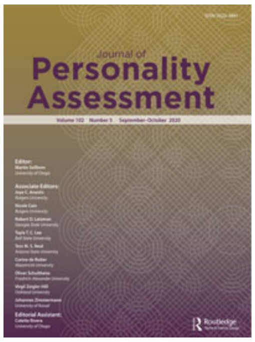

Journal of Personality Assessment

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

The HEXACO–100 Across 16 Languages: A Large-Scale Test of Measurement Invariance

Isabel Thielmann, Nazar Akrami, Toni Babarović, Amparo Belloch, Robin Bergh, Antonio Chirumbolo, Petar Čolović, Reinout E. de Vries, Daniel Dostál, Marina Egorova, Augusto Gnisci, Timo Heydasch, Benjamin E. Hilbig, Kung-Yu Hsu, Paweł Izdebski, Luigi Leone, Bernd Marcus, Janko Međedović, János Nagy, Oksana Parshikova, Marco Perugini, Boban Petrović, Estrella Romero, Ida Sergi, Kang-Hyun Shin, Snežana Smederevac, Iva Šverko, Piotr Szarota, Zsofia Szirmák, Arkun Tatar, Akio Wakabayashi, S. Arzu Wasti, Tereza Záškodná, Ingo Zettler, Michael C. Ashton & Kibeom Lee

To cite this article: Isabel Thielmann, Nazar Akrami, Toni Babarović, Amparo Belloch, Robin Bergh, Antonio Chirumbolo, Petar Čolović, Reinout E. de Vries, Daniel Dostál, Marina Egorova, Augusto Gnisci, Timo Heydasch, Benjamin E. Hilbig, Kung-Yu Hsu, Paweł Izdebski, Luigi Leone, Bernd Marcus, Janko Međedović, János Nagy, Oksana Parshikova, Marco Perugini, Boban Petrović, Estrella Romero, Ida Sergi, Kang-Hyun Shin, Snežana Smederevac, Iva Šverko, Piotr Szarota, Zsofia Szirmák, Arkun Tatar, Akio Wakabayashi, S. Arzu Wasti, Tereza Záškodná, Ingo Zettler, Michael C. Ashton & Kibeom Lee (2020) The HEXACO–100 Across 16 Languages: A Large-Scale Test of Measurement Invariance, Journal of Personality Assessment, 102:5, 714-726, DOI: 10.1080/00223891.2019.1614011

To link to this article: https://doi.org/10.1080/00223891.2019.1614011

JOURNAL OF PERSONALITY ASSESSMENT
2020, VOL. 102, NO. 5, 714–726
https://doi.org/10.1080/00223891.2019.1614011

The HEXACO–100 Across 16 Languages: A Large-Scale Test of Measurement Invariance

Isabel Thielmann[1], Nazar Akrami[2], Toni Babarović[3], Amparo Belloch[4], Robin Bergh[5], Antonio Chirumbolo[6], Petar Čolović[7], Reinout E. de Vries[8], Daniel Dostál[9], Marina Egorova[10], Augusto Gnisci[11], Timo Heydasch[12], Benjamin E. Hilbig[1], Kung-Yu Hsu[13], Paweł Izdebski[14], Luigi Leone[15], Bernd Marcus[16], Janko Međedović[17], János Nagy[18], Oksana Parshikova[10], Marco Perugini[19], Boban Petrović*[17], Estrella Romero[20], Ida Sergi[11], Kang-Hyun Shin[21], Snežana Smederevac[7], Iva Šverko[3], Piotr Szarota[22], Zsofia Szirmák[23], Arkun Tatar[24], Akio Wakabayashi[25], S. Arzu Wasti[26], Tereza Záškodná[9], Ingo Zettler[27], Michael C. Ashton[28], and Kibeom Lee[29]

[1]Department of Psychology, University of Koblenz-Landau, Landau, Germany; [2]Department of Psychology, Uppsala University, Uppsala, Sweden; [3]Ivo Pilar Institute of Social Sciences, Zagreb, Croatia; [4]Department of Personality, Evaluation, and Psychological Treatments, Universidad de Valencia, Valencia, Spain; [5]Harvard University; [6]Deparment of Psychology, Sapienza University of Rome, Rome, Italy; [7]Department of Psychology, Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia; [8]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam & Department of Educational Science, University of Twente, The Netherlands; [9]Palacky University Olomouc, Olomouc, Czech Republic; [10]Lomonosov Moscow State University, Moscow, Russia; [11]Department of Psychology, University of Campania "Luigi Vanvitelli," Caserta, Italy; [12]Department of Work and Organisational Psychology, University of Hagen, Hagen, Germany; [13]Department of Psychology, National Chung Cheng University, Chiayi County, Taiwan; [14]Institute of Psychology, Department of General and Health Psychology, Kazimierz Wielki University, Bydgoszcz, Poland; [15]Department of Social & Developmental Psychology, Sapienza University of Rome, Rome, Italy; [16]Department of Business Administration, University of Rostock, Rostock, Germany; [17]Institute of Criminological and Sociological Research, Belgrad, Serbia; [18]Eötvös Loránd University, Budapest, Hungary; [19]University of Milan-Bicocca, Milan, Italy; [20]University of Santiago de Compostela, Santiago de Compostela, Spain; [21]Ajou University, Suwon, South Korea; [22]Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland; [23]Independent Practice, Berlin, Germany; [24]Fatih Sultan Mehmet University, Istanbul, Turkey; [25]Chiba University, Inage-Ku, Chiba, Japan; [26]Sabanci University, Istanbul, Turkey; [27]University of Copenhagen, Copenhagen, Denmark; [28]Department of Psychology, Brock University, St. Catharines, Ontario, Canada; [29]Department of Psychology, University of Calgary, Alberta, Canada

ABSTRACT

The HEXACO Personality Inventory–Revised (HEXACO–PI–R) has become one of the most heavily applied measurement tools for the assessment of basic personality traits. Correspondingly, the inventory has been translated to many languages for use in cross-cultural research. However, formal tests examining whether the different language versions of the HEXACO–PI–R provide equivalent measures of the 6 personality dimensions are missing. We provide a large-scale test of measurement invariance of the 100-item version of the HEXACO–PI–R across 16 languages spoken in European and Asian countries ($N = 30{,}484$). Multigroup exploratory structural equation modeling and confirmatory factor analyses revealed consistent support for configural and metric invariance, thus implying that the factor structure of the HEXACO dimensions as well as the meaning of the latent HEXACO factors is comparable across languages. However, analyses did not show overall support for scalar invariance; that is, equivalence of facet intercepts. A complementary alignment analysis supported this pattern, but also revealed substantial heterogeneity in the level of (non)invariance across facets and factors. Overall, results imply that the HEXACO–PI–R provides largely comparable measurement of the HEXACO dimensions, although the lack of scalar invariance highlights the necessity for future research clarifying the interpretation of mean-level trait differences across countries.

ARTICLE HISTORY
Received 5 October 2018
Accepted 1 April 2019

The development of instruments assessing basic personality traits has been a vital cornerstone for the study of personality and individual differences. Modern construct-based methods of test construction were being developed and refined (e.g., Jackson, 1970, 1971) at a time when even the existence of personality traits was questioned (e.g., Mischel, 1968). In the 1980s and 1990s, researchers began to focus personality assessment on the five broad dimensions called the Big Five (e.g., Goldberg, 1990), and consequently instruments like the NEO Personality Inventory–Revised (NEO PI–R; Costa & McCrae, 1992) and the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991) became widely used in psychological research. Indeed, use of such inventories assessing basic personality traits led to seminal insights on the relevance of personality for numerous real-life outcomes, including physical and psychological health, quality of interpersonal relationships, and job performance (e.g., Ozer & Benet-Martínez, 2006).

CONTACT Isabel Thielmann thielmann@uni-landau.de Department of Psychology, University of Koblenz-Landau, Fortstraße 7, Landau 76829, Germany.
*We are sad to inform you that our friend and colleague, Boban Petrović, has passed away during the publication of this article.

More recently, however, lexical studies of personality structure across a variety of languages have revealed that the largest replicable factor space consists of six rather than five dimensions. These six dimensions are captured in the HEXACO model of personality structure (e.g., Ashton & Lee, 2007; Ashton, Lee, & De Vries, 2014; Ashton et al., 2004) and assessed via the corresponding HEXACO Personality Inventory-Revised (HEXACO–PI–R; Lee & Ashton, 2004, 2006; see Moshagen, Thielmann, Hilbig, & Zettler, in press, for a recent meta-analysis). Although the HEXACO model was not developed through any a priori modification of the Big Five, the HEXACO factors can be understood as a six-factorial adaptation and extension of the Big Five personality traits: Whereas the model basically maintains three factors that closely reflect their Big Five counterparts (i.e., Extraversion, Conscientiousness, and Openness to Experience), it incorporates major changes within the other three factors by implementing rotated versions of Neuroticism (called Emotionality in the HEXACO model) and Agreeableness and by further adding a sixth factor termed Honesty-Humility. In particular, Honesty-Humility as operationalized by the HEXACO–PI–R encompasses the facets sincerity, fairness, greed-avoidance, and modesty and thus reflects individual differences in morality and prosociality. As such, it most closely aligns with Big Five Agreeableness, but captures additional content that is not fully accommodated by the Big Five (e.g., Ashton & Lee, 2008; Lee, Ogunfowora, & Ashton, 2005). HEXACO Agreeableness, in turn, also shares some content with its same-named Big Five counterpart (e.g., gentleness), but lacks the sentimentality-related aspects of this factor and instead captures (at its low pole) the anger-related aspects included in Big Five Neuroticism. Conversely, Emotionality shares some of its content with Big Five Neuroticism apart from the anger-related aspects of the latter (which are now captured by [low] HEXACO Agreeableness) and also adds the sentimentality-related aspects of Big Five Agreeableness.

Since its emergence, the HEXACO model and inventory have gained considerable attention in psychological research and beyond, and their dissemination is still steadily increasing (for an overview of research on the HEXACO model, see www.hexaco.org). Most prominently, studies have focused on the Honesty-Humility factor and consistently demonstrated its potential in accounting for individual variation in a variety of criteria related to moral and prosocial traits and behaviors, such as dishonesty and cheating (e.g., Heck, Thielmann, Moshagen, & Hilbig, 2018; Hershfield, Cohen, & Thompson, 2012; Hilbig & Zettler, 2015), delinquency (e.g., Cohen, Panter, Turan, Morse, & Kim, 2014; Dunlop, Morrison, Koenig, & Silcox, 2012; Marcus, Lee, & Ashton, 2007; Ogunfowora, Bourdage, & Nguyen, 2013), prosociality (e.g., Hilbig & Zettler, 2009; Thielmann & Böhm, 2016; Thielmann & Hilbig, 2015; Zhao, Ferguson, & Smillie, 2017; Zhao & Smillie, 2015), and "dark" personality traits (e.g., De Vries & van Kampen, 2010; Lee & Ashton, 2005; Lee et al., 2013; Moshagen, Zettler, & Hilbig, 2018; see also Liu, Zettler, & Hilbig, 2016; Zettler & Hilbig, 2015, for recent summaries). More generally, though, diverse research now consistently suggests that the HEXACO framework provides a valid representation of personality structure and a useful empirical and theoretical account of individual variation (Ashton et al., 2014).

Arguably, one reason for the steadily increasing number of studies on the HEXACO dimensions is that the HEXACO–PI–R has been translated into more and more languages in recent years, currently summing up to a total of 24 different language versions (all freely available at www.hexaco.org for academic use). Although all language versions have been carefully designed so as to provide equivalent measures of the HEXACO dimensions, a corresponding empirical test of *measurement invariance* (Byrne, Shavelson, & Muthén, 1989) is still missing. By definition, an inventory is said to be measurement invariant if it measures a construct in the same way across different groups (e.g., languages, cultures, gender, etc.) such that individuals with the same level on a given trait provide the same responses to the inventory, irrespective of the group. Transferred to languages, measurement invariance thus implies that an inventory yields the same trait scores for individuals with the same trait levels, irrespective of the language in which the inventory has been presented.

In general, measurement invariance is a vital prerequisite for the comparability of an inventory across different groups: Unless measurement invariance has been shown to hold, it is unclear (a) whether indicators reflect the same underlying construct across groups—and thus have the same meaning—and (b) whether means can be validly compared across languages and thus whether corresponding mean differences can be substantively interpreted (but see McCrae, 2015). Measurement invariance of different language versions of an inventory is thus said to constitute an essential precondition for valid cross-language and cross-country[1] comparisons (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014; Rutkowski & Svetina, 2014; Vandenberg & Lance, 2000). Correspondingly, tests of cross-language measurement invariance of trait questionnaires have gained considerable attention in prior research (e.g., Alessandri et al., 2014; Alessandri, Vecchione, Eisenberg, & Łaguna, 2015; Bowden, Saklofske, van de Vijver, Sudarshan, & Eysenck, 2016; Church et al., 2011; Dimitrova et al., 2016; McGrath, 2016; Schlotz, Yim, Zoccola, Jansen, & Schulz, 2011; Thalmayer & Saucier, 2014; Żemojtel-Piotrowska et al., 2017). With regard to the Big Five, for instance, this evidence suggests a substantial degree of cross-cultural noninvariance for well-established measures such as the NEO PI–R (Church et al., 2011) or the Big Five Questionnaire (Alessandri et al., 2014). Similar results have also been reported for six-factorial

[1]The distinction between cross-language versus cross-country measurement invariance is commonly confounded in corresponding tests given that different language versions of an inventory are usually compared across native-speaking samples in different countries. Indeed, the same also applies to the test of measurement invariance provided here. Thus, although we primarily refer to cross-language invariance in what follows, one might likewise interpret the results in terms of cross-country invariance.

alternatives such as the Questionnaire Big Six (Thalmayer & Saucier, 2014).

For measurement of the HEXACO dimensions, however, evidence on measurement invariance across languages is at best rudimentary: To date, only a single study (Ion et al., 2017) has tested measurement invariance of the 200-item HEXACO–PI–R across several languages (i.e., Hindi, Indonesian, Arabic, Romanian, and Thai; with sample sizes ranging between 210 and 482 across groups), providing support for configural invariance, but not for metric and scalar invariance (for descriptions of these terms, see the next section). However, Ion et al. (2017) did not include the more commonly applied language versions of the HEXACO–PI–R in their comparison—most prominently English, but also languages spoken in various other European or Asian countries. In addition, and even more important, in several of the included samples, the HEXACO facet scale reliabilities and intercorrelations were far smaller than in most other samples examined to date, to such an extent as it might reflect a lack of familiarity among the respondents of those samples with self-report questionnaires. In essence, this shows that evidence on cross-language measurement invariance of the HEXACO–PI–R is clearly insufficient, highlighting the necessity of corresponding tests. In this work, we seek to address this issue and to thereby provide evidence on whether cross-country and cross-cultural comparisons based on the HEXACO inventory are indeed readily interpretable or rather confounded by differences in measurement across language versions.

Testing measurement invariance

Measurement invariance is typically tested by comparing a sequence of increasingly restricted factor models (Meredith, 1993; Van de Schoot, Lugtig, & Hox, 2012; Widaman & Reise, 1997). This sequence traditionally starts with a test for *configural invariance,* which implies estimating a unique model for each group without imposing any invariance constraints. Configural invariance holds if the same model is valid in each group, meaning that the group-specific models involve the same number of latent variables and the same pattern of indicator–factor relationships (i.e., the same indicators load on the same factors across groups but the strength of loadings can differ). Next, a test of *metric invariance* (also called *weak measurement invariance*) follows that provides information on whether factor loadings are invariant across groups; that is, whether indicators show similar relations to the latent factors. To this end, factor loadings are restricted to be equal and the model is compared with the configural model. By implication, if metric invariance holds, the latent construct has the same meaning across groups because it is defined by the same indicators to the same extent. Finally, in addition to equal factor loadings, *scalar invariance* (also called *strong measurement invariance*) requires the indicator intercepts to be invariant. Correspondingly, to test scalar invariance, indicator intercepts are restricted to be equal and the resulting model is compared to the metric invariance model. If scalar invariance holds, observed mean differences (in indicators) between groups can be attributed to corresponding differences in the latent construct.[2]

To estimate and compare these increasingly restricted models—and to thus evaluate measurement invariance of an inventory across certain groups—prior research has widely relied on multigroup confirmatory factor analysis (Jöreskog, 1971). However, particularly for personality inventories, the suitability of confirmatory factor analysis (CFA)—and corresponding tests of measurement invariance—has been called into doubt: “In actual analyses of personality data … , structures that are known to be reliable showed poor fits when evaluated by CFA techniques. We believe this points to serious problems with CFA itself when used to examine personality structure” (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996, p. 568; see, e.g., Church & Burke, 1994, for similar reasoning). More specifically, because trait indicators will likely have secondary loadings on factors other than their primary factor, it is too restrictive to require each indicator to load on a single factor only, as is naturally implied by CFA. As a remedy, exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) by default allows for cross-loadings of indicators on other factors than their primary factor, thereby ensuring that potential cross-loadings no longer contribute to model misspecification. As such, ESEM provides a valuable alternative to CFA whenever cross-loadings of indicators are to be expected.[3] Correspondingly, ESEM has already been shown to produce considerably better model fit and more accurate (i.e., less positively biased) factor correlations than CFA when applied to personality data as, for instance, based on the NEO PI–R (Marsh et al., 2010). Likewise, multigroup ESEM (Marsh et al., 2009) has been established as a useful alternative to multigroup CFA for testing measurement invariance of omnibus personality inventories (e.g., Ion et al., 2017; Marsh et al., 2010; Marsh, Nagengast, & Morin, 2013; Marsh, Vallerand, et al., 2013; Moshagen, Hilbig, & Zettler, 2014). In this work, we therefore rely on multigroup ESEM as the primary approach to test cross-language measurement invariance of the HEXACO–PI–R.

However, irrespective of the specific analytic approach used, it should be noted that complete measurement invariance (especially scalar invariance) is hardly ever achieved, especially when measurement invariance is tested across a large number of groups (e.g., Davidov et al., 2014; McGrath, 2016; Rutkowski & Svetina, 2014; Thalmayer & Saucier, 2014; Zercher, Schmidt, Cieciuch, & Davidov, 2015).

[2]In addition to this sequence of nested models, further restrictions can be imposed and tested (see Marsh et al., 2009, for a taxonomy of invariance models). However, given that our primary focus is on testing whether indicator and factor means are comparable across languages, we confine our analysis to the previously mentioned sequence, with the scalar invariance model being the most restrictive (e.g., Rutkowski & Svetina, 2014; Thompson & Green, 2013).

[3]Note that it is also possible to specify cross-loadings in CFA. However, especially when multiple cross-loadings are to be expected or when there is no strong a priori theory about which cross-loadings to expect—as is typically the case with omnibus personality inventories—ESEM provides a useful alternative to CFA models.

Whenever loadings or intercepts turn out to be noninvariant, the typical procedure is to gradually relax equality constraints based on modification indexes until the models no longer differ significantly, thus establishing *partial measurement invariance* (Byrne et al., 1989). However, "particularly in large-scale studies, the stepwise selection process of relaxing invariance constraints one parameter at a time is highly cumbersome, idiosyncratic, and likely to capitalize on chance, so that the final solution is not replicable" (Marsh et al., 2018, p. 525; see also MacCallum, Roznowski, & Necowitz, 1992). Indeed, Byrne et al. (1989) themselves warned against indiscriminate post-hoc adjustment of model parameters and pointed to the necessity of "exercising sound judgment in the implementation of these procedures" (p. 465).

To overcome these inherent limitations associated with establishing partial measurement invariance, *multiple group factor analysis alignment* (Asparouhov & Muthén, 2014; Marsh et al., 2018) has recently been proposed as an alternative approach to study measurement invariance in situations in which full invariance is not achieved (as is typically the case in large-scale studies). In general, the alignment method does not assume measurement invariance; rather, it seeks an optimal pattern of measurement invariance that keeps noninvariance to a minimum, implying a few large, noninvariant model parameters and many approximately invariant parameters. Correspondingly, the alignment model does not impose any restrictions on the model parameters but is based on the configural model. A key advantage of the alignment method is that it allows determining which parameters are approximately invariant and which are not. Specifically, for each parameter (i.e., intercepts and factor loadings), "the largest invariant set of groups is found where for each group in the invariant set of groups the measurement parameter in that group is not statistically significant from the average value for that parameter across all groups in the invariant set" (Asparouhov & Muthén, 2014, p. 499). This is done using an iterative algorithm based on multiple pairwise comparison of parameters that specifies $p > .01$ as criterion to create a "starting set" of two approximately invariant groups to which additional groups are added that are sufficiently similar, meaning that the comparison between the average parameter value in the starting set is not different from the parameter value of the potentially to-be-added group at $p = .001$ (for further details, see Asparouhov & Muthén, 2014). As such, the alignment method provides information on the relative contribution of each parameter to measurement invariance and thus on the degree of noninvariance of each specific parameter. However, note that the alignment method has to date only been implemented within multigroup CFA but not within multigroup ESEM. Nonetheless, given the striking advantages as compared to post-hoc parameter adjustments (i.e., partial measurement invariance), we considered the (CFA-based) alignment method a valuable complement to our primary (and more traditional) analysis based on multigroup ESEM. However, given that the results from the alignment method can only be meaningfully interpreted in a CFA context, we also provide results from corresponding multigroup CFA.

This study

This study aimed to provide a large-scale test of cross-language measurement invariance of the 100-item version of the HEXACO–PI–R (HEXACO–100; Lee & Ashton, 2018). As sketched earlier, cross-language measurement invariance is often considered a vital prerequisite for the comparability of scale scores across countries. However, despite this importance, corresponding tests for the HEXACO–PI–R have not been conducted within the languages in which most HEXACO-based research is and has been undertaken. In this study, we aimed at closing this gap. Specifically, we compared responses on the HEXACO-100 across 16 languages, including English as well as languages from various European and Asian countries. By this means, our aim was to investigate whether the HEXACO scores are indeed measured equivalently across a variety of languages that are commonly used in personality research. Given that our goal was thus primarily exploratory in nature, we did not preregister the study.

Method

Materials

The HEXACO–100 (Lee & Ashton, 2018) is a half-length version of the HEXACO–PI–R (Lee & Ashton, 2004, 2006) including 16 items to measure each of the six HEXACO dimensions (resulting in 96 items in total).[4] Each dimension, in turn, includes four facets that are assessed by four items each. All items are answered on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Half of the items overall (i.e., 50 out of 100) are reverse-scored. Respondents' scores are computed as the average across all responses belonging to a facet or dimension, respectively, after recoding the reverse-scored items. In this study, we used 16 language versions of the HEXACO–100, namely Chinese, Croatian, Czech, Dutch, English, German, Hungarian, Italian, Japanese, Korean, Polish, Russian, Serbian, Spanish, Swedish, and Turkish. All these versions were translated from the original English version using common back-translation procedures and were finally approved by the authors of the original (English) HEXACO–PI–R (K. Lee & M. Ashton).

To test cross-language measurement invariance of the HEXACO–100, we relied on the 24 facet scores. In comparison to item-level analyses, facet-level analyses are associated with lower model complexity (due to the considerably

[4]In addition, the inventory includes four items to measure altruism as an interstitial facet, thus bringing the total number of items to 100. The altruism facet is an interstitial facet because it is expected to divide its loadings across three factors, namely Honesty-Humility, Emotionality, and Agreeableness—which are interpreted as representing different aspects of reciprocal or kin altruistic tendencies, respectively (Ashton & Lee, 2007). Therefore, we refrained from consideration of the altruism facet but focused on the six HEXACO dimensions (and the respective 24 facets) only.

Table 1. Descriptive statistics of the 16 samples included in the analysis.

Language	Country	N	Female (in %)	Age M	Age SD
Chinese	Taiwan	717	49.5	21.6	2.7
Croatian	Croatia	877	51.1	20.4	2.7
Czech	Czech Republic	2,959	75.5	22.8	4.1
Dutch	Netherlands	3,205	59.3	37.6	16.6
English	Canada	2,851	64.9	20.9	3.9
German	Germany	9,491	76.5	32.4	9.4
Hungarian	Hungary	952	64.6	32.4	13.7
Italian	Italy	940	53.9	37.0	14.3
Japanese	Japan	1,070	52.3	19.0	1.3
Korean	Korea	341	59.9	22.1	2.5
Polish	Poland	227	78.4	32.1	10.6
Russian	Russia	767	57.7	31.2	13.9
Serbian	Serbia	2,896	55.1	28.6	11.2
Spanish	Spain	1,129	59.3	38.7	14.0
Swedish	Sweden	471	64.5	27.2	8.5
Turkish	Turkey	1,591	54.9	30.9	11.5

smaller number of parameters), thus facilitating interpretability of the data. Also, given that the facet scores represent the lowest level of any trait analysis, it is particularly important to know whether facet scores are comparable across different languages. Correspondingly, facet-level analyses have been commonly used in prior large-scale measurement invariance tests of omnibus personality inventories (e.g., Church et al., 2011; Ion et al., 2017; Labouvie & Ruetsch, 1995). However, we also report corresponding item-level analyses based on multigroup CFA. Importantly, these additional analyses overall yielded highly similar results as the facet-level analyses (see post-hoc analyses below).

Samples

A total of 16 samples from different countries were included in this study[5]. Table 1 provides an overview of the sample characteristics for the specific subgroups. Overall, the sample consisted of $N = 30{,}484$ participants (65.6% female), between 13 and 88 years old ($M = 29.7$, $SD = 11.9$). Note that the samples were collected independently by different author teams and originally sought for other research purposes involving psychometric analyses (all data were anonymous). Therefore, the samples also differed in composition, including student as well as community samples with different gender ratio, age, and educational background (Table 1). A list of publications using (parts of) the data reported on herein is provided in the online supplemental materials on the Open Science Framework (OSF; https://osf.io/bwtnr).

Data analysis

Data were analyzed using M*plus* version 7.3 (Muthén & Muthén, 2012). For all models, we relied on the robust maximum likelihood (MLR) estimator to ensure that standard

Table 2. Means, standard deviations, and scale-based intercorrelations between the HEXACO dimensions.

Variable	M	SD	1	2	3	4	5	6
1. Honesty-Humility	3.48	0.60	*.81*					
2. Emotionality	3.30	0.56	.04	*.80*				
3. Extraversion	3.40	0.60	−.01	−.14	*.85*			
4. Agreeableness	2.92	0.55	.26	−.15	.12	*.80*		
5. Conscientiousness	3.47	0.55	.16	.01	.19	.02	*.81*	
6. Openness to Experience	3.43	0.60	.10	−.08	.16	.02	.10	*.80*

Note. Overall sample, $N = 30{,}484$. Alpha reliabilities (Cronbach's α) are reported on the diagonal (in italics).

errors and tests of model fit are robust against nonnormality and nonindependence of the data. In the multigroup models, the English language version served as the reference group—given that all other language versions of the HEXACO–PI–R have been translated from the English version (see earlier). In all ESEM analyses, we used the oblique target rotation criterion given its particular suitability for models involving multiple factors (Asparouhov & Muthén, 2009). Therefore, target values for all facets except those that are intended to load primarily on the respective HEXACO factor were set to zero. Furthermore, for model identification purposes, in the configural and metric (ESEM and CFA) models we fixed the factor variances to 1 and the factor means to 0 across all groups; in the scalar models, factor means were freely estimated in all but the reference (English language) group (in which it was fixed to 1).[6] The data[7] and all analysis scripts for use in M*plus* are provided in the supplemental materials (https://osf.io/bwtnr).

To evaluate absolute model fit, we referred to the descriptive fit indexes root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR), as has been particularly recommended for usage in personality research (Beauducel & Wittmann, 2005). Specifically, we relied on common guidelines to imply satisfactory model fit, namely RMSEA $\leq .05$ and SRMR $\leq .06$ (Browne & Cudeck, 1992; Hu & Bentler, 1999). As such, we refrained from consideration of the χ^2 test statistic, which has been shown to be seriously inflated for models involving a large number of variables (e.g., Herzog & Boomsma, 2009; Moshagen, 2012; Rutkowski & Svetina, 2014) and for which statistical power is typically far too large (Moshagen & Erdfelder, 2016).

In turn, to evaluate comparative model fit and thus measurement invariance, we referred to the differences between models in the comparative fit index (ΔCFI) and the ΔRMSEA. Prior research has suggested that ΔCFI $\leq .01$ and ΔRMSEA $\leq .015$ implies that two models are sufficiently similar (Chen, 2007; Cheung & Rensvold, 2002). However, we will place particular emphasis on ΔRMSEA given that

[5]The Spanish sample consists of two subsamples in which different translations of the HEXACO–PI–R were used. However, given that group-based ESEM analysis provided support for metric and scalar invariance across the two subsamples (for results, see Table S1 in the supplemental materials on the OSF), we merged them for the following analyses.

[6]In addition, we ran the ESEM scalar invariance model with factor means being fixed to zero and to be equal across groups (note that in both these models, the factor means in the reference group were fixed to be zero by default). Corresponding model fit statistics are available in Table S2 in the supplemental materials (https://osf.io/bwtnr).

[7]Note that the data set provided on the OSF does not contain the English (Canadian), Hungarian, and Russian data given that the conditions of participant consent in these data sets were not compatible with the posting of the data in an online repository. The data are, however, available on request from the first author.

this fit index incorporates a specific penalty for low parsimony of models, a criterion that is arguably particularly important when evaluating large-scale models involving a huge number of parameters (see, e.g., Marsh, 2007; Marsh et al., 2009). Moreover, note that especially for large-scale measurement invariance tests, the criteria just mentioned appear to be comparably strict (Rutkowski & Svetina, 2014) and that other researchers have come to different conclusions regarding the appropriateness and cutoff criteria of alternative fit measures to evaluate comparative model fit (e.g., Meade, Johnson, & Braddy, 2008). As such, it is generally still under debate which criteria to use best in which context. We therefore also provide information on differences in McDonald's noncentrality index (ΔNCI; McDonald, 1989) to allow readers a fully transparent overview. Cheung and Rensvold (2002) proposed $\Delta NCI \leq .02$ as implying two models to be sufficiently similar to conclude equivalence.

Results

Descriptive statistics, alpha reliabilities, and scale-based (raw) intercorrelations between the HEXACO dimensions in the overall sample are reported in Table 2 (for corresponding statistics separated per language version, see the analyses in the supplemental materials; https://osf.io/bwtnr)[8]. As is apparent, scale-based intercorrelations were generally low, with a maximum of $|r| = .26$ (between Honesty-Humility and Agreeableness), more than half of the correlations $|r| \leq .10$, and a corresponding mean $|r| = .10$. A total group ESEM model across language versions provided good fit to the data (CFI = 0.942, RMSEA = 0.044, SRMR = 0.019), and all facets showed the highest loadings on their corresponding (oblique) personality factor (see Table 3; for fit indexes separated per language version, see Table S3 in the OSF supplemental materials). An equivalent total group CFA model yielded worse fit, CFI = 0.705, RMSEA = 0.078, SRMR = 0.078, in line with prior research on omnibus personality inventories (e.g., Marsh et al., 2010).

As described earlier, to test measurement invariance of the HEXACO–100 across language versions, we relied on multigroup ESEM as our primary approach. Fit statistics of all models estimated are summarized in Table 4. As is apparent, the configural invariance model (Model 1 in Table 4) fitted the data well. This implies that the same model is valid in each group; that is, that the same facets load on the same factors across language versions. Next, we estimated the metric invariance model by restricting the factor loadings to be equal across groups. Corresponding to the higher parsimony as compared to the configural model, model fit slightly decreased, although the model still provided satisfactory fit overall (Model 2 in Table 4). Comparing the descriptive fit statistics indicated noteworthy differences for ΔCFI, suggesting that the metric invariance model might indeed not hold for some language versions. In contrast, RMSEA slightly decreased in the metric invariance model, which indicates that the metric invariance model would actually be preferred to the configural invariance model if model parsimony is taken into account (see Marsh, 2007).

To get a deeper understanding of these (somewhat mixed) findings regarding metric invariance, we decided to examine the factor structures of some samples from the configural invariance model, namely the Japanese, Turkish, and German samples (see Table 5; factor loadings for all language versions are provided in Table S4 in the OSF supplemental materials). In choosing these samples, we considered the χ^2 value for each language version obtained from the metric invariance model relative to that obtained from the configural invariance model: The Japanese and Turkish samples showed the largest relative increases in χ^2 values (hence representing the two most "unique" factor loading solutions), and the German sample the smallest increase (hence representing the most "universal" factor loading solution). As is apparent in Table 5, the factor loading solutions from these language versions showed no appreciable differences (see also Table S4 in the OSF supplemental materials). In turn, computing factor congruence coefficients among the three language versions showed that the lowest congruence coefficient was .92 (for Emotionality between the Japanese and Turkish versions), and most congruence coefficients exceeded .95. Thus, despite the overall difference in ΔCFI between the configural and metric invariance model, these results strongly suggest that the factor loadings are sufficiently invariant across language versions.

Finally, a model requiring equality of loadings and intercepts—and thus scalar invariance—did not fit the data

Table 3. Means, standard deviations, and standardized factor loadings from exploratory structural equation modeling of the 24 HEXACO facets in the total group model.

			Factor loadings					
HEXACO facets	*M*	*SD*	HH	EM	EX	AG	CO	OP
HH sincerity	3.44	0.83	**.50**	−.10	.02	−.13	.08	.04
HH fairness	3.66	0.96	**.48**	.11	.04	.04	.19	−.05
HH greed-avoidance	3.22	0.94	**.70**	−.07	−.09	.00	−.11	.07
HH modesty	3.59	0.75	**.56**	.09	−.07	.11	−.10	−.07
EM fearfulness	2.91	0.81	−.06	**.52**	−.17	.05	.02	−.14
EM anxiety	3.50	0.78	−.07	**.56**	−.31	−.14	.14	.07
EM dependence	3.24	0.84	−.04	**.62**	.18	−.01	−.08	−.02
EM sentimentality	3.57	0.77	.17	**.66**	.17	.00	.02	.09
EX social self-esteem	3.66	0.74	.00	−.14	**.64**	.07	.17	−.03
EX social boldness	3.04	0.85	−.04	−.11	**.58**	−.14	.01	.17
EX sociability	3.46	0.77	−.10	.26	**.63**	.07	−.07	−.02
EX liveliness	3.45	0.82	.04	−.08	**.76**	.06	.04	−.04
AG forgivingness	2.58	0.79	.05	−.05	.12	**.50**	−.05	.06
AG gentleness	3.19	0.71	.08	.10	.01	**.59**	−.04	.00
AG flexibility	2.79	0.70	.02	.07	.00	**.60**	−.03	−.05
AG patience	3.11	0.81	−.07	−.15	−.04	**.76**	.11	.10
CO organization	3.40	0.85	.04	.04	.10	.00	**.62**	−.17
CO diligence	3.74	0.71	.06	.02	.27	−.10	**.57**	.13
CO perfectionism	3.49	0.74	−.05	.14	−.11	−.07	**.65**	.08
CO prudence	3.24	0.73	−.01	−.12	−.12	.19	**.62**	−.02
OP aesthetic appreciation	3.47	0.88	.10	.14	−.08	.09	.03	**.71**
OP inquisitiveness	3.33	0.87	.05	−.15	−.06	.03	.12	**.49**
OP creativity	3.56	0.86	−.01	.06	.13	−.01	.01	**.65**
OP unconventionality	3.37	0.70	−.12	−.04	−.01	.00	−.13	**.63**

Note. HH = Honesty-Humility; EM = Emotionality; EX = Extraversion; AG = Agreeableness; CO = Conscientiousness; OP = Openness to Experience. Highest loading per facet is shown in bold.

[8]Note that the various samples were, in different respects, not representative of the national populations from which they were drawn. Therefore, differences in mean scores across our various samples do not necessarily imply national-level differences.

Table 4. Model fit statistics resulting from multigroup analyses testing measurement invariance across language versions of the HEXACO–100.

Model	χ^2	*df*	CFI	NCI	RMSEA	SRMR	Model comparison	χ^2_{diff}	Δdf	ΔCFI	ΔNCI	ΔRMSEA
Multigroup exploratory structural equation modeling												
1. Configural	12,535.73	2,352	0.937	0.846	0.048	0.022						
2. Metric	18,755.69	3,972	0.909	0.785	0.044	0.044	2 vs. 1	5,277.04	1,620	0.028	0.061	−0.004
3. Scalar	40,714.04	4,242	0.774	0.550	0.067	0.075	3 vs. 2	21,649.24	270	0.135	0.235	0.023
Multigroup confirmatory factor analysis												
4. Configural	51,835.86	3,792	0.703	0.455	0.082	0.081						
5. Metric	52,924.97	4,062	0.698	0.449	0.079	0.086	5 vs. 4	758.22	270	0.005	0.006	−0.003
6. Scalar	77,142.71	4,332	0.550	0.303	0.094	0.111	6 vs. 5	23,531.76	270	0.148	0.146	0.015

Note. $N = 30{,}484$. CFI = comparative fit index; NCI = McDonald's noncentrality index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; χ^2_{diff} = scaled χ^2 difference test; Δdf = difference in the degrees of freedom; ΔCFI = difference in the comparative fit index; ΔNCI = difference in McDonald's noncentrality index; ΔRMSEA = difference in the root mean square error of approximation.

Table 5. Factor loadings from Japanese, Turkish, and German samples (configural invariance model; multigroup exploratory structural equation modeling with oblique target rotation).

	HH			EM			EX			AG			CO			OP		
	GE	JP	TR	GE	JP	TR	GE	JP	TR	GE	JP	TR	GE	JP	TR	GE	JP	TR
HH sincerity	**.59**	**.44**	**.58**	−.12	−.25	.03	−.03	.02	.08	−.13	−.13	−.12	.01	.05	.09	.05	.11	.04
HH fairness	**.47**	**.44**	**.46**	.07	.15	.16	.08	.07	.08	.03	.02	.04	.21	.19	.22	−.06	.00	−.06
HH greed-avoidance	**.66**	**.53**	**.62**	−.04	−.09	−.08	−.08	−.04	−.11	.01	.03	.06	−.11	.04	−.03	.06	.06	.06
HH modesty	**.62**	**.64**	**.61**	.15	.03	−.06	−.02	−.13	−.17	.09	.07	.03	−.05	−.11	−.18	−.06	−.15	−.01
EM fearfulness	−.05	−.07	−.10	**.43**	**.50**	**.55**	−.17	−.10	−.13	−.01	−.08	.09	.03	.03	.08	−.13	−.15	−.11
EM anxiety	−.05	−.04	−.08	**.57**	**.58**	**.48**	−.34	−.29	−.27	−.13	−.10	−.18	.12	.14	.06	.04	.10	.07
EM dependence	−.02	−.09	−.06	**.60**	**.55**	**.56**	.21	.15	.11	−.03	−.04	−.01	−.09	−.08	−.15	−.04	−.07	.01
EM sentimentality	.12	.10	.25	**.67**	**.57**	**.59**	.15	.24	.15	.01	.03	.06	−.01	−.03	.06	.10	.07	.03
EX social self-esteem	.04	−.20	−.12	−.16	−.14	−.08	**.64**	**.55**	**.64**	.10	.10	.00	.12	.14	.17	−.03	−.07	−.02
EX social boldness	−.04	−.05	−.03	−.12	−.12	−.15	**.58**	**.61**	**.55**	−.13	−.12	−.11	.04	.11	.00	.15	.23	.20
EX sociability	−.09	.08	−.06	.27	.29	.27	**.66**	**.69**	**.54**	.10	.10	.05	−.04	.01	−.12	.00	−.13	.01
EX liveliness	.06	−.01	.05	−.07	−.05	−.10	**.77**	**.80**	**.69**	.02	.03	.06	.02	−.07	.01	−.02	−.05	−.05
AG forgivingness	.02	.02	.06	−.15	−.06	.01	.10	.14	.02	**.56**	**.58**	**.56**	−.11	−.03	−.11	.08	.11	.08
AG gentleness	.05	.05	.10	.15	.13	.09	−.03	−.01	−.01	**.65**	**.65**	**.53**	−.07	−.02	.03	.01	.02	−.09
AG flexibility	.01	.09	.04	.07	−.09	.06	.03	.00	.03	**.62**	**.48**	**.49**	.03	−.11	.01	−.06	−.13	.02
AG patience	−.02	−.08	−.13	−.16	−.10	−.12	−.03	−.06	−.02	**.73**	**.80**	**.82**	.12	.12	.04	.06	.09	.08
CO organization	.00	.03	.08	.01	−.10	.04	.11	.20	.04	−.04	−.04	.08	**.56**	**.63**	**.62**	−.14	−.15	−.17
CO diligence	.06	.08	.05	.04	.08	−.01	.28	.29	.22	−.09	.00	−.08	**.58**	**.42**	**.56**	.12	.26	.11
CO perfectionism	−.04	.02	−.09	.13	.15	.09	−.14	−.09	−.11	−.06	−.08	−.10	**.58**	**.66**	**.65**	.11	.07	.12
CO prudence	.00	−.02	.04	−.12	−.06	−.10	−.11	−.18	−.05	.18	.12	.13	**.64**	**.63**	**.53**	−.06	−.19	.03
OP aesthetic appreciation	.08	.09	.06	.10	.05	.18	−.06	−.06	−.07	.11	.07	.14	.04	−.01	.07	**.67**	**.63**	**.67**
OP inquisitiveness	.07	−.03	.07	−.15	−.05	−.15	−.04	−.13	−.05	.04	.05	.04	.09	.15	.17	**.47**	**.45**	**.51**
OP creativity	−.05	−.09	−.05	.10	−.07	.06	.10	.14	.16	.02	−.02	−.02	.04	−.07	−.02	**.62**	**.65**	**.58**
OP unconventionality	−.07	.01	−.04	−.06	.04	−.10	.00	−.06	.03	−.08	.06	−.04	−.12	−.11	−.13	**.63**	**.59**	**.61**

Note. HH = Honesty-Humility; EM = Emotionality; EX = Extraversion; AG = Agreeableness; CO = Conscientiousness; OP = Openness to Experience; GE = Germany; JP = Japan; TR = Turkey. Absolute factor loadings greater than .40 are shown in bold.

adequately (Model 3 in Table 4). Correspondingly, model fit substantially decreased as compared to the metric invariance model, thus failing to provide support for scalar invariance.

Given the lack of support for scalar invariance using multigroup ESEM, we further used the alignment method (see details earlier) to determine the degree of noninvariance of each specific model parameter. To this end, we first applied multigroup CFA, given that the alignment method is to date only available within the CFA framework. As summarized in Table 4, the configural model (Model 4) imposing no restrictions on factor loadings or intercepts yielded acceptable fit to the data, although fit was generally lower as compared to the corresponding ESEM model. In turn, restricting the factor loadings to be invariant across groups did not lead to a significant reduction in model fit (Model 5 in Table 4), thus once more supporting metric invariance across language versions. However, model fit substantially decreased once additionally restricting the factor intercepts to be equal across groups (Model 6 in Table 4), thus demonstrating a lack of scalar invariance—consistent with the results from multigroup ESEM.

Applying the alignment method[9] reflected this pattern of invariance across factor loadings, but noninvariance across facet intercepts: As is apparent in Table 6, the percentage of invariant loadings was generally high across facets per HEXACO factor, ranging from 78.1% for Emotionality to 96.9% for Agreeableness.[10] The only notable exception to this pattern was apparent for the anxiety facet of Emotionality. For this facet, factor loadings were only invariant across 6 of the 16 language versions, whereas for all other facets, factor loadings were invariant across at least 12 language versions. For facet intercepts, in turn, the percentage of invariant parameters was considerably lower, ranging from 46.9% for Agreeableness to 79.7% for

[9]To estimate the alignment model, we used the "free" optimization option as implemented in M*plus,* in line with recommendations (Asparouhov & Muthén, 2014). Using this option, the factor means are freely estimated.

[10]Percentages of invariant parameters for HEXACO factors represent the total number of approximate invariant groups across facets per factor divided by the total number of groups across facets (i.e., 4 facets *16 groups = 64). In turn, percentages of invariant parameters for HEXACO facets represent the total number of approximate invariant groups divided by the total number of groups (i.e., 16).

Table 6. Percentage of invariant parameters based on the alignment method.

HEXACO factors and facets	Parameter invariance status (in %)[a]	
	Loadings	Intercepts
Honesty-Humility	**92.2**	**57.8**
Sincerity	93.8	37.5
Fairness	81.3	100.0
Greed-avoidance	100.0	56.3
Modesty	93.8	37.5
Emotionality	**78.1**	**54.7**
Fearfulness	100.0	18.8
Anxiety	37.5	68.8
Dependence	93.8	68.8
Sentimentality	81.3	62.5
Extraversion	**85.9**	**57.8**
Social self-esteem	81.3	43.8
Social boldness	87.5	68.8
Sociability	75.0	50.0
Liveliness	100.0	68.8
Agreeableness	**96.9**	**46.9**
Forgivingness	100.0	31.3
Gentleness	93.8	75.0
Flexibility	100.0	37.5
Patience	93.8	43.8
Conscientiousness	**95.3**	**79.7**
Organization	100.0	68.8
Diligence	93.8	87.5
Perfectionism	93.8	81.3
Prudence	93.8	81.3
Openness	**81.3**	**73.4**
Aesthetic appreciation	75.0	81.3
Inquisitiveness	81.3	75.0
Creativity	87.5	87.5
Unconventionality	81.3	50.0

Note. Values for HEXACO factors (shown in bold) represent the total number of approximate invariant groups across corresponding facets divided by total number of groups across facets (i.e., 4 * 16 = 64).

[a]Total number of approximate invariant groups divided by total number of groups (i.e., 16).

Conscientiousness. Strikingly, the degree of invariance associated with facet intercepts varied substantially, even within one and the same factor: For instance, whereas the gentleness facet of Agreeableness yielded a relatively high degree of invariance of facet intercepts (being invariant across 12 of the 16 groups), the forgivingness facet of Agreeableness yielded a comparably low degree of invariance of facet intercepts (being invariant across only five groups). Likewise, whereas intercepts of the fairness facet of Honesty-Humility were invariant across all 16 groups, intercepts of the sincerity and modesty facets were only invariant across six groups each. Overall, this shows that there is some variation in the degree of noninvariance across HEXACO facets and factors. However, it also demonstrates that—despite the overall lack of scalar invariance—the intercepts of some facets are still associated with a fairly high degree of invariance across language versions.

Post-hoc analyses: Item-level analyses

Although we originally planned to exclusively examine the measurement invariance of the HEXACO–PI–R at the facet level, in this section we also report results from a series of item-level multigroup CFAs for the sake of completeness. The analysis including all six HEXACO personality factors in one model encountered some convergence problems, although generally replicating the facet-level results (i.e., support for configural and metric invariance, but no support for scalar invariance; see Table S6 in the OSF supplemental materials). We therefore decided to additionally conduct a multigroup CFA for each personality factor separately (we thank an anonymous reviewer for this suggestion). As such, each model included four oblique factors that were defined by four items each. Table 7 summarizes the results of these item-level analyses per HEXACO factor. As is apparent, model fit statistics of the configural models were satisfactory to good for all factors. Likewise, the model comparison provided evidence for metric invariance for all factors. However, analyses again showed no support for scalar invariance, thus replicating the facet-level results as well as the item-level results for the overall model.

Discussion

The HEXACO model of personality structure and corresponding inventory, the HEXACO–PI–R (Lee & Ashton, 2004, 2006), have become well-established in psychology and beyond and are still steadily gaining increasing attention in research. Corresponding to this development, the HEXACO–PI–R has up to now been translated into 24 different languages. Although most of these language versions have, taken individually, been thoroughly validated (e.g., Babarović & Šverko, 2013; Bergh & Akrami, 2016; De Vries, Lee, & Ashton, 2008; Lee & Ashton, 2018; Međedović, Čolović, Dinić, & Smederevac, 2017; Moshagen et al., 2014; Romero, Villar, & López-Romero, 2015; Roncero, Fornés, & Belloch, 2016; Tatar, 2018; Wakabayashi, 2014; Wasti, Lee, Ashton, & Somer, 2008; Záškodná & Dostál, 2016), evidence on whether the different language versions provide equivalent measures of the six broad personality dimensions is largely missing. Strikingly, though, measurement invariance of an inventory across different groups—especially with regard to the general structure (configural invariance) and loadings (metric invariance)—is often considered a vital prerequisite for the comparability of trait scores obtained in these groups. Given this importance, we aimed at providing a large-scale test of measurement invariance of the HEXACO–PI–R across diverse languages. Specifically, we investigated whether and to what extent the 100-item version of the HEXACO–PI–R—the HEXACO–100—provides comparable measurement of the HEXACO dimensions across 16 languages spoken in European and Asian countries.

Overall, results from multigroup ESEM and multigroup CFA provided consistent support for configural and metric invariance of the HEXACO–100 across language versions. This implies that (a) the factor structure of the HEXACO dimensions is similar across languages, meaning that the same facets load on the same factors; and (b) the latent HEXACO factors have the same meaning across languages, given that the factors are described by the same facets in equal measure (i.e., equivalent factor loadings). However, analyses did not provide support for scalar invariance; that is, equivalence of facet intercepts across languages. This raises the question of whether observed differences in facet

Table 7. Model fit statistics resulting from item-level multigroup confirmatory factor analyses testing measurement invariance across language versions of the HEXACO–100 for each HEXACO factor.

Model	χ^2	*df*	CFI	NCI	RMSEA	SRMR	Model comparison	χ^2_{diff}	Δdf	ΔCFI	ΔNCI	ΔRMSEA
Honesty-Humility												
1. Configural	7,637.17	1,568	0.939	0.905	0.045	0.022						
2. Metric	9,009.77	1,748	0.927	0.888	0.047	0.044	2 vs. 1	1,204.89	180	0.012	0.018	0.002
3. Scalar	21,751.58	1,928	0.802	0.722	0.073	0.081	3 vs. 2	12,056.34	180	0.125	0.165	0.026
Emotionality												
4. Configural	11,558.64	1,568	0.890	0.849	0.058	0.051						
5. Metric	13,387.90	1,748	0.871	0.826	0.059	0.062	5 vs. 4	1,600.43	180	0.019	0.023	0.001
6. Scalar	26,506.44	1,928	0.728	0.668	0.082	0.087	6 vs. 5	12,550.72	180	0.143	0.158	0.023
Extraversion												
7. Configural	12,983.15	1,568	0.898	0.829	0.062	0.048						
8. Metric	15,195.58	1,748	0.880	0.802	0.064	0.067	8 vs. 7	1,920.57	180	0.018	0.027	0.002
9. Scalar	30,094.73	1,928	0.749	0.630	0.088	0.098	9 vs. 8	14,201.18	180	0.131	0.172	0.024
Agreeableness												
10. Configural	14,014.81	1,568	0.856	0.815	0.065	0.051						
11. Metric	15,822.48	1,748	0.837	0.794	0.065	0.061	11 vs. 10	1,554.46	180	0.019	0.021	0
12. Scalar	35,599.43	1,928	0.610	0.576	0.096	0.106	12 vs. 11	18,532.43	180	0.227	0.218	0.031
Conscientiousness												
13. Configural	14,933.04	1,568	0.841	0.803	0.067	0.054						
14. Metric	16,216.67	1,748	0.828	0.789	0.066	0.063	14 vs. 13	1,062.25	180	0.013	0.014	−0.001
15. Scalar	30,878.87	1,928	0.656	0.622	0.089	0.089	15 vs. 14	14,004.52	180	0.172	0.167	0.023
Openness												
16. Configural	10,522.70	1,568	0.884	0.863	0.055	0.044						
17. Metric	11,799.03	1,748	0.869	0.848	0.055	0.054	17 vs. 16	1,094.82	180	0.015	0.015	0
18. Scalar	26,581.19	1,928	0.679	0.667	0.082	0.083	18 vs. 17	14,381.46	180	0.190	0.181	0.027

Note. $N = 30{,}484$. CFI = comparative fit index; NCI = McDonald's noncentrality index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; χ^2_{diff} = scaled χ^2 difference test; Δdf = difference in the degrees of freedom; ΔCFI = difference in the comparative fit index; ΔNCI = difference in McDonald's noncentrality index; ΔRMSEA = difference in the root mean square error of approximation.

and factor means between countries can actually be attributed to "true" differences in the latent constructs or rather to differences in the measurement of these constructs.

Indeed, some researchers have argued that "meaningful comparisons of mean scores across cultures … *require* scalar invariance" (Church et al., 2011, p. 1069, italics added; see, e.g., Steinmetz, 2013; Van de Schoot et al., 2012, for similar reasoning). In turn, an absence of scalar invariance would imply undesirable biases in measurement and therefore prevent meaningful comparison of indicator and factor means across groups. According to such a view, however, cross-cultural mean comparisons are virtually impossible, given that scalar noninvariance in cross-cultural studies has been a rule rather than an exception (e.g., Davidov et al., 2014; McGrath, 2016; Rutkowski & Svetina, 2014; Thalmayer & Saucier, 2014; Zercher et al., 2015).

In contrast to this view, others (e.g., Davidov et al., 2014; McCrae, 2015; Vandenberg & Lance, 2000) have argued that scalar noninvariance does not necessarily prevent meaningful mean-level comparison. Interested readers can consult with Davidov et al.'s (2014) suggestions, which are useful in case of the violation of scalar invariance. Of these suggestions, the following is arguably most relevant: Measurement noninvariance itself might be a phenomenon of substantive interest. For example, McCrae (2015) pointed out that a lack of scalar invariance might be "the result of [actual] group differences in specific variance associated with the item" (p. 107). In other words, "intercept differences may not reflect biases (undesirable) but response threshold differences that might be predicted based on known group differences (desirable)" (Vandenberg & Lance, 2000, p. 38). Results from the alignment method reported in this research might provide some initial clues from which such exploration on true differences in the HEXACO facets between countries can begin. In any case, the alignment results also showed that, on average, intercepts were approximately invariant across the majority (i.e., ∼60%) of languages. Nonetheless, future research is needed to clarify the sources of limited scalar invariance as observed across the different language versions of the HEXACO–PI–R under scrutiny herein.

Regarding the invariance of factor loadings, results from the alignment method further indicated that loadings of the HEXACO facets tend to be invariant across most of the languages, with one major exception being the anxiety facet of the Emotionality factor. Specifically, this facet was invariant only for about one third of the languages. Inspection of the factor loadings of the anxiety facet resulting from CFA per language (i.e., configural model; Table S4 in the OSF supplemental materials) revealed that in some languages this facet showed very strong loadings on the Emotionality factor, and in these cases, there were also unusually strong negative correlations between Emotionality and Extraversion. This combination of results reflects the fact that the anxiety facet of the Emotionality factor typically shows a substantial negative secondary loading on the Extraversion factor; conversely, the sociability facet of the Extraversion factor typically shows a substantial positive secondary loading on the Emotionality factor (Table 3; see also Lee & Ashton, 2018). Because the alignment method can be performed only on a model assuming a perfect simple structure, the inability to allow those secondary loadings to be freely estimated might have contributed to differences between languages in the loadings of facets on these factors (especially in the loadings of the anxiety facet on Emotionality) and, in turn, in the correlation between factors. Consistent with this suggestion, when we reran a multigroup CFA in which the previously mentioned secondary loadings were allowed to be estimated, the anomalous results described earlier—very high loadings of the anxiety facet on Emotionality and strong negative correlations between Emotionality and Extraversion—disappeared (for corresponding factor intercorrelations, see Table S5 in the OSF supplemental materials). We thus suggest that results from the alignment method should be interpreted with caution when variable sets to be analyzed are not simple structured.

Another potential source of differences between language versions of the HEXACO–100 as identified by our analyses is variation in the composition of the samples included. That is, whereas some of the samples were student samples, others were community samples. As a consequence, the samples differed in distribution of male and female participants, mean and range of age, and educational background (Table 1). The degree of measurement noninvariance as implied by our results should thus be taken as an upper-bound estimate of the "true" degree of measurement noninvariance. In other words, it is likely that the differences between language versions would have been smaller if the samples were collected in the same way across countries. Related to this point, samples also showed considerable differences in size, ranging from $n = 227$ in the Polish sample to $n = 9{,}491$ in the German sample, causing some (larger) samples to receive greater weight in the measurement invariance test than other, relatively smaller samples. Importantly, however, when repeating the analyses with a random subsample of $n = 200$ per language group (i.e., total $N = 3{,}200$), results remained virtually the same (see Table S7 in the OSF supplemental materials). This suggests that our results were not biased by differences in sample sizes across language groups. Nonetheless, future research would profit from investigating measurement invariance across languages when ruling out sample differences with regard to both composition and size. This might be realized by recruiting and comparing nationally representative samples in different countries or by asking bilingual participants to fill in the HEXACO–PI–R in both their native languages.

Finally, it should be noted that we focused on facet-level measurement invariance in our primary analyses, and only reported post-hoc supplementary analyses for item-level measurement invariance. Although facet-level analyses have been commonly used in prior large-scale measurement invariance tests of omnibus personality inventories (e.g., Church et al., 2011; Ion et al., 2017; Labouvie & Ruetsch, 1995)—arguably because they are associated with lower model complexity and because facet scores typically represent the lowest level of any trait analysis—interpretability of facet-level measurement invariance tests hinges to some extent on item-level measurement invariance. That is, it is conceivable that a certain degree of noninvariance at the facet level is attributable to a certain degree of noninvariance at the item level. In other words, facet loadings and intercepts might indeed be invariant across groups, but a corresponding test of measurement invariance might nonetheless indicate some degree of noninvariance because item loadings and intercepts are noninvariant. We can therefore not rule out that our results implying a lack of scalar invariance at the facet level might be—at least to some extent—attributable to a lack of scalar invariance at the item level. That said, it is important to note that both facet- and item-level results provided consistent support for configural and metric invariance of the HEXACO–100 across languages.

Conclusion

Our large-scale test of measurement invariance of the 100-item HEXACO–PI–R suggests that this inventory provides largely comparable measurement of the six broad personality dimensions across languages. Although facet intercepts showed a substantial degree of noninvariance (i.e., a lack of scalar invariance), the factor structure of the HEXACO dimensions strongly converged across the 16 language versions under scrutiny (i.e., configural and metric invariance). We thus conclude that findings on the HEXACO dimensions from different language versions of the HEXACO–PI–R can be interpreted in much the same way. Nonetheless, researchers aiming at direct cross-country comparisons should be careful when interpreting mean level differences for some HEXACO facets and factors, respectively.

Acknowledgments

This article has earned the Center for Open science badges for Open Data. The data and materials are openly accessible at https://osf.io/bwtnr.

ORCID

Isabel Thielmann http://orcid.org/0000-0002-9071-5709
Nazar Akrami http://orcid.org/0000-0002-9641-6275
Amparo Belloch http://orcid.org/0000-0002-4280-9946
Antonio Chirumbolo http://orcid.org/0000-0002-4274-2489
Petar Čolović http://orcid.org/0000-0003-1212-3131
Reinout E. de Vries http://orcid.org/0000-0002-4252-5839
Augusto Gnisci http://orcid.org/0000-0003-0429-3405
Luigi Leone http://orcid.org/0000-0003-3397-8016
Marco Perugini http://orcid.org/0000-0002-4864-6623
Estrella Romero http://orcid.org/0000-0002-9239-2544
Ida Sergi http://orcid.org/0000-0001-8073-1150
Snežana Smederevac http://orcid.org/0000-0002-3780-0576
Arkun Tatar http://orcid.org/0000-0002-2369-9040
Ingo Zettler http://orcid.org/0000-0001-6140-7160
Kibeom Lee http://orcid.org/0000-0003-2775-5596

References

Alessandri, G., Vecchione, M., Donnellan, B. M., Eisenberg, N., Caprara, G. V., & Cieciuch, J. (2014). On the cross-cultural replicability of the resilient, undercontrolled, and overcontrolled personality types. *Journal of Personality*, *82*(4), 340–353. doi:10.1111/jopy.12065

Alessandri, G., Vecchione, M., Eisenberg, N., & Łaguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem

Scale. *Psychological Assessment*, *27*(2), 621–635. doi:10.1037/pas0000073

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. doi:10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The prediction of honesty-humility-related criteria by the HEXACO and five-factor models of personality. *Journal of Research in Personality*, *42*(5), 1216–1228. doi:10.1016/j.jrp.2008.03.006

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139–152. doi:10.1177/1088868314523838

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., … De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. doi:10.1037/0022-3514.86.2.356

Asparouhov, T., & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*(3), 397–438. doi:10.1080/10705510903008204

Asparouhov, T., & Muthén, B. O. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling*, *21*(4), 495–508. doi:10.1080/10705511.2014.919210

Babarović, T., & Šverko, I. (2013). The HEXACO personality domains in the Croatian sample. *Drustvena Istrazivanja*, *22*(3), 397–411. doi:10.5559/di.22.3.01

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*(1), 41–75. doi:10.1207/s15328007sem1201_3

Bergh, R., & Akrami, N. (2016). Are non-agreeable individuals prejudiced? Comparing different conceptualizations of agreeableness. *Personality and Individual Differences*, *101*, 153–159. doi:10.1016/j.paid.2016.05.052

Bowden, S. C., Saklofske, D. H., van de Vijver, F. J. R., Sudarshan, N. J., & Eysenck, S. B. G. (2016). Cross-cultural measurement invariance of the Eysenck Personality Questionnaire across 33 countries. *Personality and Individual Differences*, *103*, 53–60. doi:10.1016/j.paid.2016.04.028

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. doi:10.1177/0049124192021002005

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. doi:10.1037//0033-2909.105.3.456

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/107055107013. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. doi:10.1207/S15328007SEM0902_5

Church, A. T., Alvarez, J. M., Mai, N. T. Q., French, B. F., Katigbak, M. S., & Ortiz, F. A. (2011). Are cross-cultural comparisons of personality profiles meaningful? Differential item and facet functioning in the revised NEO personality inventory. *Journal of Personality and Social Psychology*, *101*(5), 1068–1089. doi:10.1037/a0025290

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology*, *66*(1), 93–114. doi:10.1037/0022-3514.66.1.93

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, *107*(5), 943–963. doi:10.1037/a0037245

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five–Factor Inventory (NEO–FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. doi:10.1146/annurev-soc-071913-043137

De Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO personality inventory: Psychometric properties, self-other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment*, *90*(2), 142–151. doi:10.1080/00223890701845195

De Vries, R. E., & van Kampen, D. (2010). The HEXACO and 5DPT models of personality: A comparison and their relationships with psychopathy, egoism, pretentiousness, immorality, and Machiavellianism. *Journal of Personality Disorders*, *24*(2), 244–257. doi:10.1521/pedi.2010.24.2.244

Dimitrova, R., Crocetti, E., Buzea, C., Jordanov, V., Kosic, M., Tair, E., … Uka, F. (2016). The Utrecht-Management of Identity Commitments Scale (U-MICS): Measurement invariance and cross-national comparisons of youth from seven European countries. *European Journal of Psychological Assessment*, *32*(2), 119–127. doi:10.1027/1015-5759/a000241

Dunlop, P. D., Morrison, D. L., Koenig, J., & Silcox, B. (2012). Comparing the Eysenck and HEXACO models of personality in the prediction of adult delinquency. *European Journal of Personality*, *26*(3), 194–202. doi:10.1002/per.824

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. doi:10.1037/0022-3514.59.6.1216

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment & Decision Making*, *13*, 356–371.

Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision Processes*, *117*(2), 298–310. doi:10.1016/j.obhdp.2011.11.002

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling*, *16*(1), 1–27. doi:10.1080/10705510802561279

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty-humility, social value orientations, and economic behavior. *Journal of Research in Personality*, *43*(3), 516–519. doi:10.1016/j.jrp.2009.01.003

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72–88. doi:10.1016/j.jrp.2015.04.003

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

Ion, A., Iliescu, D., Aldhafri, S., Rana, N., Ratanadilok, K., Widyanti, A., & Nedelcea, C. (2017). A cross-cultural analysis of personality structure through the lens of the HEXACO model. *Journal of Personality Assessment*, *99*(1), 25–34. doi:10.1080/00223891.2016.1187155

Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61–96). New York, NY: Academic Press.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review*, *78*(3), 229–248. doi:10.1037/h0030852

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory: Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. doi:10.1007/BF02291393

Labouvie, E., & Ruetsch, C. (1995). Testing for equivalence of measurement scales: Simple structure and metric invariance reconsidered.

Multivariate Behavioral Research, *30*(1), 63–76. doi:10.1207/s15327906mbr3001_4

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. doi:10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2005). Psychopathy, machiavellianism, and narcissism in the five-factor model and the HEXACO model of personality structure. *Personality and Individual Differences*, *38*(7), 1571–1582. doi:10.1016/j.paid.2004.09.016

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO personality inventory: Two new facet scales and an observer report form. *Psychological Assessment*, *18*(2), 182–191. doi:10.1037/1040-3590.18.2.182

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. doi:10.1177/1073191116659134

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the Dark Triad and Honesty–Humility. *European Journal of Personality*, *27*(2), 169–184. doi:10.1002/per.1860

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality traits beyond the Big Five: Are they within the HEXACO space? *Journal of Personality*, *73*(5), 1437–1463. doi:10.1111/j.1467-6494.2005.00354.x

Liu, J., Zettler, I., & Hilbig, B. E. (2016). Honesty-Humility. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Springer International Publishing, Cham, Germany. Advance online publication. https://doi.org/10.1007/978-3-319-28099-8_1228-1.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504. doi:10.1037/0033-2909.111.3.490

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology*, *60*(1), 1–34. doi:10.1111/j.1744-6570.2007.00063.x

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology*. (pp. 774–798). Hoboken, NJ: John Wiley & Sons Inc.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B. O., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524–545. doi:10.1037/met0000113

Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–491. doi:10.1037/a0019227

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*(3), 439–476. doi:10.1080/10705510903008220

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194–1218. doi:10.1037/a0026913

Marsh, H. W., Vallerand, R. J., Lafrenière, M.-A. K., Parker, P., Morin, A. J. S., Carbonneau, N., … Paquet, Y. (2013). Passion: Does one scale fit all? Construct validity of two-factor passion scale and psychometric invariance over different activities and languages. *Psychological Assessment*, *25*(3), 796–809. doi:10.1037/a0032573

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

McCrae, R. R., Zonderman, A. B., Costa, P. T. J., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552–566. doi:10.1037/0022-3514.70.3.552

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*(1), 97–103. doi:10.1007/BF01908590

McGrath, R. E. (2016). Measurement invariance in translations of the VIA inventory of strengths. *European Journal of Psychological Assessment*, *32*(3), 187–194. doi:10.1027/1015-5759/a000248

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568–592. doi:10.1037/0021-9010.93.3.568

Međedović, J., Čolović, P., Dinić, B. M., & Smederevac, S. (2017). The HEXACO personality inventory: Validation and psychometric properties in the Serbian language. *Journal of Personality Assessment. Advance Online Publication*, https://doi.org/10.1080/00223891.2017.1370426.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi:10.1007/BF02294825

Mischel, W. (1968). *Personality and assessment.* Hoboken, NJ: John Wiley & Sons Inc.

Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling*, *19*(1), 86–98. doi:10.1080/10705511.2012.634724

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 54–60. doi:10.1080/10705511.2014.950896

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorstruktur, psychometrische Eigenschaften und Messinvarianz der deutschen Version des 60-Item HEXACO Persönlichkeitsinventars [Factor structure, psychometric properties, and measurement invariance of the German-language version of the 60-item HEXACO personality inventory]. *Diagnostica*, *60*(2), 86–97. doi:10.1026/0012-1924/a000112

Moshagen, M., Thielmann, I., Hilbig, B. E., & Zettler, I. (in press). Meta-analytic investigations of the HEXACO personality inventory(-revised): Reliability generalization, self-observer agreement, intercorrelations, and relations to demographic variables. *Zeitschrift für Psychologie*.

Moshagen, M., Zettler, I., & Hilbig, B. E. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. doi:10.1037/rev0000111

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Ogunfowora, B., Bourdage, J. S., & Nguyen, B. (2013). An exploration of the dishonest side of self-monitoring: Links to moral disengagement and unethical business decision making. *European Journal of Personality*, *27*(6), 532–544. doi:10.1002/per.1931

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*(1), 401–421. doi:10.1146/annurev.psych.57.102904.190127

Romero, E., Villar, P., & López-Romero, L. (2015). Assessing six factors in Spain: Validation of the HEXACO-100 in relation to the five factor model and other conceptually relevant criteria. *Personality and Individual Differences*, *76*, 75–81. doi:10.1016/j.paid.2014.11.056

Roncero, M., Fornés, G., & Belloch, A. (2016). HEXACO: Una nueva aproximación a la evaluación de la personalidad en Español [HEXACO: A new approach to the personality assessment in Spanish]. *Revista de Psicopatología y Psicología Clínica*, *25*(1), 205–217. doi:10.5944/rppc.vol.19.num.1.2014.12929

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. doi:10.1177/0013164413498257

Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L., & Schulz, P. (2011). The perceived stress reactivity scale: Measurement invariance, stability, and validity in three countries. *Psychological Assessment*, *23*(1), 80–94. doi:10.1037/a0021148

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *9*(1), 1–12.

Tatar, A. (2018). Psychometric properties of the HEXACO Personality Inventory Revised (HEXACO-PI-R) long and short Turkish forms. *Anatolian Journal of Psychiatry*, *19*(Supplement 1), 5–13. doi:10.5455/apd.283098

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire big six in 26 nations: Developing cross-culturally applicable big six, big five and big two inventories. *European Journal of Personality*, *28*(5), 482–496. doi:10.1002/per.1969

Thielmann, I., & Böhm, R. (2016). Who does (not) participate in intergroup conflict? *Social Psychological and Personality Science*, *7*(8), 778–787. doi:10.1177/1948550616660160

Thielmann, I., & Hilbig, B. E. (2015). The traits one can trust: Dissecting reciprocity. *Personality and Social Psychology Bulletin*, *41*(11), 1523–1536. doi:10.1177/0146167215600530

Thompson, M. S., & Green, S. B. (2013). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 163–218). Charlotte, NC: IAP Information Age Publishing.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. doi:10.1080/17405629.2012.686740

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–69. doi:10.1177/109442810031002

Wakabayashi, A. (2014). A sixth personality domain that is independent of the Big Five domains: The psychometric properties of the HEXACO personality inventory in a Japanese sample. *Japanese Psychological Research*, *56*(3), 211–223. https://doi.org/10.1111/jpr.12045.

Wasti, S. A., Lee, K., Ashton, M. C., & Somer, O. (2008). Six Turkish personality factors and the HEXACO model of personality structure. *Journal of Cross-Cultural Psychology*, *39*(6), 665–684. doi:10.1177/0022022108323783

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association. https://doi.org/10.1037/10222-009.

Záškodná, T., & Dostál, D. (2016). Šestifaktorový model osobnosti a psychometrické vlastnosti Revidovaného osobnostního inventáře HEXACO. [Six-factor model of personality and psychometric properties of The HEXACO personality inventory–revised]. *Psychologie a Její Kontexty*, *7*(2), 31–42.

Żemojtel-Piotrowska, M. A., Piotrowski, J. P., Cieciuch, J., Calogero, R. M., Van Hiel, A., Argentero, P., … Wills-Herrera, E. (2017). Measurement of psychological entitlement in 28 countries. *European Journal of Psychological Assessment*, *33*(3), 207–217. doi:10.1027/1015-5759/a000286

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. approximate measurement invariance. *Frontiers in Psychology*, *6*, 733.

Zettler, I., & Hilbig, B. E. (2015). Honesty and humility. In J. D. Wright (Ed.), *International encyclopedia of social and behavioral sciences* (2nd ed., Vol. 11, pp. 169–174). Oxford, UK: Elsevier.

Zhao, K., Ferguson, E., & Smillie, L. D. (2017). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *Journal of Personality*, *85*(2), 244–256. doi:10.1111/jopy.12237

Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review*, *19*(3), 277–302. doi:10.1177/1088868314553709