RESEARCH ARTICLE:

The Cross-cultural Structural Validity of the Big Five Personality Inventory (BFI-10) in a South African Sample

Takawira Munyaradzi Ndofirepi[1] and Renier Steyn[2]

Abstract

The study sought to assess the structural validity of the 10-item measure of the B5P instrument in the South African context. World Values Survey data, collected in South Africa (N = 3 531), were analysed using exploratory and confirmatory factor analyses, to ascertain the factorial structure of the data, including across sub-groups, focusing on measurement invariance. The theorised factorial structure of the B5P survey did not mirror the theorised structure in the South African sample. This was demonstrated in the inspection report, as well as in the tests of measurement invariance. Even sub-groups, typifying the Westernised, educated, industrialised, wealthy, and democratic part of the South African society, did not provide structures that mirrored the theorised model. The assumption that well-established instruments are valid in settings, different to the one where they were initially developed in, should be questioned, and such instruments should not be used, unless thoroughly tested.

Keywords: Big Five personality; World Values Survey; measurement invariance; structural validity; cross-cultural research

Introduction

The Big Five personality (B5P) model has a long history, stretching back to Fiske (1949), and including researchers such as Norman (1967), Smith (1967) and Goldberg (1981), as well as McCrae and Costa (1987). Methodologically, the B5P model is based on psycho-lexical studies, such as those conducted by Allport and Odbert (1936), and, later, Cattell (1950). It is a method still used extensively in contemporary studies of personality, including South Africa (Hill *et al*., 2013). John (2021: 35) states that, after decades of research, and long debates about the right number of factors and the best labels for these factors, the field has now achieved an initial consensus on a general taxonomy of personality traits: the 'Big Five' personality dimensions. Many versions of B5P measures were published, of which the 240-item Revised NEO Personality Inventory, the 60-item NEO Five-Factor Inventory, and the 44-item Big Five Inventory, all from Costa and McCrae (1992), are the most well-known versions. The Basic Traits Inventory (BTI) questionnaire (including 193 items), which is a South African instrument, also assessed the B5P assessment. Several shorter-versions of the B5P are available, such as the Big Five Personality Inventory (BFI-10) (Rammstedt and John, 2007) and the 20-item mini–International Personality Item Pool-Five-Factor Model (mini-IPIP-FFM) (Donnellan, Oswald *et al*., 2006), which has gained popularity in psychology research in recent times (Rammstedt *et al*., 2018; Soto and John, 2017; Nunes *et al*., 2018). The World Values Survey (WVS) (6th Wave) included a 10-item version of the B5P (Rammstedt and John, 2007) and applied it across 24 countries. It could be presumed that the designers of the WVS believed that this version of the B5P were applicable to South Africa, or they may have wished to gain data on the psychometric characteristics of the instrument to assess its applicability. However, no research could be located with specific reference to the (structural) validity of this short version of the B5P in South Africa. This arguably limits the use of this handy instrument within the South African context.

---

[1]University of South Africa, takandofirepi@gmail.com
[2]University of South Africa, steynr@unisa.ac.za

Despite their popularity and widespread use in psychology and related fields of study, personality measurement scales, based on the B5P model of personality, have come under fire, particularly in cross-cultural studies (Marsh *et al.*, 2012; Dong and Dumas, 2020). The model's general flaw is that each of its five proposed personality traits is far too broad (Hough, 1992) and that, while the wide-ranging categories provide an adequate summary of personality, additional and specific traits should be added to the model to capture the concept of personality completely. Another frequently levelled criticism is that it was developed from the perspective of Westernised, educated, industrialised, wealthy and democratic societies (WEIRD) (Laajaj, 2019; Doğruyol *et al.*, 2019). It is argued that views formed in such contexts are not directly comparable to those formed in societies with different cultural characteristics (Spector *et al.*, 2015). Additionally, it is cautioned that the items measuring the B5P characteristics may not have the same meaning in non-WEIRD societies as they do in WEIRD cultural settings (Ludeke and Larsen, 2017). Considering this, equivalence cannot be assumed with confidence in cross-cultural studies (Laajaj *et al.*, 2019).

To illustrate the foregoing challenge, Gurven's (2013) study in Bolivia, which tested the B5P model using the 44-item BFI tool, produced results that were inconsistent with those obtained in other Western countries, using the same instrument. Additionally, Lajaaj *et al.* (2019) study of data from face-to-face surveys, conducted in 23 low- and middle-income countries, on commonly used measures of the BP5 personality model, discovered that the construct measures failed to measure the intended personality traits, indicating a low validity of the instrument. This finding casts doubt on the cross-cultural validity of the BP5 scales, prompting a call for additional research on them; particularly in non-WEIRD environments, to resolve the discrepancy. Numerous explanations have been advanced in the literature to explain such cross-cultural discrepancies in measurement non-variance or cross-cultural validity of commonly used personality measures. These include participants' systematic response patterns (Boer *et al.*, 2018), the effect of enumerator interaction (Laajaj *et al.*, 2019) and, finally, participants' low levels of education (Gurven *et al.*, 2013). Taken together, these and other factors distort personality measures, particularly in large-scale studies; thereby, increasing the risk of misinterpretation of personality traits in multi-group studies.

While it is preferable to use well-established, comprehensive, and longer multi-item measurement instruments for B5P-related studies, due to their superior content validity and reliability, time constraints may make this impossible, in some instances (Rammstedt and John, 2007). The use of lengthy instruments may result in participation exhaustion, annoyance and inattentive responses, and there is a world-wide trend to use shorter instruments (Kemper, 2019; Rossi-Ferrario, 2019). Ultimately, the reliability of the results of longer instruments can potentially be compromised (Soto and John, 2017). As a result, shorter versions of instruments, such as Rammstedt and John's (2007) 10-item BFI-10, Donnellan *et al.*'s (2006) 20-item mini-IPIP-FFM, and Gerlitz and Schupp's (2005) 15-item Big Five Personality Inventory (BFI-S) have emerged over the years. Robins *et al.* (2001) attest to the value of brief instruments, particularly in large-scale, multi-group surveys and longitudinal studies. They posit that, in general, brief versions have been found to be just as valid as comprehensive and sophisticated scales. They are also more attractive, due to their lower cost and efficiency. Additionally, the psychometric superiority of longer measures is not always reflected in practice (Burisch, 1997).

Despite the advantages of brief instruments in psychology research, some concerns have been raised about them and about brief B5P measures. To begin with, Langford (2003) observes that longer B5P questionnaires are, on average, more representative of broad personality constructs and have greater measurement superiority when compared to shorter versions. Secondly, Gosling *et al.* (2003) observe that shorter versions of the B5P have inferior psychometric properties to this standard multi-item instrument. Lajaaj *et al.* (2019), who used a brief 15-item instrument to analyse data from 94 751 participants in 23 low- and middle-income countries (non-Western, educated, industrialised, rich and democratic populations), encountered similar instrument validity issues. In the study, commonly used personality questions failed to measure the desired personality traits and demonstrated low validity. The WVS, a multi-country research study that examines people's values and beliefs, temporal evolution, and socio-political implications, incorporated the B5P measures for the first time in its sixth wave (covering the years from 2010 to 2014). Rammstedt and John's (2007) 10-item BFI-10 was used to develop the personality measures. The BFI-10 scale adequately measures the extraversion, agreeableness, openness, conscientiousness and neuroticism dimensions of human personality, as demonstrated by several previous studies (Balgiu, 2018; Guido *et al.*, 2015; Sudzina, 2016).

While the BFI-10 has received widespread support for its adequate measurement properties, Ludeke and Larsen (2017), as well as Simha and Parboteeah (2020), expressed reservations about the credibility of the data elicited from participants in the sixth wave of the WVS. Ludeke and Larsen (2017) noted that, in the sixth wave of the WVS, certain measurement items associated with the same personality trait tended not to correlate as expected (in fact, they correlated negatively), contradicting findings from previous research. Disparities were discovered across all five personality dimensions. In the same study, a factor analysis of the ten items did not produce the five dimensions predicted by the B5P model. Rather, the factor analysis of the BFI-10 scales items produced an unwieldy three-factor solution that was difficult to interpret. Another finding from the study was that BFI-10 measures are unreliable in non-Western countries, among less wealthy and educated populations. As a result, caution should be exercised in future studies that make use of WVS data. These results corroborate Chapman and Elliot's (2019) conclusions, when looking at the relationship between BFI-10 measures and mortality in the General Social Survey (USA). They found that using abbreviated versions of longer personality tests, in cross-cultural studies, produced unusual results that did not replicate those obtained using the original versions.

Literature reveals that the global acceptance of the B5P model is questioned; the use of short versions of psychometric tests are sometimes questioned (Chapman and Elliot, 2019; Kemper *et al*., 2019; and the use of the 10-item measure in the WVS is questioned by some. The literature affirms the need for an analysis of the structural validity of the 10-item measure in the WVS, including in the South African context. This research aims to assess the structural validity of the 10-item measure of the B5P instrument in the South African context, with the intention of advocating for, or against, its use. The following research question guided the study:

- Is the 10-item B5P instrument structurally valid when applied to the South African context?

The purpose is to explore whether factors such as language proficiency (Ægisdóttir, 2008; Abrahams and Mauer, 1999; Király *et al*., 2019; McDonald, 2011) can result in the BFI-10 not being valid within this national context.

Methodology

Cross-sectional secondary data, obtained from the WVS database, was used in this study. The data was collected during the WVS wave 6 survey, where a probability-based sample of 3 531 participants were interviewed across South Africa. The sample comprised mainly women, 1 824 (51.7%), and the mean age of all the participants was 37.72 years, with a standard deviation of 15.675. Worldometer (live data) report that women are a small majority in South Africa (51.35%), while the median age is 27.10 years (Worldometer, 2023). It should, however, be noted that the inclusion criteria for participation in the study was 18 years, and a large portion of South Africans are below this age. This means that the sample was somewhat representative in terms sex and the age of the South African population. Since WEIRD was used to create subgroups in the sample, the 427 white participants were deemed to be more Western than the other groups and were used to represent the W (Western) dimension in WEIRD; 316 participants, with university-level education, with or without a degree, were used to represent the E (Educated) dimension in WEIRD; 436 participants from the higher and upper middle class groups were combined to represent the R (Rich) dimension in WEIRD; and, the participants, totalling 692, who selected responses '9' and '10' to the questionnaire item regarding participants' perceptions of an active democracy in the country were combined (278 and 414), which represented the D (Democratic) dimension in WEIRD. It is acknowledged that "Democratic" should be a consistent within a country, but, given the absence of any other option in the dataset, this variable was included.

Participants completed a pre-validated instrument on the Big Five personality factors and self-reported their demographics (these were used to stratify the data). For the personality variable, participants completed the B5P measures of the WVS (6th Wave). The version used was the 10-item BFI-10, by Rammstedt and John (2007). The participants responded to measures, scored on a 5-point Likert-type scale, ranging from 1=disagree strongly to 5=strongly agree. See table 1 for the items of the questionnaire and the way it was coded in the survey.

Table 1: B5P-traits and how they were measured

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

From table 1, two items, one positively worded and the other negatively, represents each of the five traits of the B5P model. With regard to reliability, Rammstedt and John (2007) report test-retest reliability coefficients of at least 0.7 for each of the Big Five traits. Other researchers also report respectable test-retest coefficients, namely .515-.873 (Carciofo *et al.*, 2016, 66-87) (Rammstedt and Krebs, 2007) and an aggregated alpha coefficient of as low as .55 (Erdle and Rushton, 2011). An alpha coefficient of .55 is hardly acceptable, and was a red light noticed by the authors of this paper.

Considering factorial validity, Rammstedt and John (2007) found the expected five-factor structure theorised by the B5P model. Similar positive findings were reported by Bagliu (2018) and Guido *et al.* (2015), all of whom agreed that the theoretical structure of the BFI-10 is supported by data. However, studies by John *et al.* (2019), and Ludeke and Larsen (2017), did not demonstrate adequate empirical support for the five-factor structure in non-WEIRD contexts.

The following variables were used as direct measures of WEIRD or non-WEIRD:

- V254 - Ethnic group (5-point categorical data): Whites were seen as more representative of the dimension, Western, than black and Indian - as examples of other groups.
- V248 - Highest educational level attained (8-point ordinal data): Those with "Some university level education (without a degree)" and "University level education (with a degree)" were selected as more representative of educated, rather than those who had completed secondary school, or who had lower educational achievements.
- V225 - How often one made use of a PC (4-point categorical data): Those who used their PCs frequently were deemed as more industrialised than those who used it occasionally and never, or those who did not have access to PCs.
- V238 - Social class (subjective) (8-point ordinal data): Those in the higher and upper-middle classes were categorised as rich, compared to those in the lower classes.
- V141 - How democratically is this country being governed today? (10-point interval scale): Though it is reasonable to assume that a country, in this case South Africa, has a certain level of democracy, excluding democracy as a variable will give rise to varying perceptions regarding the level of democracy within a country. As a measure of democracy, those with scores of 9 and 10 were seen as representative of the D (Democratic) in the WEIRD acronym.

All the data on the measurements were extracted from the WVS Wave 6 (Inglehart *et al.*, 2014). The study was approved by the University of South Africa's School of Business Leadership's Research Ethics Review Committee. The data needed to address the research question was freely and readily available online, on the WVS website. Thus, the researchers downloaded and proceeded to analyse it. In accordance with the requirements of the owners of the database, the researchers acknowledged the data's source.

The researchers used IBM SPSS 28 to perform exploratory factor analyses (EFA), to first establish whether the data would fit the B5P model using the entire South African sample, as recommended by De Roover and Vermunt (2019).

The natural fit of the data was then determined, using Kaiser's rule regarding eigenvalues. As a rule, the EFA is followed by a confirmatory factor analysis (CFA), and then a multilevel multi-group confirmatory factor analysis (MGCFA), in a stepwise and hierarchical process, in the test for structural validity (Vandenberg and Lance, 2000; Gouveia *et al*., 2015; Putnick and Bonstein, 2016). Varimax rotation was used. Because the EFA result demonstrated that the data did not fit the total dataset for the B5P model well, the sample was stratified into five smaller samples, using the WEIRD acronym. Following this, the same sequence of statistical tests was performed to assess whether the B5P model would fit these sub-populations. The expectation was that the same number of factors would be extracted and that loadings would be similar across groups (Bialosiewicz *et al*., 2013). Neither EFA nor CFA could confirm configural fit; thus, no further tests were performed.

Results

For the full South African sample, the outcome of the EFA showed a Kaiser-Meyer-Olkin measure of sampling adequacy of 0.872, and the Bartlett's Test of Sphericity approximate Chi-Square was 10557.903 (df=45), which was statistically significant (p<.0001) (N=3 525). When applying Kaiser's criterion of retaining factors with loading, with eigenvalues greater than 1, two factors were retained, explaining 55.64% of the variance. When "forcing" the data into the five-factor solution, the variance observed was 76.60%. Table 2 presents these results.

Table 2: Exploratory factor analyses: Based on eigenvalues larger than 1, and on the theorised structure of five factors (Full South African sample, N=3 525)

No	Item	Factors			Factors				
		1	2		1	2	3	4	5
1E (R)	Reserved	**0.632**	0.128		0.183	0.142	0.127	**0.892**	0.033
2A	Trusting	**0.778**	-0.024		-0.081	0.306	0.280	**0.683**	0.237
3C (R)	Lazy	-0.062	**0.824**		**0.846**	-0.111	0.145	0.018	0.052
4S	Relaxed	**0.683**	0.188		0.104	0.280	**0.846**	0.159	0.048
5O (R)	Few interests	**0.639**	0.340		0.196	0.130	**0.723**	0.221	0.326
6E	Outgoing	0.581	0.364		0.096	0.261	0.250	0.133	**0.825**
7A (R)	Find fault	0.256	**0.764**		**0.627**	0.135	0.044	0.114	0.564
8C	Thorough job	**0.733**	0.073		-0.004	**0.836**	0.189	0.189	0.171
9S (R)	Nervous	0.345	**0.693**		**0.699**	0.479	0.123	0.123	0.063
10O	Imagination	**0.715**	0.253		0.164	**0.626**	0.295	0.271	0.245

Note: In the first column, the items are indicated as numbers, with the measured construct indicated as follows: E – Extraversion, A – Agreeableness; C – Conscientiousness; N – Neuroticism; O – Openness. Reverse coding is indicated by R. The "filled" cells represent the theorised structure of the BFI-10.

An indicator of factorial equivalence is when the same number of factors emerge, as proposed in the theoretical model (Bialosiewicz *et al*., 2013). As shown in table 2, two instead of five factors were extracted when applying Keiser's criterion of retaining factors, with loading with eigenvalues greater than 1. Another indicator of structural validity and measurement invariance is when items associated with the same latent variable load on the same factor (McGovern and Lowe, 2018). To test this premise, the data was forced into a five-factor solution. The anticipation was that the B5P traits would each load on a separate factor. Considering the "pointer loadings", which are the items with the high loading (in this case higher than 0.60, and bolded in table 2), it was found that none of the loadings mirrored the theoretical B5P model. Also absent from the results in table 2 were the expected positive and negative loadings per factor, since each personality trait was measured with two items, one of which was reverse coded. To test for structural validity in a comprehensive manner, a confirmatory factor analyses (CFA) were performed, where five factors were proposed, with two items loading on each factor. The results were as follows: CFI=0.870 (cut off CFI>0.90); RMSEA=0.122 (cut off RMSEA<0.08) (Hu and Bentler, 1999). The results revealed a poor fit.

The full South African dataset seemed not to mirror the theorised model. Given the WEIRD principle, it was assumed that groups that meet this criterion may be better in reflecting the B5P model. The sample was thus stratified using indicators in the WVS, which could be deemed proxies to Western, educated, industrialised, rich, and democratic. When applying the same statistical testing procedure to the full sample, the following results (see table 3) emerged for those from a Western background.

Table 3: Exploratory factor analyses: Based on eigenvalues larger than 1, and on the theorised structure of five factors (Western backgrounds sample, N=427) see table 3.

No	Item	Factors			Factors				
		1	2		1	2	3	4	5
1E (R)	Reserved	0.513	0.215		0.129	0.224	0.154	0.064	**0.912**
2A	Trusting	**0.761**	-0.102		-0.005	**0.849**	0.145	0.086	0.284
3C (R)	Lazy	-0.086	**0.821**		**0.853**	-0.055	-0.048	0.064	-0.043
4S	Relaxed	**0.647**	0.123		0.038	0.478	0.091	**0.765**	-0.066
5O (R)	Few interests	0.471	0.528		0.307	-0.048	0.273	**0.806**	0.171
6E	Outgoing	**0.746**	0.090		0.066	0.356	**0.791**	0.202	-0.064
7A (R)	Find fault	0.112	**0.851**		**0.849**	0.002	0.118	0.137	0.088
8C	Thorough job	**0.774**	-0.188		-0.112	**0.760**	0.346	0.141	0.041
9S (R)	Nervous	0.067	**0.841**		**0.840**	-0.030	0.083	0.092	0.136
10O	Imagination	**0.704**	0.182		0.089	0.153	**0.819**	0.131	0.314

Note: In the first column the items are indicated as numbers with the measured construct indicated as follows: E – Extraversion, A – Agreeableness; C – Conscientiousness; N – Neuroticism; O – Openness. Reverse coding is indicated by R. The "filled" cells represent theorised structure of the BFI-10.

The number of factors extracted, applying the Kaiser's standard, was two, signalling structural fit problems (Bialosiewicz *et al*., 2013). After "forcing" the data into five factors, neither the "pointer loadings" nor the expected positive and negative loadings per factor were present in the Western sample. When testing for structural validity, using confirmatory factor analyses with five factors, the following results were produced: CFI=1 (cut off CFI>0.90) (a CFI=1 describes an over-identified model, which suggests that less than 5-factor model could describe the data better); RMSEA= 0.226 (cut off RMSEA<0.08). The results revealed a poor fit.

It was then proposed that those in South Africa, who are more educated, may resemble the B5P configuration. These results are presented in table 4.

Table 4: Exploratory factor analyses: Based on eigenvalues larger than 1, and on the theorised structure of five factors (Educated sample, N=316). See bottom of table 4.

No	Item	Factors			Factors				
		1	2		1	2	3	4	5
1E (R)	Reserved	0.510	0.291		0.170	0.109	0.207	**0.885**	-0.058
2A	Trusting	**0.726**	0.038		-0.003	0.466	0.059	**0.638**	0.446
3C (R)	Lazy	-0.059	**0.800**		**0.772**	0-.214	-0.043	0.078	0.404
4S	Relaxed	**0.709**	0.102		0.065	0.349	0.568	-0.001	**0.647**
5O (R)	Few interests	**0.671**	0.313		0.203	0.151	**0.852**	0.226	0.144
6E	Outgoing	**0.739**	0.032		0.039	0.596	0.532	0.110	-0.022
7A (R)	Find fault	0.179	**0.812**		**0.802**	0.078	0.147	0.203	-0.056
8C	Thorough job	**0.778**	-0.078		0.012	.874	0.080	0.117	0.254
9S (R)	Nervous	0.199	**0.773**		**0.817**	0.235	0.159	-0.034	-0.084
10O	Imagination	**0.704**	0.226		0.240	**0.609**	0.371	0.272	-0.098

Note: In the first column, the items are indicated as numbers, with the measured construct indicated as follows: E – Extraversion, A – Agreeableness; C – Conscientiousness; N – Neuroticism; O – Openness. Reverse coding is indicated by R. The "filled" cells represent the theorised structure of the BFI-10.

As per the previous analyses, the number of factors extracted, applying Kaiser's rule, was two, signalling structural fit problems. After "forcing" the data into five factors, neither the "pointer loadings" nor the expected positive and negative loadings per factor were present in the educated category. When testing for structural validity, using confirmatory factor analyses with five factors, the following results were produced: CFI= 0.758 (cut off CFI>0.90); RMSEA= 0.172; (cut off RMSEA<0.08). The results revealed a poor fit.

Following the WEIRD principle, attention was next given to the industrialised group. It was expected that those in a more industrialised environment, would structurally be close to the B5P. The results are summarised in table 5.

Table 5: Exploratory factor analyses: Based on Eigenvalues larger than 1, and on the theorised structure of five factors (Industrialised sample, N=485)

No	Item	Factors			Factors				
		1	2		1	2	3	4	5
1E (R)	Reserved	0.507	0.150		0.152	0.054	**0.906**	0.064	0.009
2A	Trusting	**0.770**	-0.038		-0.041	0.437	0.610	0.199	0.262
3C (R)	Lazy	-0.061	**0.839**		**0.854**	0.032	-0.040	0.050	-0.097
4S	Relaxed	**0.670**	0.132		0.095	**.796**	0.006	0.451	0.029
5O (R)	Few interests	0.533	0.444		0.283	0.222	0.049	**0.826**	0.087
6E	Outgoing	**0.659**	0.106		0.101	0.230	0.070	0.175	**0.876**
7A (R)	Find fault	0.110	**0.834**		**0.823**	-0.074	0.065	0.143	0.195
8C	Thorough job	**0.745**	-0.081		-0.004	**0.733**	0.307	-0.050	0.382
9S (R)	Nervous	0.199	**0.800**		**0.804**	0.118	0.154	0.136	0.059
10O	Imagination	**0.699**	0.170		0.048	0.070	0.382	**0.611**	0.445

Note: In the first column, the items are indicated as numbers, with the measured construct indicated as follows: E – Extraversion, A – Agreeableness; C – Conscientiousness; N – Neuroticism; O – Openness. Reverse coding is indicated by R. The "filled" cells represents the theorised structure of the BFI-10.

The number of factors extracted, applying Kaiser's criterion, was again two, signalling structural fit problems. When forcing the data into five factors, neither the "pointer loadings" nor the expected positive and negative loadings per factor, were present in the industrialised category. When testing for structural validity, using affirmative factor analyses with five factors, the following results were produced: CFI=0.764 (cut-off CFI>0.90); RMSEA=0.166 (cut off RMSEA<0.08). The results revealed a poor fit.

Next, the rich category was explored as a potential group that reflects the B5P taxonomy.

Table 6: Exploratory factor analyses: Based on eigenvalues larger than 1, and on the theorised structure of five factors (Rich sample, N=436)

No	Item	Factors			Factors				
		1	2		1	2	3	4	5
1E (R)	Reserved	**0.633**	0.252		0.162	0.183	0.204	**0.896**	0.099
2A	Trusting	**0.832**	-0.003		0.011	0.341	**0.685**	0.459	0.059
3C (R)	Lazy	-0.040	**0.821**		**0.890**	0.147	-0.009	-0.050	-0.149
4S	Relaxed	**0.758**	0.076		0.104	**0.814**	0.398	0.122	0.001
5O (R)	Few interests	**0.727**	0.259		0.155	**.767**	0.100	0.257	0.389
6E	Outgoing	**0.717**	0.217		0.094	0.392	0.372	0.089	**0.699**
7A (R)	Find fault	0.264	**0.776**		**0.672**	-0.052	0.104	0.214	0.535
8C	Thorough job	**0.780**	0.029		0.065	0.190	**0.881**	0.110	0.244
9S (R)	Nervous	0.221	**0.818**		**0.783**	0.105	0.073	0.193	0.234
10O	Imagination	**0.771**	0.175		0.097	0.392	0.407	0.435	0.354

Note: In the first column, the items are indicated numbers with the measured construct indicated as follows: E – Extraversion, A – Agreeableness; C – Conscientiousness; N – Neuroticism; O – Openness. Reverse coding is indicated by R. The "filled" cells represent the theorised structure of the BFI-10.

As summarised in table 6, the number of factors extracted, using Kaiser's criterion, numbered two, signalling structural fit problems. When the data was forced into five factors, neither the "pointer loadings" nor the expected positive and negative loadings per factor were present in the rich. When testing for structural validity, using affirmative factor analyses with five factors, the following results were produced: CFI=0.811 (cut off CFI>0.90); RMSEA=0.164 (cut off RMSEA<0.08). The results revealed a poor fit.

Lastly, the idea was explored that those who believe they live within a democratic dispensation, may share the personality characteristics portrayed in the B5P theory of personality.

Table 7: Exploratory factor analyses: Based on eigenvalues larger than 1, and on the theorised structure of five factors (Democratic sample, N=692)

No	Item	Factors			Factors				
		1	2		1	2	3	4	5
1E (R)	Reserved	**0.666**	0.150		0.205	0.126	**0.886**	0.036	0.125
2A	Trusting	**0.806**	0.011		0.011	0.341	**0.755**	0.222	0.169
3C (R)	Lazy	-0.013	**0.826**		**0.851**	0.076	0.060	-0.084	-0.027
4S	Relaxed	**0.707**	0.152		0.079	**0.802**	0.213	0.194	0.119
5O (R)	Few interests	**0.717**	0.251		0.202	**0.712**	0.271	-0.046	0.423
6E	Outgoing	0.488	**0.527**		0.451	0.549	0.112	0.474	-0.102
7A (R)	Find fault	0.238	**0.820**		**0.795**	0.227	0.080	0.146	0.138
8C	Thorough job	0.575	0.204		0.084	0.146	0.147	**0.880**	0.233
9S (R)	Nervous	0.321	**0.743**		**0.711**	0.024	0.185	0.284	0.346
10O	Imagination	**0.698**	0.257		0.177	0.260	0.239	0.260	**0.807**

Note: In the first column the items are indicated as numbers with the measured construct indicated as follows: E – Extraversion, A – Agreeableness; C – Conscientiousness; N – Neuroticism; O – Openness. Reverse coding is indicated by R. The "filled" cells represent the theorised structure of the BFI-10.

The number of factors extracted, based on Kaiser's criterion, was two (see table 7), signalling structural fit problems (Bialosiewicz *et al.*, 2013). When the data was "forced" into five factors, neither the "pointer loadings" nor the expected positive and negative loadings per factor, were present in the "democratic" category. The following results were obtained from the confirmatory factor analyses with five factors: CFI=0.861 (cut off CFI> 0.90); RMSEA=0.134 (cut off RMSEA<0.08).

Discussion

The literature is clear that the B5P model is still well respected, but with several concerns, particularly with the shorter BFI-10 version. These challenges included it having been established from the standpoint of Westernised cultural settings and its interpretation, thus, not directly comparable to those formed in populations with distinct cultural traits. Also, the BFI-10, which was used to measure personality traits in the WVS wave 6, was seen to have inferior psychometric properties and, in some instances, failed to capture the Big Five personality traits. The limitations of the research instrument have thus necessitated this study. The integrity of the data collected via the WVS was not questioned, as it was collected by a well-established and reputable service provider. The sample, as per the demographic data, was representative of the South African population.

The results revealed that the B5P model did not fit the data for the total population. Longer versions seem to work well (Abrahams and Mauer, 1999; McDonald, 2011; Taylor and De Bruin, 2006). When EFA was applied to the data, the results revealed that the number of factors extracted, applying Keiser's criterion was 2, and not the expected 5, as per the B5P model; thus, signalling structural fit problems. Even when the data was "forced" in EFA into a five-factor model, none of the loadings mirrored the theoretical B5P model. Also absent from these EFA results were the expected positive and negative loadings per factor. When testing for CFA, the results echoed those found with the EFA, with none of the indicators fitting the theorised model. These results are contrary to previous research by Guido *et al.* (2015) and Sudzina (2016), where it was found that the BFI-10 scale properly measured the extraversion, agreeableness, openness, conscientiousness and neuroticism characteristics of human personality. However, the results affirm the research by Ludeke and Larsen (2017), as well as Simha and Parboteeah (2020), where the empirical data in the sixth wave of the WVS was found not to adequately reflect the B5P model.

It was envisaged that creating subgroups (of the South African sample) informed by the WEIRD principle (Westernised, educated, industrialised, wealthy, and democratic) could yield results more in line with expectations. These mirrored the total population results, and this was demonstrated with the EFA and confirmed with the CFA results. These affirm some studies done across cultures and nations (Chapman and Elliot, 2019; Ludeke and Larsen, 2017), where it was found that the BFI-10 research instrument was not structurally valid and the measure demonstrated equivalence. This research is also novel in that it used the WEIRD principle within a specific population and provided insight into the

applicability of the BFI-10 research to population enclaves with WEIRD attributes but situated in a non-Western country. This can be seen as a unique methodological contribution to the study.

Our study has a limitation in that it relies only on secondary sources, and we had no influence over the data collecting process. However, multiple research publications, in credible journals, that have used the WVS datasets, corroborate the data's legitimacy. Another issue is that the research instrument employed was written in English. However, several of the participants were not native English speakers, which might have hampered their interpretation of the questionnaire items. Future researchers are encouraged to assess the structural validity of all instruments "imported" to their cultures, before using them. This may not only be a concern in South Africa, but also in other countries, which are not WEIRD. Specifically, regarding the B5P assessments and the BFI-10, the functionality of the theory, across cultures, may require further investigation, as well as the utility of short versions of instruments, across cultures and languages. It may be difficult to capture meaning in a few words or sentences across the divide of culture and language (Djiwandono, 2006).

Conclusions

It was concluded that the B5P model was not replicable in the short format in South Africa, or within any of the subgroups created using the WEIRD principle. The analyses revealed that the BFI-10 factor structures are not replicated in the South African and South African sub-samples and that the BFI-10 lacks measurement invariance within the South African population. The assumption that well-established instruments are valid in settings different to the ones they were initially developed in, should be questioned, and such instruments should not be used, unless thoroughly tested. This study exposes the extent of measurement non-invariance when using an instrument in a foreign setting, and shows that, even in equitable subgroups, invariance does not occur. Those working with foreign individuals or conducting cross-cultural research should be particularly aware of these threats to validity. Though only some research, specific to measurement invariance and the WVS, is available, the decision to include the BFI-10 in the 7th wave of the survey seems wise.

References

Abrahams, F. and Mauer, K. F. 1999. The Comparability of the Constructs of the 16PF in the South African Context. *Journal of Industrial Psychology*, 25(1): 53–59.

Ægisdóttir, S., Gerstein, L. H. and Çinarbaş, D. C. 2008. Methodological Issues in Cross-Cultural Counselling Research: Equivalence, Bias, and Translations. *The Counselling Psychologist,* 36(2): 188-219.

Allport, G. W. and Odbert, H. S. 1936. Trait-names: A Psycho-Lexical Study. *Psychological Monographs*, 47(1): 1–171.

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.

Bialosiewicz, S., Murphy, K. and Berry, T. 2013. *An Introduction to Measurement Invariance Testing: Resource Packet for Participants*. Washington, DC: American Evaluation Association.

Boer, D., Hanke, K. and He, J. 2018. On Detecting Systematic Measurement Error in Cross-Cultural Research: A Review and Critical Reflection on Equivalence and Invariance Tests. *Journal of Cross-Cultural Psychology*, 49(5): 713-734.

Burisch, M. 1997. Test Length and Validity Revisited. *European Journal of Personality*, 11(4): 303-315.

Carciofo, R., Yang, J., Song, N., Du, F. and Zhang, K. 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): 1-26.

Cattell, R. B. 1950. *Personality: A Systematic Theoretical and Factual Study*. New York, NY: McGraw Hill.

Chapman, B. P. and Elliot, A. J. 2019. Brief report: How Short is Too Short? An Ultra-Brief Measure of the Big-Five Personality Domains Implicates "Agreeableness" as a Risk for All-Cause Mortality. *Journal of Health Psychology*, 24(11): 1568-1573.

Costa, P. T. and McCrae, R. R. 1992. Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment*, 4(1): 5-13.

De Roover, K. and Vermunt, J. K. 2019. On the Exploratory Road to Unravelling Factor Loading Non-Invariance: A New Multigroup Rotation Approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6): 905-923.

Djiwandono, P. I. 2006. Cultural Bias in Language Testing. *TEFLIN Journal,* 17(1): 81-88.

Doğruyol, B., Alper, S. and Yilmaz, O. 2019. The Five-Factor Model of the Moral Foundations Theory is Stable Across WEIRD and Non-WEIRD Cultures. *Personality and Individual Differences*, 151: 1-6.

Dong, Y. and 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: 1-23.

Donnellan, M. B., Oswald, F. L., Baird, B. M. and 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.

Erdle, S. and 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.

Fiske, D. W. 1949. Consistency of the Factorial Structures of Personality Ratings from Different Sources. *The Journal of Abnormal and Social Psychology*, 44(3): 329-344.

Gerlitz, J. Y. and Schupp, J. 2005. On the Survey of the Big Five-Based Personality Characteristics in the SOEP. *DIW Research Notes*, 4: 1-36.

Goldberg, L. R. 1981. Language and Individual Differences: The Search for Universals in Personality Lexicons. In: Wheeler, L. ed. *Review of Personality and Social Psychology*, 2(1): 141-165.

Gosling, S. D., Rentfrow, P. J. and Swann Jr, W. B. 2003. A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality*, 37(6): 504-528.

Gouveia, V. V., Vione, K. C., Milfont, T. L., and Fischer, R. 2015. Patterns of Value Change During the Life Span: Some Evidence from a Functional Approach to Values. *Personality and Social Psychology Bulletin*, 41(9): 1276-1290.

Guido, G., Peluso, A. M., Capestro, M. and 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: 135-140.

Gurven, M., Von Rueden, C., Massenkoff, M., Kaplan, H. and Lero Vie, M. 2013. How Universal is the Big Five? Testing the Five-Factor Model of Personality Variation among Forager–Farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, 104(2): 354-370.

Hair, J., Black, W., Babin, B. and Anderson, R. 2010. *Multivariate Data Analysis*. 7th edition. Upper Saddle River, NJ: Prentice-Hall, Inc.

Hill, C., Adams, B. G., De Bruin, G. P., Nel, J. A., Van de Vijver, F. J., Valchev, V. H. and Meiring, D. 2013. Developing and Testing Items for the South African Personality Inventory (SAPI). *SA Journal of Industrial Psychology*, 39(1): 1-13.

Hough, L. M. 1992. The Big Five Personality Variables-Construct Confusion: Description Versus Prediction. *Human Performance*, 5(1-2): 139-155.

Hu, L. T. and 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.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin and B. Puranen, B. eds. 2014. World Values Survey: Round Six - Country-Pooled Datafile Version. Available: www.worldvaluessurvey.org/WVSDocumentationWV6.jsp (Assessed 22 February 2023)

John, O. P. 2021. History, Measurement, and Conceptual Elaboration of the Big-Five Trait Taxonomy: The Paradigm Matures. In: John, O. P. and Robins, R. W. eds. *Handbook of Personality: Theory and Research.* Washington DC: The Guilford Press, 35–82.

Kemper, C. J., Trapp, S., Kathmann, N., Samuel, D. B. and Ziegler, M. 2019. Short Versus Long Scales in Clinical Assessment: Exploring the Trade-off between Resources Saved and Psychometric Quality Lost Using Two Measures of Obsessive–Compulsive Symptoms. *Assessment,* 26(5): 767-782.

Király, O., Bőthe, B., Ramos-Diaz, J., Rahimi-Movaghar, A., Lukavska, K., Hrabec, O. and Demetrovics, Z. 2019. Ten-item Internet Gaming Disorder Test (IGDT-10): Measurement Invariance and Cross-cultural Validation Across Seven Language-based Samples. *Psychology of Addictive Behaviors,* 33(1): 91-103.

Kunnel John, R., Xavier, B., Waldmeier, A., Meyer, A. and Gaab, J. 2019. Psychometric Evaluation of the BFI-10 and the NEO-FFI-3 in Indian Adolescents. *Frontiers in Psychology*, 10: 1-10.

Laajaj, R., Macours, K., Hernandez, D. A. P., Arias, O., Gosling, S. D., Potter, J. and Vakis, R. 2019. Challenges to Capture the Big Five Personality Traits in Non-WEIRD Populations. *Science Advances*, 5(7): 1-13.

Langford, P. H. 2003. A One-minute Measure of the Big Five? Evaluating and Abridging Shafer's (1999a) Big Five Markers. *Personality and Individual Differences*, 35(5): 1127-1140.

Ludeke, S. G. and Larsen, E. G. 2017. Problems with the Big Five Assessment in the World Values Survey. *Personality and Individual Differences*, 112: 103-105.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U. and Nagengast, B. 2010. A New Look at the Big Five Factor Structure through Exploratory Structural Equation Modeling. *Psychological Assessment*, 22(3): 471-491.

McCrae, R. R. and Costa, P. T. 1987. Validation of the Five-Factor Model of Personality across Instruments and Observers. *Journal of Personality and Social Psychology*, 52(1): 81-90.

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.

Meisenberg, G. 2015. Do We Have Valid Country-level Measures of Personality?. *Mankind Quarterly*, 55(4): 360-382.

Minkov, M., van de Vijver, F. J. and Schachner, M. 2019. A Test of a New Short Big-Five Tool in Large Probabilistic Samples from 19 Countries. *Personality and Individual Differences*, 151: 1-6.

Norman, W. T. 1967. *2800 Personality Trait Descriptors: Normative Operating Characteristics for a University Population*. Michigan: University of Michigan.

Nunes, A., Limpo, T., Lima, C. F. and Castro, S. L. 2018. Short Scales for the Assessment of Personality Traits: Development and Validation of the Portuguese Ten-Item Personality Inventory (TIPI). *Frontiers in Psychology*, 9: 1-5.

Putnick, D. L. and Bornstein, M. H. 2016. Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review*, 41: 71-90.

Rammstedt, B. and 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.

Rammstedt, B. and 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.

Rammstedt, B., Danner, D., Soto, C. J. and John, O. P. 2018. Validation of the Short and Extra-short Forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment,* 36(1): 149–161.

Robins, R. W., Hendin, H. M. and Trzesniewski, K. H. 2001. Measuring Global Self-esteem: Construct Validation of a Single-item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin,* 27(2): 151-161.

Rossi-Ferrario, S., Panzeri, A., Anselmi, P. and Vidotto, G. 2019. Development and Psychometric Properties of a Short Form of the Illness Denial Questionnaire. *Psychology Research and Behavior Management*, 12: 727-739.

Schmitt, D. P., Realo, A., Voracek, M. and Allik, J. 2008. Why Can't a Man be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *Journal of Personality and Social Psychology*, 94(1): 168-182.

Simha, A. and 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.

Smith, G. M. 1967. Usefulness of Peer Ratings of Personality in Educational Research. *Educational and Psychological Measurement*, 27(4): 967-984.

Soto, C. J. and 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.

Spector, P. E., Liu, C. and Sanchez, J. I. 2015. Methodological and Substantive Issues in Conducting Multinational and Cross-Cultural Research. *Annual Review of Organizational Psychology and Organizational Behavior,* 2(1): 101-131.

Sudzina, F. 2016. Do Gender and Personality Traits (BFI-10) Influence Trust? A Replication. *Central European Journal of Management*, 3(1): 47-53.

Taylor, N. and De Bruin, G. P. 2006. *Basic Traits Inventory*. Johannesburg: Jopie van Rooyen.

Van de Vijver, F. J. and Poortinga, Y. H. 2002. Structural Equivalence in Multilevel Research. *Journal of Cross-Cultural Psychology*, 33(2): 141-156.

Vandenberg, R. J. and 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): 4-70.

Worldometer. 2023. South Africa Population. Available: https://www.worldometers.info/world-population/south-africa-population/ (Accessed 22 February 2023).