonism, stimulation, autonomy of thought* and *autonomy of action*) the previous results have been mixed, which means that both scenarios are possibilities (Thorisdottir et al. 2007, Aspelund et al. 2013).

Table 1 shows the correlations between *income* and *left-right placement* and 17 values before and after the correction for measurement error. For both variables, the first column contains correlations before correction for measurement error and the second column correlations after correction for measurement error.

*Table 1. Correlations between the values and the age before and after the correction for measurement error.*

| | Correlations between values & income | | Correlations between values & left-right self-placement | |
|---|---|---|---|---|
| | Before correction for measurement error | After correction for measurement error | Before correction for measurement error | After correction for measurement error |
| Autonomy of thought | .139** | .188** | .089 | .130** |
| Autonomy of action | .054 | .072* | .161** | .247** |
| Stimulation | .062 | .106** | -.024 | -.059* |
| Hedonism | .056 | .078* | .077 | .141** |
| Achievement | .121** | .171** | .074 | .077* |
| Wealth | .069* | .119** | .061 | .109** |
| Power | .122** | .208** | -.073 | -.127** |
| Societal security | .028 | .047 | .071 | .123** |
| Personal security | -.044 | -.074* | .051 | .088* |
| Health | -.023 | -.039 | -.052 | -.093* |
| Tradition | -.021 | -.034 | -.124** | -.221** |
| Conformity | -.111** | -.177** | -.034 | -.037 |
| Humility | -.184** | -.307** | -.081* | -.141** |
| Benevolence | -.105* | -.134** | .032 | .047 |
| Equality | -.022 | -.039 | -.061 | -.087* |
| Preserving nature | -.06 | -.11** | -.037 | -.065 |
| Tolerance | .006 | .011 | .007 | .012 |

**Correlation is significant at the 0.05 level; **Correlation is significant at the 0.01 level*

As it can be seen from table 2, all the correlations between values and the external variables get stronger after correction for measurement error. As majority of the relationships are rather weak, the increases in correlations are mostly modest, although the results are still substantially quite interesting.

Based on the raw data, the Schwartz' hypothesis on relations between *income* and value preferences are only partially confirmed. Respondents with lower income do hold indeed significantly more importance on *conformity, humility and benevolence* values while respondents with high income emphasize *autonomy of thought*, *achievement* and interestingly *power* values. However in contradiction with his hypothesis, there do not exist significant correlations between *income* and *stimulation, hedonism, autonomy of action*, *personal security* and *maintaining tradition*. After correction for measurement error all these correlations, except one with tradition, become statistically significant. The deviation with tradition might be related to overall low scoring of tradition item measuring importance of religion as discussed earlier.

The effect of correction for measurement error becomes even more evident in the context of values and *left-right placement*. Before correction for measurement error, there appears only 3 statistically significant correlations – *autonomy of action* correlates positively with right wing placement and *tradition* and *humility* with left wing placement. However after correction for measurement error 13 correlations out of 17 stand out as statistically significant. Based on given results in Estonia right-wing placement correlates positively (besides autonomy of action) with higher assessment of *autonomy of thought, hedonism, achievement, wealth, societal* and *personal security* while left-wing placement correlates positively with higher assessment of *stimulation, power, tradition, humility* and *equality*. Interestingly many of these correlations contradict with expectations. Latter can be related with peculiarity of Estonian party-system (Kulik & Pshizova), which in turn shapes citizens perception of left-right wing distinction.

## 5 Conclusions

This study examined the effect of the measurement error on the structure of basic human values and found that based on the Estonian representative sample the configural structure of basic values remained similar, although the final selection of value indicators differed a bit, while the correlation between the values changed substantially. On the one hand the change was related to relatively low quality of value items and on the other hand the existence of a considerable common method bias, due to the same response scale and similar formulation of value questions. In general the relations between the values were more in line with the theoretical structure after correction for the measurement error – while some very high positive correlations decreased, the theory-based opposition of conflicting values (Schwartz, 1992) became much more evident through the stronger negative correlations between opposing values.

Secondly, the correction of the model's measurement part allowed for testing the equivalence of strictly cognitive differences, which revealed 7 non-invariant loadings between Estonian-speaking and Russian-speaking samples. These results are very interesting, as analyses of the same dataset without correcting for measurement error showed full metric equivalence (Lilleoja & Saris 2014), which is a similar result to most of the earlier cross-cultural value comparisons obtained (example Davidov & Schmidt, 2007; Davidov, 2010; Davidov & Depner, 2011; Cieciuch & Davidov, 2012). It seems that in the context of basic values, the measurement error tends to hide the real cognitive differences across cultural groups. The revealed differences in understanding of the same items by Estonian-speakers and Russian-speakers also helped to explain some of the substantial differences in the relationships between the values.

While comparing the circular ordering of value types before and after correction for measurement error, we found that for both Estonian-speakers and Russian-speakers, the congruity with the theoretical value structure proposed by Schwartz and colleagues (2012), increased after the correction procedure.

And finally, when analysing relationships between the values and external variables (income and left-right self placement) before and after correction for the measurement error, we saw that all correlations became stronger. And even more importantly, many of relations that were statistically insignificant before, became significant after correction for measurement error, which changed results substantially.

All these aspects demonstrate that the correction for measurement error provides additional support for the validity of the new value structures and sheds some new light on their cross-cultural equivalence. These results emphasize also the general importance of the correction for measurement error while studying the relationships between different concepts. Latter is especially true, if the quality of variables is low, which now can easily be detected and if needed, corrected with the help of SQP2.

All in all, the correction for measurement error, not only for random errors, but also for method effects, provides a clear effect on the structure of basic human values, confirming more convincingly that the tested model holds in the Estonian context and it is, with some limitations, comparable across cultural groups.

## 5.1 Limitations and Future Research

This paper applied the correction for the measurement error, although there exists also other potential sources for errors, like nonresponse, coverage problems, data collection, etc., which all can have an additional effect on the final results. Therefore, in future research these aspects could be addressed as well.

The current study is based on the PVQ-40 value scale, which lacks indicators for some of the value types and it also has some problems with some phrasings (double-barrelled wording, etc.). As Schwartz and colleagues have recently developed a new value scale (PVQ-R) with a changed wording and a larger number of indicators (Schwartz et al., 2012), it would be interesting to repeat this study on the basis of the results of the new questionnaire. Latter would also be an important validation of the new value structure.

In the light of our results, we once more encourage researchers to use the procedure of correction for the measurement error, which can bring new insights to even extensively explored topics.

## References


Alwin, D. F. (2007). Margins of Error: A Study of Reliability in Survey Measurement. Hoboken, NJ: John Wiley & Sons.

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement the influence of question and respondent attributes. *Sociological Methods & Research*, *20*(1), 139-181.

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public opinion quarterly*, *48*(2), 409-442.

Aspelund, A., Lindeman, M., & Verkasalo, M. (2013). Political Conservatism and Left-Right Orientation in 28 Eastern and Western European Countries: Conservatism and Left-Right. *Political Psychology*, *34*(3), 409–417. doi:10.1111/pops.12000

Beierlein, C., Davidov, E., Schmidt, P., Schwartz, S. H., & Rammstedt, B. (2012). Testing the discriminant validity of Schwartz' portrait value questionnaire items—a replication end extension of Knoppen and Saris (2009). Survey Research Methods, 6(1), 25–36.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, *56*(2), 81.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*, *100*(3), 525.

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., & Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political psychology*, *27*(1), 1-28.

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*(1), 37-48.

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

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

Davidov, E. (2010). Testing for comparability of human values across countries and time with the third round of the European Social Survey. *International Journal of Comparative Sociology*, *51*(3), 171-191.

Davidov, E. & Schmidt, P. (2007). *Are values in the Benelux countries comparable? Testing for equivalence with the European Social Survey.* In: Loosveldt, G; Swyngedouw, M; Cambré, B. Measuring Meaningful Data in Social Research. Leuven, 373-386.

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

Davidov, E., Schmidt, P., & Schwartz, S. (2008). Bringing values back in: the adequacy of the European social survey to measure values in 20 countries. *Public opinion quarterly, 72*(3), 420–445.

De Castellarnau, A., & Saris, W. E. (2014). A simple procedure to correct for measurement errors in survey research. European Social Survey Education Net (ESS EduNet). Retrieved from http://essedunet.nsd.uib.no/cms/topics/measurement/

De Clercq, S., & Fontaine, J. R. J. (2006). Extending the Schwartz value theory for assessing supplementary person-organization fit. *Unpublished PhD. dissertation, Universiteit Ghent, Belgium*.

Fischer, R., Vauclair, C. M., Fontaine, J. R. J., & Schwartz, S. H. (2010). Are individual-level and country- level value structures different? Testing Hofstede's legacy with the Schwartz Value Survey. *Journal of Cross-Cultural Psychology, 41*(2), 135–151.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American sociological review*, 93-101.

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

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

Knoppen, D., & Saris, W. E. (2009c). Schwartz's Theory of Human Values: Balancing Homogeneity of Reflective Items and Theoretical Coverage. *RECSM Working Paper 9*. Retrieved from http://www.upf.edu/survey/_pdf/RECSM_wp009.pdf

Költringer, R. (1995). Measurement Quality in Austrian Personal Interview Surveys. In W. E. Saris & A. Münnich (Eds.), *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments* (pp. 207–224). Budapest: Eötvös University Press.

Kulik, A., & Pshizova, S. (2005). *Political parties in post-soviet space: Russia, Belarus, Ukraine, Moldova, and the Baltics* (Vol. 5). Greenwood Publishing Group.

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*(1), 153-172. doi:10.1007/s11205-013-0272-4

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores* (pp. 395-479). Reading, MA: Addison-Wesley.

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

Mohler, P. P., & Wohn, K. (2005). Personliche Wertorientierungen im European Social Survey. ZUMA- Arbeitsbericht 2005/01.

Oberski, D. (2010). Jrule for Mplus. Retrieved from http://wiki.github.com/daob/JruleMplus/

Oberski, D.T., Gruner, T., & Saris, W. E. (2011). The prediction of the quality of the questions. In W. E Saris, D. Oberski, M. Revilla, D. Zavalla, L. Lilleoja, I. Gallhofer, & T. Grüner. (Eds.), The development of the Program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working paper 24, chapter 6,* Universitat Pompeu Fabra, Barcelona

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

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

R Development Core Team. (2009). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Ringvee, R. (2014). Religion: Not declining but changing. What do the population censuses and surveys say about religion in Estonia? *Religion*, *44*(3), 502-515.

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

Saris, W. E. (2001). The strength of the causal relationship between living conditions and satisfaction. *Sociological Methods & Research*, *30*(1), 11-34.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research ( 2nd ed.)*. Wiley, Hoboken, NJ.

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

Saris, W., & Revilla, M. (2015). Correction for Measurement Errors in Survey Research: Necessary and Possible. *Social Indicators Research*, 1–16. http://doi.org/10.1007/s11205-015-1002-x

Saris, W., Knoppen, D. and Schwartz, S.H. (2013). "Operationalizing the Theory of Human Values: Balancing Homogeneity of Reflective Items and Theoretical Coverage", *Survey Research Methods*, 7(1); pp. 29-44.

Scherpenzeel, A.C. (1995). The quality of indicators of satisfaction across Europe: A meta-analysis of multitrait-multimethod studies. In: Scherpenzeel, A.C. A Question of Quality. Evaluating survey questions by multitrait-multimethod studies. Dissertation. Royal PTT Nederland NV, KPN Research.

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

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), 1–65.

Schwartz, S. H. (1994). Are There Universal Aspects in the Structure and Contents of Human Values? *Journal of Social Issues*, *50*(4), 19–45. doi:10.1111/j.1540-4560.1994.tb01196.x

Schwartz, S. H. (2005). Robustness and Fruitfulness of a Theory of Universals in individual human values. In A. Tamayo & J. B. Porto (Eds.), Valores e comportamento nas Organizações [Values and Behavior in Organizations] (pp. 56–95). Petro´polis, Brazil: Vozes.

Schwartz, S. H. (2006). Les valeurs de base de la personne: The´orie, mesures et applications [Basic human values: Theory, measurement, and applications]. Revue Française de Sociologie, 47, 249–288.

Schwartz, S. H. (2007). Value orientations: measurement, antecedents and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally. Lessons from the European Social Survey* (p. 169-203). London: Sage Publications.

Schwartz, S. H. (2011). Studying Values: Personal adventure, future directions. Journal of *Cross-Cultural Psychology, 42*(2), 307–319.

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

Schwartz, S. H., & 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., 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., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., ... & Konty, M. (2012). Refining the theory of basic individual values. *Journal of personality and social psychology*, *103*(4), 663.

Sõmer, M. (2011). The stability of Estonian value hierarchies based on two Schwartz value surveys. In I. Tart (Ed.), *Basic human values in Estonia and Baltic Sea Countries* (pp. 137–147). Tartu: Tartu University Press.

Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.

Survey Quality Predictor (2015). SQP Users' Manual. Barcelona, Universitat Pompeu Fabra. Retrieved from http://sqp.upf.edu/media/files/sqp_users_manual.pdf

Thorisdottir, H., Jost, J.T., Liviatan, I., & Shrout, P. (2007). Psychological needs and values underlying left-right political orientation: Cross-national evidence from Eastern and Western Europe. *Public Opinion Quarterly,* 71, 175-203.

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

Vecchione, M., Schwartz, S. H., Caprara, G. V., Schoen, H., Cieciuch, J., Silvester, J., ... Alessandri, G. (2015). Personal values and political activism: A cross-national study. *British Journal of Psychology, 106*, 84–106.

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, 112-117.

Appendix 1. The new ordering of the items and values of the PVQ based on Estonian sample (Lilleoja & Saris 2014).

| Abbreviation | Value items | Value types based on study of Lilleoja & Saris (2014) | Affiliation after correction for measurement error |
|---|---|---|---|
| be12 | It's very important to him to help the people around him. He wants to care for other people. | Benevolence | Benevolence |
| be18 | It is important to him to be loyal to his friends. He wants to devote himself to people close to him. | Benevolence | Benevolence |
| be27 | It is important to him to respond to the needs of others. He tries to support those he knows. | Benevolence | - |
| be33 | Forgiving people who might have wronged him is important to him. He tries to see what is good in them and not to hold a grudge. | Benevolence | Benevolence |
| un3 | He thinks it is important that every person in the world be treated equally. He wants justice for everybody, even for people he doesn't know. | Social equality | Social equality |
| un8 | It is important to him to listen to people who are different from him. Even when he disagrees with them, he still wants to understand them. | Tolerance | Tolerance |
| un19 | He strongly believes that people should care for nature. Looking after the environment is important to him. | Preserving nature | Preserving nature |
| un23 | He believes all the worlds' people should live in harmony. Promoting peace among all groups in the world is important to him. | - | Social equality |
| un29 | He wants everyone to be treated justly, even people he doesn't know. It is important to him to protect the weak in society. | Social equality | Social equality |
| un40 | It is important to him to adapt to nature and to fit into it. He believes that people should not change nature. | Preserving nature | Preserving nature |
| sd1 | Thinking up new ideas and being creative is important to him. He likes to do things in his own original way. | Autonomy of thought | Autonomy of thought |
| sd11 | It is important to him to make his own decisions about what he does. He likes to be free to plan and to choose his activities for himself. | Autonomy of action | Autonomy of action |
| sd22 | He thinks it's important to be interested in things. He likes to be curious and to try to understand all sorts of things. | Autonomy of action | Autonomy of action |
| sd34 | It is important to him to be independent. He likes to rely on himself. | Autonomy of thought | Autonomy of thought |
| st6 | He thinks it is important to do lots of different things in life. He always looks for new things to try. | - | Stimulation |
| st15 | He likes to take risks. He is always looking for adventures. | Stimulation | - |
| st30 | He likes surprises. It is important to him to have an exciting life. | Stimulation | Stimulation |
| he10 | He seeks every chance he can to have fun. It is important to him to do things that give him pleasure. | Hedonism | - |
| he26 | Enjoying life's pleasures is important to him. He likes to 'spoil' himself. | Hedonism | Hedonism |
| he37 | He really wants to enjoy life. Having a good time is very important to him. | Hedonism | Hedonism |
| ac4 | It's very important to him to show his abilities. He wants people to admire what he does. | Achievement | Achievement |
| ac13 | Being very successful is important to him. He likes to impress other people. | Achievement | Achievement |
| ac24 | He thinks it is important to be ambitious. He wants to show how capable he is. | Achievement | - |
| ac32 | Getting ahead in life is important to him. He strives to do better than others. | Achievement | Achievement |
| po2 | It is important to him to be rich. He wants to have a lot of money and expensive things. | Wealth | Wealth |
| po17 | It is important to him to be in charge and tell others what to do. He wants people to do what he says. | Power | Power |
| po39 | He always wants to be the one who makes the decisions. He likes to be the leader. | Power | - |
| se5 | It is important to him to live in secure surroundings. He avoids anything that might endanger his safety. | Personal security | Personal security |
| se14 | It is very important to him that his country be safe. He thinks the state must be on watch against threats from within and without. | Societal security | Societal security |
| se21 | It is important to him that things be organized and clean. He doesn't want things to be a mess. | - | - |
| se31 | He tries hard to avoid getting sick. Staying healthy is very important to him. | Health | Health |
| se35 | Having a stable government is important to him. He is concerned that the social order be protected. | Societal security | Societal security |
| co7 | He believes that people should do what they're told. He thinks people should follow rules at all times, even when no-one is watching. | Conformity | Conformity |
| co16 | It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong. | Conformity | Conformity |
| co28 | It is important to him to be obedient. He believes he should always show respect to his parents and to older people. | Conformity | Conformity |
| co36 | It is important to him to be polite to other people all the time. He tries never to disturb or irritate others. | Conformity | - |
| tr9 | He thinks it's important **not** to ask for more than what you have. He believes that people should be satisfied with what they have. | Humility | Humility |
| tr20 | Religious belief is important to him. He tries hard to do what his religion requires. | - | Maintain traditions |
| tr25 | He believes it is best to do things in traditional ways. It is important to him to follow the customs he has learned. | Maintain traditions | Maintain traditions |
| tr38 | It is important to him to be humble and modest. He tries not to draw attention to himself. | Humility | Humility |

Appendix 2. Reliability and validity and method-effects predictions obtained by SQP2.

| Abbreviation | Reliability coefficient | | Validity coefficient | | Method-effect | |
|---|---|---|---|---|---|---|
| | Estonian-speakers | Russian-speakers | Estonian-speakers | Russian-speakers | Estonian-speakers | Russian-speakers |
| sd1 | .82 | .84 | .93 | .93 | .37 | .37 |
| po2 | .8 | .81 | .92 | .93 | .38 | .36 |
| un3 | .79 | .8 | .93 | .93 | .38 | .36 |
| ac4 | .79 | .8 | .93 | .93 | .38 | .36 |
| se5 | .81 | .82 | .94 | .94 | .35 | .33 |
| st6 | .8 | .81 | .93 | .94 | .36 | .34 |
| co7 | .8 | .81 | .93 | .94 | .36 | .34 |
| un8 | .81 | .82 | .94 | .94 | .35 | .33 |
| tr9 | .81 | .82 | .93 | .94 | .36 | .34 |
| he10 | .8 | .81 | .93 | .94 | .36 | .34 |
| sd11 | .81 | .82 | .93 | .94 | .36 | .34 |
| be12 | .81 | .82 | .94 | .94 | .35 | .33 |
| ac13 | .81 | .82 | .93 | .94 | .36 | .34 |
| se14 | .82 | .83 | .92 | .93 | .39 | .38 |
| st15 | .81 | .82 | .93 | .94 | .36 | .34 |
| co16 | .8 | .81 | .93 | .94 | .36 | .34 |
| po17 | .81 | .82 | .93 | .94 | .35 | .33 |
| be18 | .81 | .82 | .93 | .94 | .36 | .34 |
| un19 | .79 | .8 | .92 | .93 | .38 | .36 |
| tr20 | .79 | .8 | .92 | .93 | .38 | .36 |
| se21 | .79 | .8 | .92 | .94 | .38 | .36 |
| sd22 | .8 | .81 | .93 | .93 | .36 | .34 |
| un23 | .81 | .81 | .93 | .94 | .36 | .34 |
| ac24 | .81 | .82 | .93 | .94 | .38 | .36 |
| tr25 | .8 | .81 | .92 | .93 | .37 | .36 |
| he26 | .79 | .80 | .92 | .93 | .38 | .36 |
| be27 | .81 | .82 | .92 | .93 | .38 | .36 |
| co28 | .82 | .83 | .92 | .93 | .36 | .34 |
| un29 | .81 | .82 | .92 | .93 | .38 | .36 |
| st30 | .8 | .81 | .93 | .94 | .38 | .35 |
| se31 | .8 | .81 | .93 | .94 | .38 | .35 |
| ac32 | .79 | .80 | .93 | .94 | .36 | .34 |
| be33 | .81 | .81 | .93 | .93 | .38 | .35 |
| sd34 | .82 | .83 | .93 | .93 | .38 | .35 |
| se35 | .8 | .81 | .93 | .93 | .38 | .35 |
| co36 | .8 | .81 | .93 | .94 | .37 | .35 |
| he37 | .82 | .83 | .92 | .93 | .39 | .38 |
| tr38 | .81 | .82 | .93 | .94 | .36 | .34 |
| po39 | .8 | .81 | .93 | .94 | .36 | .34 |
| un40 | .81 | .82 | .93 | .94 | .35 | .33 |

*Appendix 3. Comparison of model fits*

| Dimensions | Chi2 | DF | CFI | RMSEA | Nr of modifications | Nr of remaining misspecifications[1] |
|---|---|---|---|---|---|---|
| Openness to change (A) | 50 | 18 | .994 | .054 | 9 | 1 (10/1) |
| Self-enhancement (B) | 83 | 13 | .989 | .093 | 6 | 1 (4/6) |
| Conservation (C) | 294 | 58 | .947 | .081 | 8 | 5(19/38) |
| Self-transcendence (D) | 264 | 58 | .970 | .076 | 15 | 4(13/18) |
| A + B | 516 | 66 | .965 | .105 | 22 | 4(16/17) |
| C + D | 1012 | 200 | .949 | .081 | 47 | 9(30/73) |
| A + C | 799 | 166 | .950 | .078 | 35 | 0(45/59) |
| B + D | 1124 | 152 | .950 | .102 | 46 | 11(26/60) |
| A + D | 1343 | 186 | .933 | .100 | 51 | 9(43/48) |
| B + C | 1729 | 175 | .904 | .120 | 29 | 18(47/107) |

[1] Number of matching misspecifications (std. epc. >.1) between samples (number of misspecifications in Estonian-speaking sample/ number of misspecifications in Russian-speaking sample)