Structural validity and measurement invariance of the short version of the Big Five Inventory (BFI-10) in selected countries

Renier Steyn & Takawira Munyaradzi Ndofirepi |

Published online: 01 Jul 2022.

Submit your article to this journal ⧉

Article views: 210

View related articles ⧉

View Crossmark data ⧉

cogent psychology

*Corresponding author: Takawira Munyaradzi Ndofirepi, School of Business Leadership, University of South Africa, Midrand, South Africa E-mail: takandofirepi@gmail.com

Reviewing editor:
Michael Daly, Maynooth University: National University of Ireland Maynooth, IRELAND

Additional information is available at the end of the article

SOCIAL PSYCHOLOGY | RESEARCH ARTICLE

Structural validity and measurement invariance of the short version of the Big Five Inventory (BFI-10) in selected countries

Renier Steyn[1] and Takawira Munyaradzi Ndofirepi[1]*

Abstract: We sought to determine the applicability and structural equivalence of a personality instrument developed in western, educated, industrialised, rich and democratic (WEIRD) contexts in non-WEIRD environments. The data for this study came from interviews conducted during the sixth wave of the World Values Survey in the Netherlands (N = 1902), Germany (N = 2046), Rwanda (N = 1527), and South Africa (N = 3531). We conducted exploratory and confirmatory factor analyses to assess structural validity and measurement invariance. The findings from the Big Five Inventory 10 (BFI-10) instrument did not support a perfect five-factor model as theorised by the Big Five Personality model across all countries, even though Germany and the Netherlands obtained better As a result, the findings do not support structural validity and do not demonstrate measurement invariance between WEIRD and non-WEIRD countries. The findings indicate that while the concise BFI-10 instrument partially replicates the structure of the B5P model in WEIRD countries, it falls short in non-WEIRD countries. Users of the instrument should therefore proceed with caution in both WEIRD and non-WEIRD contexts, bearing in mind the instrument's structural flaws.

Subjects: Testing, Measurement and Assessment; Cross-Cultural/ Multicultural Testing and Assessment; Psychometrics/ Testing & Measurement Theory

ABOUT THE AUTHORS

Renier Steyn is a full professor of leadership and organisational behaviour at the Graduate School of Business Leadership at the University of South Africa. He holds four doctorates from different South African universities. He conducts research on human resource management, workplace diversity, research design, psychometrics, supervision, and publication ethic

Takawira Munyaradzi Ndofirepi is a postdoctoral researcher at the Graduate School of Business Leadership at the University of South Africa. He earned a Doctor of Business Administration degree from South Africa's Central University of Technology. His research interests include entrepreneurship, consumer and organisational behaviour.

Takawira Munyaradzi Ndofirepi

PUBLIC INTEREST STATEMENT

Personality tests are important for determining a person's strengths and shortcomings. The outcomes of such examinations can be utilised to make potentially life-changing career decisions. However, failing to account for the fact that various cultural groups may attribute different meanings to items in a rating instrument can lead to erroneous judgments about people's personalities. As a result, it is critical to guarantee that the tests being utilised have been scientifically proven to be valid and reliable. The study described in this article looks at the psychometric features of the BFI-10 test in diverse cultural settings. The findings show that structural validity exists in WEIRD countries but not in non-WEIRD countries. The findings show that practitioners and researchers should proceed with caution when employing purportedly globally accepted assessment tools in contexts for which they were not designed.

1. Background and introduction

Since Tupes & Christal's, 1992/Tupes & Christal, 1992) naming, and Norman's (1963) replication of the five-factor Big Five Personality (B5P) traits model, it has become one of the central theoretical lenses in personality-related research. The model is derived from the psycholexical methodological approach to personality research, which emphasises the use of language to describe human personality traits. The significance of the five-factor structure stems from its robust ability to explain and predict individual differences regarding diverse topics such as mental health (Anglim & Horwood, 2021; Sun et al., 2018), job satisfaction (Bui, 2017), academic performance (Trapmann et al., 2007; Vedel, 2016), and work performance (Barrick & Mount, 1991; Pletzer et al., 2019).

The five major dimensions of the B5P model are extraversion, agreeableness, openness, conscientiousness, and neuroticism (Costa & McCrae, 1985). Costa and McCrae (1985) start by defining extraversion, stating it to be a personality trait comprising energy, talkativeness, and assertiveness. Second, agreeableness is explained as the degree of friendliness, cooperation, and compassion. Third, openness entails being perceptive and imaginative, in addition to possessing a diverse range of interests. Fourth, conscientiousness refers to the human characteristics of orderliness and thoroughness. Finally, neuroticism refers to an individual's emotional stability and susceptibility to negative emotions.

Research interest in the B5P has coincided with the proliferation of a variety of measurement instruments claiming to assess the big five traits. Subjectively, the measures can be classified as either long or short versions. Best known are Costa and McCrae's (1992) 240-item Revised NEO Personality Inventory, their 60-item NEO-Five Factor Inventory and their 44-item Big-Five Inventory. Taylor and De Bruin (2006) developed a South African version of the test, the Basic Traits Inventory (BTI), with 193 items. Short versions include Donnellan et al.'s (2006) 20-item International Personality Item Pool–Five Factor Model (IPIP–FFM) and Gerlitz and Schupp's (2005) 15-item Big Five Personality Inventory (BFI-S). Used in the World Values Survey (sixth wave) was Rammstedt and John's (2007) 10-item Big Five Personality Inventory (BFI-10).

While scholars of personality have agreed that personality traits generally fall within the five categories proposed by the B5P (John, 2021), a methodological concern is whether respondents react similarly to B5P items (Hahn et al., 2012). Within the South African context, Abrahams and Mauer (1999) and McDonald (2011) demonstrated that individuals from different cultural backgrounds may differ in their interpretations of the terminology used to categorise the big five personality traits. Consistent with this, Grobler and De Beer (2015, p. 50) observe that when participants from diverse cultural backgrounds are included in studies, the likelihood of "measurement bias from item interpretation differences is high, and empirical investigation of the items is important."

This study examined the psychometric properties of the BFI-10 instrument, a brief instrument for measuring the B5P factors, using data from the World Values Survey (WVS) of four culturally diverse countries, namely Germany, the Netherlands, Rwanda, and South Africa. The first two countries can be classified as western, educated, industrialised, rich, and democratic (WEIRD), and the second two as non-WEIRD. The goal was to determine whether a five-structure model similar to that proposed in the B5P is suited to application in both WEIRD and non-WEIRD countries. Based on data from the sixth wave of the WVS, Ludeke and Larsen (2017) as well as Simha and Parboteeah (2020) have raised concerns about the structural equivalence of the BFI-10 questionnaire. Ludeke and Larsen (2017) reported that when the questionnaire was used, indicators from the same scale tended to correlate negatively. Additional testing is necessary, as structural equivalence, which exists when a factor model is applicable across groups, is

a necessary condition for accurate statistical analyses across cultural groups, and this requirement must be objectively established (Fontaine et al., 2008).

Although Ludeke and Larsen (2017) found significant item-correlation issues with the Big Five measures in the WVS data, their analysis ignores the pattern of loadings and factor structures that emerge from the data in order to determine the aspects of the theorised model that are applicable to different countries. To analyse the factor structures and patterns of factor loadings in four countries with differing degrees of WEIRD-ness, we revisited the data for four countries with varying degrees of WEIRD-ness. This study is expected to contribute to a better understanding of the Big Five model's applicability in various countries. Considering this, the objectives of the study are as follows.

1.1. Main objective
To examine the psychometric properties of the BFI-10 instrument, a brief instrument for measuring the B5P factors, using data from the World Values Survey (WVS) of four culturally diverse countries, namely Germany, the Netherlands, Rwanda, and South Africa.

1.2. Sub-objectives
-to determine whether a five-structure model similar to that proposed in the B5P is suited for application in both WEIRD and non-WEIRD countries.
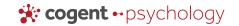
-to assess if the short version of the Big Five Inventory is measurement invariant in both WEIRD and non-WEIRD countries.

2. Literature review
Personality tests, as well as career aptitude and competence assessments, are useful for ascertaining people's strengths and weaknesses. The results of such tests can be used as a basis for potentially life-altering career decisions (Cascio & Aguinis, 2011). However, concerns have been raised about the direct transfer of assessment tools developed in developed economies to disadvantaged and culturally distinct contexts without considering the instrument validity implications in the contrasting contexts (Allik et al., 2017; Laajaj et al., 2019; Meiring et al., 2005). Inadequate consideration of the fact that different cultural groups may ascribe different meanings to items in a rating instrument can result in inaccurate judgements about individuals' personalities (Fontaine et al., 2008). In South Africa, the Employment Equity Act (No. 55 of 1988) offers a measure of protection in stipulating that "psychometric testing and other similar assessments of an employee are prohibited unless the test or test being used has been scientifically shown to be valid and reliable".

3. Measurement bias and personality assessment
From the foregoing, it is evident that measurement bias (incomparability or inequivalence) is an ever-present threat to the reliability and validity of personality scales in cross-cultural studies. In the literature, three categories of measurement bias are identified, namely construct, method and item bias (Berry et al., 2011). Construct bias occurs when a construct applies uniquely to a particular cultural group, or when construct indicators cannot be used across different groups (Fontaine et al., 2008). Method bias arises when supposed construct measures in an instrument do not measure the construct they are supposed to measure (Meiring et al., 2005). This may be due to translation errors, acquiescent responding, or group-influenced response patterns. Lastly, item bias occurs when a construct indicator systematically demonstrates a higher or lower score than expected with a particular group (F. Van de Vijver & Tanzer, 2004). While measurement bias and its consequence can emanate from issues such as different norms across different cultures, problems with translation issues, and language issues (Nye et al., 2008; Sass, 2011; Saucier et al., 2014; Thalmayer & Saucier, 2014; Thalmayer et al., 2020), a useful way to identify the possibility of such a problem with an instrument is by evaluating its measurement invariance (Jak et al., 2014). This concept is explained in the next subsection.

3.1. The measurement invariance concept

Wang et al. (2018) define measurement invariance as a statistical property of a research instrument that indicates whether it consistently measures a latent variable (construct) across groups of respondents. Wu et al. (2007) provide a similar characterisation, stating that "measurement invariance holds if and only if the probability of an observed score, given the true score and the group membership, is equal to the probability of that given only the true score" (p. 2). Measurement invariance is seen as a reliable indicator of the structural equivalence of a research instrument (F. J. Van de Vijver & Poortinga, 2002). Selig et al. (2008) offer a comparable and perhaps the most complete definition, stating that measurement invariance entails testing an evaluating tool to ascertain whether a latent variable it seeks to test across population groups provides comparable equivalence information. In other words, measurement invariance examines whether questionnaire items measuring a particular construct, such as personality, are understood in the same way by different groups. Groups typically used in invariance studies are age (Dong & Dumas, 2020), ethnic origin (Selig et al., 2008), socio-economic status (Hughes et al., 2021), level of education (Patel et al., 2019), nation (F. J. Van de Vijver & Poortinga, 2002) and occupation (Spurk et al., 2015).

The B5P assessment tools are examples of where measurement invariance testing is applicable, with a number of studies having been done to test the measurement invariance of the B5P scales (Chiorri et al., 2016; Laverdière et al., 2013; Schmitt et al., 2011). Measurement invariance is applicable in the case of the B5P scales, as it employs multiple and conceptually related clusters of items to assess the five personality characteristics. When measurement invariance holds, respondents from diverse groups assign the same meaning to each of the five personality clusters, allowing for the comparability of research results.

3.2. Assessing measurement invariance

The literature on measurement invariance has grown over the years. This is particularly evident from a study of the articles in the *Journal of Cross-Cultural Psychology* [0022–0221 (print); 1552–5422 (web)]. Essentially, assessing measurement invariance is affirmed through a stepwise process, using mainly confirmatory factor analysis, and follows a series of increasingly restrictive equality constraints hypotheses (Berry et al., 2011). In this sub-section, we discuss the categories of measurement invariance according to the different levels of model restriction. A literature search identified anything from three to five distinct levels/types of measurement invariance. Here the focus will be on the most comprehensive typology, which comprises conceptual, configural, metric, scalar, and strict invariance.

- Conceptual invariance implies that the domain or trait should make sense in all the groups to be compared (Berry et al., 2011). When a measured construct is specific to a particular context, it would thus be impossible to find a comparable operational pattern of relationships with other constructs across the groups (Fontaine et al., 2008). Although conceptual invariance is based mainly on theoretical arguments, and although no statistical tests directly test conceptual equivalence, Berry et al. (2011) state that evidence of configural invariance supports claims regarding conceptual equivalence.

- Configural (configurational) invariance is the fundamental type of invariance and is examined first, before any other types of invariance are considered. Configural invariance (pattern invariance) is confirmed when the number of factors and their loading patterns is consistent across groups (Bialosiewicz et al., 2013). However, with configural invariance the strength of the factor loadings may vary across population groups, and this does not guarantee structural equivalence across respondents' groups in a multigroup study, with additional tests being required to confirm group comparability.

- Metric invariance (also known as weak invariance) is a type of measurement invariance that is at a higher level than configural invariance. Unlike configural invariance, which requires only a similar number of factors and an identical pattern of factor loadings to confirm measurement invariance, metric invariance requires that the strength/size of the factor loadings be

equal across population groups (Davidov et al., 2014). Where there is no metric invariance, any comparison of constructs across groups should be performed with caution, as the constructs themselves are not identical (Marsh et al., 2012).

- Scalar invariance is the most robust and desired level of measurement invariance, according to many authors. Apart from the conditions of configural and metric invariance, the intercepts of the scale items must be equivalent across respondent groups (Melipillán & Hu, 2020), and only then will scalar invariance be indicated. When this level of invariance is achieved, meaningful comparisons of constructs across groups of respondents are possible (Li et al., 2018; Wang et al., 2018).

- The final level of measurement invariance, known as strict invariance, is concerned with the equivalence of residual error between groups (Bialosiewicz et al., 2013). Unlike the previously discussed levels of measurement invariance, strict invariance has two sublevels. The invariance of factor variances is the first level of strict invariance, where the error variances of the factors are equal across groups. The second level of strict invariance refers to the invariance of the error terms of an indicator variable, indicating that the unique errors of the indicator variable are equal across groups. Therefore, when you test for strict invariance, you are in essence determining whether your residual error is comparable across administrations (Bialosiewicz et al., 2013).
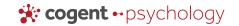
In measurement invariance studies, typically multi-group confirmatory factor analysis (MGCFA) is performed in phases. The first phase tests for configural invariance, the second for metric invariance, the third for scalar invariance, and the last for strict invariance. These phases are inextricably linked, and researchers frequently abandon testing when any of these steps exhibits non-invariance. While strict invariance is often tested for in standard syntaxes, most scholars agree that scalar invariance is sufficient for drawing meaningful conclusions about group comparability (Wang et al., 2016).

Confirmatory factor analysis indicators are used to determine an acceptable level of model fit and, ultimately, measurement invariance. Among the most commonly used goodness-of-fit indices are the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR), as well as the chi-square statistic (Bibi et al., 2020; Browne & Cudeck, 1993; Kim, 2017; Kong, 2017; Sun, 2005). For each of the indicators, Hu and Bentler (1999) propose the following cut-off criteria: CFI and TLI values of 0.9 or greater; RMSEA and SRMSR values of less than 0.08; and a chi-square statistic that is not statistically significant. Last-mentioned is seldom met. Lavaan (R; R Core Team, 2020) can be used to test measurement invariance up to the level of strict invariance, and both Svetina et al. (2020) and Steyn and De Bruin (2019) used Lavaan (R) to analyse their data, and applied the abovementioned guidelines to interpret the findings of their respective studies. Once measurement invariance is proven, group comparisons can be made to ascertain whether different categories of respondents interpret the multiple scale items measuring a particular construct similarly (Rhudy et al., 2020).

3.3. Analyses of abridged versions of B5P personality models

Although using detailed and often longer multi-item rating scales for B5P-related investigations is ideal due to the increased content validity and reliability, time limitations sometimes make this difficult (Rammstedt & John, 2007). Long questionnaires may tire participants, frustrate them, and elicit inattentive responding. In the end, the veracity the instruments' results may be jeopardised (Soto & John, 2017). As a result, abridged instrument versions have evolved with time.

Although short questionnaires are convenient in psychological studies, reservations over them, particularly the brief B5P scales, have been highlighted. For instance, it is claimed that lengthier B5P tests seem to be more reflective of broad personality constructs and have better measuring capacity than shorter ones (Langford, 2003). According to Gosling, Rentfrow, and Swann Jr. (2003),

abridged versions of the B5P show poorer psychometric qualities than the regular multi-item measure. Similar instrument validity concerns were observed by Laajaj et al. (2019) whose research based on a 15-item instrument could not accurately assess the target personality traits and found that the instrument had low validity.

Several studies have found the Rammstedt and John's (2007) 10-item BFI-10 scale to be a reliable measure of extraversion, agreeableness, openness, conscientiousness, and neuroticism in individuals (Balgiu, 2018; Guido et al., 2015. Other studies based on the WVS sixth wave (from 2010 to 2014), however, reveal the shortcomings of the BFI-10 scale's psychometric qualities when measuring the Big Five personality traits in cross-national scenarios (Ludeke & Larsen, 2017; Simha & Parboteeah, 2017). The conclusions regarding the challenges of the BFI-10 instrument were substantiated by findings from Chapman and Elliot's (2019) study based on General Social Survey data, which revealed odd results that did not replicate those of original big five instruments. Given the foregoing, the B5P model's universal application becomes problematic. Furthermore, the 10-item measure's persistent use in cross-cultural research without a consensus on its generalisability seems contentious. As a result, the literature supports the necessity for additional research into the structural validity of the WVS's 10-item measure.

4. Method
In this section the design, sampling, research instrument used, procedure, statistical analyses and ethical concerns are discussed.

4.1. Design
This study is based on the analysis of cross-sectional data on the B5P model collected during the World Values Survey (sixth wave) in Germany, the Netherlands, Rwanda, and South Africa. The data was quantitative in nature and collected by means of Rammstedt and John's (2007) 10-item Big Five Personality Inventory (BFI-10). The purpose of the analysis was either to confirm or refute the structural validity of the BFI-10 instrument using data from selected WEIRD and non-WEIRD nations. We acknowledge that the study's concentration on a small number of nations and a single dataset limits the universal applicability of its findings.

4.2. Sampling
The WVS website provides a detailed description of how individuals were sampled across countries (Inglehart et al., 2014), with every effort being made to sample individuals randomly per country.

The data from four countries, namely Germany, the Netherlands, Rwanda, and South Africa, was used in the study. The selection of countries for analysis was by no means arbitrary. As the focus of the study was on the B5P construct, and as this measure was only used in the sixth wave, the countries were selected from those which completed the questionnaires during that period. We chose only two Western European countries to represent WEIRD since their WVS data was found by Ludeke and Larsen (2017) to reflect the Big Five Model's five-factor structure. The inclusion of Rwanda and South Africa in the non-WEIRD category was motivated by the desire to include contrasting comparisons. The demographics of Rwandan respondents were substantially skewed toward non-WEIRD, but those of South African respondents were mixed, allowing for a good comparison of countries with diverse characteristics.

4.3. Measurement instrument
The BFI-10 designed by Rammstedt and John (2007) was used in the WVS to collect data on the theorised B5P constructs, namely extraversion, agreeableness, openness, conscientiousness, and neuroticism. The BFI-10 items are presented as five affirming statements (e.g., "I see myself as someone who is outgoing, sociable") and five disaffirm statements (e.g., "I see myself as someone who is reserved"), two statements per personality dimension. Respondents were required to rate the statements on a Likert scale with five response categories ranging from 1 = strongly disagree

cogent • psychology

Table 1. B5P-traits and their quantification		
Items: I see myself as someone who …	**B5P traits**	**Reverse coding**
… is reserved	Extraversion	Yes
… is generally trusting	Agreeableness	No
… tends to be lazy	Conscientiousness	Yes
… is relaxed, handles stress well	Neuroticism	No
… has few artistic interests	Openness	Yes
… is outgoing, sociable	Extraversion	No
… tends to find fault with others	Agreeableness	Yes
… does a thorough job	Conscientiousness	No
… gets nervous easily	Neuroticism	Yes
… has an active imagination	Openness	No

Table 2. Model fit cut-off criteria			
Measure	**Terrible**	**Acceptable**	**Excellent**
CMIN/DF	> 5	> 3	> 1
P-value	0.05<	>0.05	
CFI	<0.90	<0.95	>0.95
RMSEA	>0.08	>0.06	<0.06
PClose	<0.01	<0.05	>0.05

Source: Hu and Bentler (1999).

to 5 = strongly agree. The table below summarises the questionnaire items and how they were classified in the survey.

It can be observed from Table 1 that two items each, one positively worded and the other negatively worded, represent each of the five traits of the B5P-model.

Literature on the reliability and validity of the BFI-10 paints a mixed picture. Studies by Rammstedt and John (2007), Carciofo et al. (2016), Rammstedt and Krebs (2007), and Erdle and Rushton (2011) yielded respectable reliability coefficients for each of the five constructs in the five-factor model. Also, studies by Rammstedt and John (2007), Balgiu (2018), and Guido et al. (2015) yielded data which supported a five-factor model as was theorised in the B5F-model. Notwithstanding the preceding findings, later research by Ludeke and Larsen (2017) disconfirmed positive reliability findings and the validity of the five-factor hypothesis.

4.4. Procedure

The primary aim of the study was to test the structural validity of the B5P-model across countries, some of which were WEIRD, and others non-WEIRD. As stated, data from the WVS was used, specifically the SPSS file available on the WVS website (https://www.worldvaluessurvey.org/wvs.jsp).

As the aim of the study was to assess the fit of the B5P-model with data from four countries, exploratory factor analyses (EFA) were first performed, to visually assess whether the data fitted the B5P-model in each country. First, following Kaiser's rule regarding eigenvalues, the natural fit of the data was established. Then, using the SPSS software (IBM Corp, 2020), the data was "forced" into a five-factor solution, again one country at a time. As a final preparation before entering into multi-group confirmatory factor analyses (MGCFA), confirmatory factor analyses (CFA) were performed per country. Given the aforementioned results, multi-group

confirmatory factor analyses (MGCFA) were envisaged as a means to test for the levels of measurement invariance.

4.5. Statistical analyses

It was planned that exploratory factor analyses (EFA), confirmatory factor analyses (CFA), and multi-group confirmatory factor analyses (MGCFA) were to be used successively to assess the levels of measurement invariance. With CFA, the number of factors and their loading patterns (Bialosiewicz et al., 2013) were considered, while for CFA and MGCFA the guidelines below were used. Table 2 presents the criteria for evaluating the CFA results.

4.6. Ethical considerations

The use of the WVS data was open to all interested parties, subject to referencing the database in the reference list (see, Inglehart et al., 2014). No data specific to this research was collected.

5. Results

Data for the sixth wave of the WVS was collected in Germany (2013), the Netherlands (2012), Rwanda (2012), and South Africa (2013). In all, 8724 responses were collected: 2010 from Germany, 1739 from the Netherlands, 1527 from Rwanda, and 3448 from South Africa.

Table 3. Demographic variables and WEIRD classification					
		Germany	**Nether-lands**	**Rwanda**	**South Africa**
Sex	Men	48.9	46.5	49.6	48.3
	Women	51.1	53.5	50.4	51.7
Age	Mean	49.63	53.34	33.77	37.72
	Std. Dev.	17.94	16.44	11.23	15.67
W— Westernised [a]	White [b]	87.1	75.7	Low [d]	12.1
	Black [c]	.7	1.3	High [d]	76.5
E—Educated	Graduated [e]	11.7	11.5	7.7	4.2
I— Industrialised	PC use [f]	58.7	80.5	12.7	13.7
R—Rich	Savings [g]	57.0	50.5	30.6	19.6
D—Democratic	Perceived [h]	50.2	47.5	42.9	36.6

[a]Germany and the Netherlands are countries in Europe, and can therefore be classified as Western, whilst Rwanda and South Africa are African countries. However, ethnicity is also used as an indicator of being defined as either Western or non-Western. This aspect is reported here as well, with "white" interpreted as Western, and "black" as non-Western.

[b]Listed as Caucasian white in Germany and the Netherlands, and as white in South Africa.

[c]Listed as African in Germany, black Negro in the Netherlands, and black in South Africa.

[d]No questions on ethnic groups were posed to the Rwandan sample. However, when considering V247, namely "What language do you speak at home", 97.6% of respondents indicated Kinyarwanda and 0.8% Swahili, with 0.1% indicating English and 1.5% French. It would be a reasonable assumption that most respondents in the Rwandan sample would have been classified as African, and therefore as non-Western.

[e]V248 was used to create the variable "Educated". Presented here is the percentage of respondents who indicated that they had completed a university degree.

[f]V225 was used to create the "Industrialised" variable. The percentage of respondents who indicated that they use their personal computer frequently, rather than occasionally or seldom, is presented here.

[g]V141 was used to create the "Rich" variable. The percentage of respondents who provided answers to 8, 9 and 10 is relevant in this case.

[h]V237 was used to create the "Democratic" variable. Presented here is the percentage of respondents who stated that they had sufficient savings.

5.1. Demographic variables and WEIRD classification

In the table below, demographic data are presented per country. Apart from the customary reported data on age and sex, the report also includes WEIRD data per country. This was done because frequent criticism is levelled against psychometric instruments developed from the perspective of WEIRD societies (Doğruyol et al., 2019; Laajaj et al., 2019).

From Table 3 it can be observed that men are marginally underrepresented, but this applies across all the groups. Noteworthy are the differences in average age between the European and the African sample, namely 15 years, which is about the same as the standard deviation in the groups. This equates to a large effect size and a practically significant difference on age.

The data captured in Table 3 affirms that the European countries can be classified as WEIRD and the African countries as non-WEIRD. Stated differently, the European countries scored higher on all the WVS's proxies of WEIRD than the African countries did.

5.2. Statistics testing for measurement invariance

First the data from Germany was analysed using EFA. The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy was .617 for the German sample, and the Bartlett's Test of Sphericity (BTS) approximated chi-square was 1939.332 (df = 45), which was statistically significant (p < .001) (N = 2046). Kaiser's criterion of retaining factors with eigenvalues greater than one was used first, and resulted in the retention of four factors, accounting for 59.414 per cent of the variance. When the data was "forced" into a five-factor solution, the declared variance was 68.875 per cent. These findings are summarised in Table 4.

What can be observed in Table 4 is that the a priori model yielded the same number of factors and patterns of loadings as the theorised model. The loadings also fitted the positive–negative alternation, as was expected given the reverse wording of every other item. However, for component five (A(R)), the loading weight for that scale item was lower than what was theoretically expected. The model based on eigenvalues exposed four factors, where the first factor was more complex, but with the remaining three following the B5P conceptualisation.

The data for the Netherlands was analysed next. The KMO measure of sampling adequacy was .526 for the Netherlands sample, and the BTS approximated chi-square was 1510.445 (df = 45),

Table 4. Germany: EFA with components (a priori) and component (by eigenvalues greater than 1)										
G	**Five components (a priori)**					**Component with eigenvalues >1**				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**-**
E(R)	**−.769**	.091	−.042	.041	.157	**−.789**	−.035	.039	−.017	-
A	−.003	−.085	.008	−.09	**.878**	−.123	.342	.416	.215	-
C(R)	−.09	−.014	−.004	**.872**	.025	−.167	.131	**−.794**	.100	-
N	−.073	**−.713**	.12	.233	.33	−.100	**.801**	−.088	.164	-
O(R)	.026	.165	**−.815**	.177	.126	−.032	−.081	−.126	**−.709**	-
E	**.752**	−.128	.185	−.083	.187	**.728**	.17	.132	.233	-
A(R)	**.731**	.176	−.011	.097	**−.047**	**.707**	−.164	−.125	.031	-
C	−.15	−.042	.138	**−.603**	.444	−.158	.098	**.731**	.166	-
N(R)	−.08	**.853**	.028	.164	.086	−.153	**−.757**	−.116	.137	-
O	.221	.107	**.748**	.093	.199	.186	−.038	−.007	**.794**	-

Note: G = Germany, E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neuroticism, O = Openness, R = Reverse coding. Also, in **bold** in the table are the two highest loadings per component, and "shadowed" are the a priori loadings, as per the B5P model, where the highest loading was used as a marker item.

cogent · · psychology

Table 5. Netherlands: EFA with components (a priori) and component (by eigenvalues greater than 1)

N	Five components (a priori)					Component with eigenvalues >1				
	1	2	3	4	5	1	2	3	4	5
E(R)	.109	.117	−.042	**−.774**	.296	.109	.117	−.042	**−.774**	.296
A	−.120	−.090	−.022	.008	**.785**	−.120	−.090	−.022	.008	**.785**
C(R)	.138	−.059	**.825**	−.134	.138	.138	−.059	**.825**	−.134	.138
N	**−.781**	.065	.045	.136	.297	**−.781**	.065	.045	.136	.297
O(R)	.050	**.846**	.028	.017	.074	.050	**.846**	.028	.017	.074
E	−.018	−.049	−.152	**.736**	.327	−.018	−.049	−.152	**.736**	.327
A(R)	.461	.241	.345	.297	.152	.461	.241	.345	.297	.152
C	.090	.080	−.673	−.043	.407	.090	.080	**−.673**	−.043	.407
N(R)	**.853**	−.050	.062	−.055	.047	**.853**	−.050	.062	−.055	.047
O	.113	**−.701**	.140	.172	.200	.113	**−.701**	.14	.172	.200

Note: N = Netherlands, E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neuroticism, O = Openness, R = Reverse coding. Also, in **bold** in the table are the two highest loadings per component, and "shadowed" are the a priori loadings, as per the B5P model, where the highest loading was used as a marker item.

Table 6. Rwanda: EFA with components (a priori) and component (by eigenvalues greater than 1)

R	Five components (a priori)					Component with eigenvalues >1				
	1	2	3	4	5	1	2	3	4	-
E(R)	.261	**.819**	.132	.054	−.100	.323	.251	**.753**	.001	-
A	−.058	**.869**	.039	.009	.218	−.036	.038	**.883**	.129	-
C(R)	−.097	.119	**.834**	.015	.135	−.237	**.718**	.174	.097	
N	.187	.211	.034	.272	**.808**	.047	−.169	.352	**.683**	-
O(R)	.010	−.008	.100	**.927**	.098	.071	.259	−.132	**.825**	-
E	.462	.202	−.376	.482	.207	**.523**	−.273	.148	**.526**	-
A(R)	.421	.322	.320	.326	**−.534**	**.528**	**.593**	.136	−.008	-
C	**.824**	.107	.008	.072	−.062	**.816**	.083	.073	.054	-
N(R)	.315	.067	**.665**	.075	−.277	.272	**.732**	.008	−.071	-
O	**.843**	.032	.093	.006	.119	**.770**	.075	.058	.102	-

Note: R = Rwanda, E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neuroticism, O = Openness, R = Reverse coding. Also, in **bold** in the table are the two highest loadings per component, and "shadowed" are the a priori loadings, as per the B5P model, where the highest loading was used as a marker item.

which was statistically significant (p < .001) (N = 1902). When Kaiser's criterion of retaining factors with eigenvalues greater than one was applied, five factors were retained, accounting for 66.859 per cent of variance. These findings are summarised in Table 5.

Table 5 shows that, as was the case with the German sample, the a priori model had five factors with patterns of loadings reflecting those theorised in the B5P model. However, for component five (A(R)), the loading weight of the scale items was lower than expected. This is the same item for which loading was not satisfactory in the German sample. Worse in the case of the Netherlands was that the loading of the items was not signed contrary to that of the other marker variable. When the model was based on eigenvalues, five factors were found, which implies that the "forced" and the eigenvalue models were identical, both providing support to the B5P conceptualisation.

Table 7. South Africa: EFA with components (a priori) and component (by eigenvalues greater than 1)										
p	Five components (a priori)					Component with eigenvalues >1				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**-**	**-**	**-**
E(R)	.183	.142	.127	**.892**	.033	**.632**	.128	-	-	-
A	−.081	.306	.28	**.683**	.237	**.778**	−.024	-	-	-
C(R)	**.846**	−.111	.145	.018	.052	−.062	**.824**	-	-	-
N	.104	.28	**.846**	.159	.048	**.683**	.188	-	-	-
O(R)	.196	.13	**.723**	.221	.326	**.639**	.34	-	-	-
E	.096	.261	.25	**.133**	**.825**	**.581**	.364	-	-	-
A(R)	**.627**	.135	.044	.114	**.564**	.256	**.764**	-	-	-
C	−.004	**.836**	.189	.189	.171	**.733**	.073	-	-	-
N(R)	**.699**	.479	**.123**	.123	.063	.345	**.693**	-	-	-
O	.164	**.626**	.295	.271	.245	**.715**	.253	-	-	-

Note: S = South Africa, E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neuroticism, O = Openness, R = Reverse coding. Also, in **bold** in the table are the two highest loadings per component, and "shadowed" are the a priori loadings, as per the B5P model, where the highest loading was used as a marker item.

The data for Rwanda was analysed next. The KMO measure of sampling adequacy was .611 for the Rwanda sample, and the BTS estimated chi-square was 1693.435 (df = 45), which was statistically significant (p < .001) (N = 1527). When Kaiser's criterion for maintaining variables with eigenvalues greater than one was applied, four factors were maintained, accounting for 61.395 per cent of the variance. The total variance explained was 82.646 per cent when "forcing" the data into the five-factor solution. Table 6 summarises these findings.

What can be observed is that the number of factors and patterns of loadings on the factor components based on the eigenvalues greater than one do not match those theorised in the B5P model. The indicator items loaded haphazardly across the factors such that no definite personality construct in line with the B5P model could be identified. Also, when the data was "forced" into the five-factor solution, the patterns so pronounced in the WEIRD-countries were absent.

Lastly, data for South Africa was analysed. The KMO measure of sampling adequacy for the South African sample was .872, and the BTS approximate chi-square was 1057.903 (df = 45), which was statistically significant (p < .001) (N = 3531). When Kaiser's criterion of retaining factors with eigenvalues greater than one was applied, two factors were retained, accounting for 55.644 per cent of the variance. When the data was "forced" into the five-factor solution, the declared variance was 76.606 per cent. These findings are shown in Table 7.

Only two components were derived based on the strategy of selecting eigenvalues greater than one. As shown in Table 7, the patterns of the item loadings do not reflect any particular personality construct as proposed in the five-factor model. Also, when considering the forced five-factor solution, none of the patterns typical of B5P, as observed in the samples from Germany and the Netherlands, were observed.

The results presented in Tables 4 to 8 show compelling evidence that the BFI-10 functions at different levels of effectiveness: in the WEIRD countries it follows the B5P conceptualisation, while this does not occur in the non-WEIRD countries.

The aforementioned conclusion is based on exploratory factor analyses (EFA), and the interpretation involved some subjectivity. To quantify these tentative conclusions, confirmatory factor analysis (CFA) was used to evaluate for structural validity in a thorough and comprehensive

Table 8. CFA model fit measures for four countries					
Measure	**Threshold**	**South Africa**	**Rwanda**	**Nether-lands**	**Germany**
Chi-square, degrees of freedom, p-value	<0.05	Chi-square = 974.554, df = 34, p-value = 0.000	Chi-square = 947.158, df = 25, p-value = 0.000	Chi-square = 284.189, df = 25, p-value = 0.000	Chi-square = 474.078, df = 26, p-value = 0.000
CMIN/DF	1< and >3	∞	∞	NaN	NaN
CFI	>.95	.817	.684	.831	.789
RMSEA	<.06	.122	.155	.074	.092
PClose	>.05	0	0	0	0

Note: ∞ = infinity; NaN = Not a Number

manner, again per country. Five factors were postulated, with two items loading on each factor, as explained in Table 1. Hu and Bentler's (1999) cut-off criteria for fit indices, as presented in Table 3, were applied. The test results for each of the four countries are presented in Table 8.

According to the cut-off criteria, the findings revealed a poor fit across all counties, including those which had comparatively better patterns of loadings in EFA. These results were surprising, to say the least, and will be discussed in the discussion below.

6. Discussion

The overall aim of the study was to establish the structural validity and measurement of variance of the BFI-10 instrument in WEIRD and non-WEIRD countries. The novelty of the study lies in its attempt to compare and contrast the psychometric properties of the BFI-10 instrument in the culturally different environments.

The literature review focused on the B5P conceptualisation, as well as the concept of measurement invariance and the need to test for it. It was revealed that although the B5P conceptualisation is well accepted, it is not without critique, particularly in terms of its use as a universal theory of personality. With regard to measurement invariance, the concept, as well as how it should be assessed, was considered.

The data revealed that the countries included in the study were clearly differentiable based on the WEIRD concept, with the two Western-European counties being classified as WEIRD, and the two from sub-Saharan Africa as non-WEIRD. Another characteristic which differentiated the countries was mean age, which was substantially higher in the WEIRD countries. It may well be asked whether the acronym WEIRDO, in which O stands for old, might perhaps be applicable. The implications of including older respondents from the WEIRD group in the analyses of personality can only be speculated on. Lang et al. (2011) hypothesised that the mental strain associated with personality studies could preclude elderly respondents from providing valid self-report responses, thereby decreasing the likelihood of deriving a compact five-factor model. They did note, however, that more educated elderly people are likely to cope better with mental strain than less educated elderly people, and thus provide more consistent item responses during surveys. Given that B5P conceptualisation is based largely on trait theory, the fact that data was extracted from a more mature sample may be irrelevant.

The result of the EFA suggests a partially valid model in the WEIRD contexts and an invalid one in the non-WEIRD contexts (see, Tables 4 to 7). The results reveal that, at a configural level of measurement invariance, WEIRD countries met the criteria (to a large degree), but that the non-WEIRD data did not support the proposed theoretical structure at all.

These results underscore the psychometric difficulties relating to comparison of personality score levels prevalent in cross-cultural studies. This corroborates Hofstede and McCrae's (2004) affirmation that perception of personality dimensions is not divorced from cultural context. Hence, researchers should not assume that personality instruments are equally valid in settings other than those in which they were developed.

Even though the outcome of the EFA for the WEIRD countries suggested a factor structure almost equivalent to the theorised model, it still fell short of the ideal. When a more comprehensive statistic is used to test for configural fit, the model fit indicators (CFI and RMSEA) reflected a poor model fit for all four countries. These results were surprising, as the EFA results were quite satisfactory for Germany and the Netherlands. Given the CFA results, it should be stated that at the most basic form of measurement invariance, that is, configural invariance, the BFI-10 failed.

All further tests of measurement invariance were abandoned, as testing for measurement invariance is a sequential process, where testing for higher levels of invariance is done only once certain milestones have been achieved (Berry et al., 2011). Thus, no tests were performed on configural, metric, scalar, and strict invariance.

Other studies, notably those of Ludeke and Larsen (2017)have since shown the measurement limitations of using the BFI-10 instrument. However, Ludeke and Larsen (2017) found it to be a reliable and valid tool in Germany and the Netherlands (the WEIRD countries). Previous studies within the WVS domain (Balgiu, 2018; Rammstedt & John, 2007) found satisfactory levels of structural validity and measurement invariance. Several studies outside the WVS domain have found the BFI-10 to be a reliable, valid, convenient and useful tool for gauging self-reported personality traits (Erdle & Rushton, 2011; Guido et al., 2015; Rammstedt & Krebs, 2007).

These results affirm the problematic nature of the BFI-10 tool, and this may be the reason why its use in the WVS has since been discontinued, and it does not appear in the WVS seventh wave questionnaire. Notwithstanding its limitations identified in the WVS sixth wave data, the BFI-10 remains a useful instrument for researchers, perhaps particularly so in WEIRD environments. Thus, it is recommended that the psychometric properties of even well-established personality rating tools need to be examined, particularly if they are used in environments foreign to where they were developed. There is also evidence that disparities in schooling affect structural validity even within a country (Lang et al., 2001; Rammstedt et al., 2010). As a result, the findings may be attributed less to WEIRD vs. non-Weird and more to educational differences. Future studies should probe further the possible influence of the variations in respondents' levels of education and other demographic variables on the structural validity of the BFI-10 instrument.

7. Conclusion

The article discusses the importance of assessing the psychometric properties of well-established personality rating instruments when they are used across diverse cultural groups. Preliminary results (using EFA) indicate that the BFI-10 instrument has configural structural validity and measurement invariance in WEIRD countries, but not in non-WEIRD countries. Further investigations (using CFA) revealed that not even in WEIRD countries was the BF-10 structurally valid. If nothing else, this indicates the importance of using multiple statistical techniques to gain information on a specific question. The results suggest that practitioners and researchers should adopt a cautious approach when applying ostensibly globally accepted tools in contexts for which they were not designed, and that, particularly as research findings concerning the BFI-10 are so contradictory, further research on this instrument be conducted.

Lastly, it seems that the WVS data supported the WEIRD categorisation of countries, even though the results did not fit the theorised factor structure perfectly. It was noted that

respondents in these countries were also older than those in non-WEIRD countries, opening up the opportunity for further studies on whether an O (for old) can be added to the acronym.

Author details
Renier Steyn[1]
ORCID ID: http://orcid.org/0000-0002-2446-3662
Takawira Munyaradzi Ndofirepi[1]
E-mail: takandofirepi@gmail.com
ORCID ID: http://orcid.org/0000-0001-7409-2241
[1] School of Business Leadership, University of South Africa, Midrand, South Africa.

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

Data availability statement
The data that support the findings of this study are openly available on the following website: http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp

Citation information
Cite this article as: Structural validity and measurement invariance of the short version of the Big Five Inventory (BFI-10) in selected countries, Renier Steyn & Takawira Munyaradzi Ndofirepi, *Cogent Psychology* (2022), 9: 2095035.

References
Abrahams, F., & Mauer, K. F. (1999). The comparability of the constructs of the 16PF in the South African context. *Journal of Industrial Psychology, 25*, 53–59. https://doi.org/10.4102/sajip.v25i1.679

Allik, J., Church, A. T., Ortiz, F. A., Rossier, J., Hřebíčková, M., De Fruyt, F., Realo, A., & McCrae, R. R. (2017). Mean profiles of the NEO personality inventory. *Journal of Cross-Cultural Psychology, 48*(3), 402–442. https://doi.org/10.1177/0022022117692100

Anglim, J., & Horwood, S. (2021). Effect of the COVID-19 pandemic and Big Five personality on subjective and psychological well-being. *Social Psychological and Personality Science, 12*(8), 1527–1537. https://doi.org/10.1177/1948550620983047

Balgiu, B. A. (2018). The psychometric properties of the Big Five Inventory-10 (BFI-10) including correlations with subjective and psychological well-being. *Global Journal of Psychology Research: New Trends and Issues, 8*(2), 61–69.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Berry, J. W., Poortinga, Y. H., Breugelmans, S. M., Chasiotis, A., & Sam, D. L. (2011). *Cross-cultural psychology: Research and applications* (3rd ed.). Cambridge University Press.

Bialosiewicz, S., Murphy, K., & Berry, T. (2013). *An introduction to measurement invariance testing: Resource packet for participants*. American Evaluation Association.

Bibi, A., Lin, M., Zhang, X. C., & Margraf, J. (2020). Psychometric properties and measurement invariance of Depression, Anxiety and Stress Scales (DASS-21) across cultures. *International Journal of Psychology, 55*(6), 916–925. https://doi.org/10.1002/ijop.12671

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Bui, H. T. (2017). Big Five personality traits and job satisfaction: Evidence from a national sample. *Journal of General Management, 42*(3), 21–23. https://doi.org/10.1177/0306307016687990

Carciofo, R., Yang, J., Song, N., Du, F., Zhang, K., & Qiu, J. (2016). Psychometric evaluation of Chinese-language 44-item and 10-item Big Five personality inventories, including correlations with chronotype, mindfulness and mind wandering. *PloS One, 11*(2), e0149963. https://doi.org/10.1371/journal.pone.0149963

Cascio, W. F., & Aguinis, H. (2011). *Applied psychology in human resource management talent management*. Pearson.

Chapman, B. P., & Elliot, A. J. (2019). Brief report: How short is too short?. *Journal of Health Psychology, 24*(11), 1568–1573. https://doi.org/10.1177/1359105317720819

Chiorri, C., Marsh, H. W., Ubbiali, A., & Donati, D. (2016). Testing the factor structure and measurement invariance across gender of the Big Five Inventory through exploratory structural equation modelling. *Journal of Personality Assessment, 98*(1), 88–99. https://doi.org/10.1080/00223891.2015.1035381

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO personality inventory manual*. Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. 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. https://doi.org/10.1146/annurev-soc-071913-043137

Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences, 1511*, 109547. https://doi.org/10.1016/j.paid.2019.109547

Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences, 160*(1), 109956. https://doi.org/10.1016/j.paid.2020.109956

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Elliot A J and Chapman B P. (2016). Socioeconomic status, psychological resources, and inflammatory markers: Results from the MIDUS study. *Health Psychology, 35*(11), 1205–1213. https://doi.org/10.1037/hea0000392

Employment Equity Act [No. 55 of 1998]. https://www.labour.gov.za/DocumentCenter/Acts/Employment%20Equity/Act%20-%20Employment%20Equity%201998.pdf Retrieved 22 November 2021, from

Erdle, S., & Rushton, J. P. (2011). Does self-esteem or social desirability account for a general factor of personality (GFP) in the Big Five? *Personality and Individual Differences, 50*(7), 1152–1154. https://doi.org/10.1016/j.paid.2010.12.038

Fontaine, J. R., Poortinga, Y. H., Delbeke, L., & Schwartz, S. H. (2008). Structural equivalence of the values domain across cultures: Distinguishing sampling fluctuations from meaningful variation. *Journal of Cross-Cultural Psychology, 39*(4), 345–365. https://doi.org/10.1177/0022022108318112

Gerlitz, J. Y., & Schupp, J. (2005). *Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP*. DIW.

Gosling S D, Rentfrow P J and Swann W B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37(6), 504–528. 10.1016/S0092-6566(03)00046-1

Grobler, S., & De Beer, M. (2015). Psychometric evaluation of the basic traits inventory in the multilingual South African environment. *Journal of Psychology in Africa, 25*(1), 50–55. https://doi.org/10.1080/14330237.2014.997033

Guido, G., Peluso, A. M., Capestro, M., & Miglietta, M. (2015). An Italian version of the 10-item Big Five Inventory: An application to hedonic and utilitarian shopping values. *Personality and Individual Differences, 76* 1 , 135–140. https://doi.org/10.1016/j.paid.2014.11.053

Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality, 46*(3), 355–359. https://doi.org/10.1016/j.jrp.2012.03.008

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38*(1), 52–88. https://doi.org/10.1177/1069397103259443

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

Hughes, B. T., Costello, C. K., Pearman, J., Razavi, P., Bedford-Petersen, C., Ludwig, R. M., & Srivastava, S. (2021). *The Big Five across socioeconomic status: Measurement invariance, relationships, and age trends. Collabra: Psychology*, 7(1). University of Chicago Press. https://psyarxiv.com/wkhfx/download?format=pdf

IBM Corp. (2020). *IBM SPSS Statistics for Windows, Version 27.0*.

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B. (eds.). (2014). *World values survey: Round six – country-pooled datafile version*. JD Systems Institute. www.worldvaluessurvey.org/WVSDocumentationWV6.jsp

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(1), 31–39. https://doi.org/10.1080/10705511.2014.856694

John, O. P. (2021). History, measurement, and conceptual elaboration of the Big Five trait taxonomy: The paradigm matures. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (pp. 35–82). Guilford Press.

Kim, S. (2017). Developing an item pool and testing measurement invariance for measuring public service motivation in Korea. *International Review of Public Administration, 22*(3), 231–244. https://doi.org/10.1080/12294659.2017.1327113

Kong, F. (2017). The validity of the Wong and Law Emotional Intelligence Scale in a Chinese sample: Tests of measurement invariance and latent mean differences across gender and age. *Personality and*

Individual Differences, 116(1), 29–31. https://doi.org/10.1016/j.paid.2017.04.025

Laajaj, R., Macours, K., Hernandez, D. A. P., Arias, O., Gosling, S. D., Potter, J., Rubio-Codina, M., & Vakis, R. (2019). Challenges to capture the Big Five personality traits in non-WEIRD populations. *Science Advances, 5* (7), eaaw5226. https://doi.org/10.1126/sciadv.aaw5226

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica, 47*(3), 111–121. https://doi.org/10.1026//0012-1924.47.3.111

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior Research Methods, 43*(2), 548–567. https://doi.org/10.3758/s13428-011-0066-z

Laverdière, O., Morin, A. J., & St-Hilaire, F. (2013). Factor structure and measurement invariance of a short measure of the Big Five personality traits. *Personality and Individual Differences, 55*(7), 739–743. https://doi.org/10.1016/j.paid.2013.06.008

Li, M., Wang, M., Shou, Y., Zhong, C., Ren, F., Zhang, X., & Yang, W. (2018). Psychometric Properties and Measurement Invariance of the Brief Symptom Inventory-18 Among Chinese Insurance Employees. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00519

Ludeke, S. G., & Larsen, E. G. (2017). Problems with the Big Five assessment in the world values survey. *Personality and Individual Differences, 112*(1), 103–105. https://doi.org/10.1016/j.paid.2017.02.042

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2012). 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. Advance online publication. https://doi.org/10.1037/a0026913

McDonald, E. (2011). *Comparing a native English-speaking group's and non-native English-speaking group's understanding of the vocabulary used in the 16PF5*. [Unpublished Master's Thesis]. University of South Africa.

Meiring, D., Van de Vijver, A. J. R., Rothmann, S., & Barrick, M. R. (2005). Construct, item and method bias of cognitive and personality tests in South Africa. *SA Journal of Industrial Psychology, 31*(1), 1–8. https://doi.org/10.4102/sajip.v31i1.182

Melipillán, E. R., & Hu, M. (2020). *Measurement invariance across groups*. SAGE.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology, 66*(6), 574–686. https://doi.org/10.1037/h0040291

Nye, C. D., Roberts, B. W., Saucier, G., & Zhou, X. (2008). Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality, 42*(6), 1524–1536. https://doi.org/10.1016/j.jrp.2008.07.004

Patel, J. S., Oh, Y., Rand, K. L., Wu, W., Cyders, M. A., Kroenke, K., & Stewart, J. C. (2019). Measurement invariance of the patient health questionnaire-9 (PHQ-9) depression screener in US adults across sex, race/ethnicity, and education level: Nhanes 2005–2016. *Depression and Anxiety, 36*(9), 813–823. https://doi.org/10.1002/da.22940

Pletzer, J. L., Bentvelzen, M., Oostrom, J. K., & De Vries, R. E. (2019). A meta-analysis of the relations

between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behavior*, *112*(1), 369–383. https://doi.org/10.1016/j.jvb.2019.04.004

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rammstedt, B., & Krebs, D. (2007). Does response scale format affect the answering of personality scales? Assessing the Big Five dimensions of personality with different response scales in a dependent sample. *European Journal of Psychological Assessment*, *23*(1), 32–38. https://doi.org/10.1027/1015-5759.23.1.32

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

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, *44*(1), 53–61. https://doi.org/10.1016/j.jrp.2009.10.005

Rhudy, J. L., Arnau, R. C., Huber, F. A., Lannon, E. W., Kuhn, B. L., Palit, S., Payne, M. F., Sturycz, C. A., Hellman, N., Guereca, Y. M., Toledo, T. A., & Shadlow, J. O. (2020). Examining configural, metric, and scalar invariance of the pain catastrophizing scale in native American and non-Hispanic White adults in the Oklahoma study of Native American pain risk (OK-SNAP). *Journal of Pain Research*, *13*(1), 961. https://doi.org/10.2147/JPR.S242126

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Saucier, G., Thalmayer, A. G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., Church, A. T., Katigbak, M. S., Somer, O., Szarota, P., Szirmák, Z., & Zhou, X. (2014). A basic bivariate structure of personality attributes evident across nine languages. *Journal of Personality*, *82*(1), 1–14. https://doi.org/10.1111/jopy.12028

Schmitt, N., Golubovich, J., & Leong, F. T. (2011). Impact of measurement invariance on construct correlations, mean differences, and relations with external correlates: An illustrative example using Big Five and RIASEC measures. *Assessment*, *18*(4), 412–427. https://doi.org/10.1177/1073191110373223

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modelling in cross-cultural research: multi-group and multi-level approaches . In F. J. R Van de Vijver, D. A Van Hemert, Y. H Poortinga. (Eds.), *Multilevel Analysis of Individuals and Cultures* (pp. 93–119). Lawrence Erlbaum.

Simha, A., & Parboteeah, K. P. (2020). The big 5 personality traits and willingness to justify unethical behavior: A cross-national examination. *Journal of Business Ethics*, *167*(3), 451–471. https://doi.org/10.1007/s10551-019-04142-7

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68* 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Spurk, D., Abele, A. E., & Volmer, J. (2015). The career satisfaction scale in context: A test for measurement invariance across four occupational groups. *Journal*

of *Career Assessment*, *23*(2), 191–209. https://doi.org/10.1177/1069072714535019

Steyn, R., & De Bruin, G. (2019). The structural validity of the innovative work behaviour questionnaire: Comparing competing factorial models. *The Southern African Journal of Entrepreneurship and Small Business Management*, *11*(1), 1–11. https://doi.org/10.4102/sajesbm.v11i1.291

Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, *37*(4), 240–256. https://doi.org/10.1080/07481756.2005.11909764

Sun, J., Kaufman, S. B., & Smillie, L. D. (2018). Unique associations between Big Five personality aspects and multiple dimensions of well-being. *Journal of Personality*, *86*(2), 158–172. https://doi.org/10.1111/jopy.12301

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the Lavaan/SEM tools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Taylor, N., & De Bruin, G. P. (2006). *Basic traits inventory*. Jopie van Rooyen.

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. https://doi.org/10.1002/per.1969

Thalmayer, A. G., Saucier, G., Ole-Kotikash, L., & Payne, D. (2020). Personality structure in east and West Africa: Lexical studies of personality in Maa and Supyire-Senufo. *Journal of Personality and Social Psychology*, *119*(5), 1132. https://doi.org/10.1037/pspp0000264

Trapmann, S., Hell, B., Hirn, J. O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Zeitschrift Für Psychologie/Journal of Psychology*, *215*(2), 132–151. https://doi.org/10.1027/0044-3409.215.2.132

Tupes, E. C., & Christal, R. E. Recurrent personality factors based on trait ratings. (1992). *Journal of Personality*, *60*(2), 225–251. Original work published 1961. https://doi.org/10.1111/j.1467-6494.1992.tb00973.x

Van de Vijver, F. J., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology*, *33*(2), 141–156. https://doi.org/10.1177/0022022102033002002

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Vedel, A. (2016). Big Five personality group differences across academic majors: A systematic review. *Personality and Individual Differences*, *92*(1), 1–10. https://doi.org/10.1016/j.paid.2015.12.011

Wang, S., Chen, C. C., Dai, C. L., & Richardson, G. B. (2018). A call for, and beginner's guide to, measurement invariance testing in evolutionary psychology. *Evolutionary Psychological Science*, *4*(2), 166–178. https://doi.org/10.1007/s40806-017-0125-5

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research, and Evaluation*, *12*(3), 1–26. https://doi.org/10.7275/mhqa-cd89

cogent ·· psychology

Cogent Psychology (ISSN: 2331-1908) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com