




An Examination of the Spanish Translation of the 50-item International Personality Item Pool Big-five Inventory in a Spanish Speaking Peruvian Sample

David J. Hughes¹ , Daniel Pizarro de Olazabal², Ioannis K. Kratsiotis¹ , Ricardo Twumasi²  and Tom Booth³

¹ The University of Manchester (UK)

² King's College London (UK)

³ The University of Edinburgh (UK)

Abstract. The International Personality Item Pool (IPIP) five-factor model inventories are widely used for personality research and have been translated into multiple languages. However, the extent of the psychometric assessment of translated scales is variable, often minimal. The lack of psychometric scrutiny is particularly problematic because translation is an inherently complex process. Here, we present a structural analysis of one Spanish translation of the 50-item IPIP five-factor inventory in a sample of Peruvian, non-university educated, working adults ($n = 778$). A global confirmatory factor analytic (CFA) model of the a priori five factors failed to fit. So too did single factor models for four of the five factors, the exception being Neuroticism. Fit was improved via use of an exploratory structural equation measurement model, but the resultant solution showed very poor theoretical coherence. So, we explored the data for systematic measurement artefacts and sought to model them to improve the psychometric properties of the scale. Specifically, the pattern of factor loadings suggested that the lack of coherence might be due to the effects of the valence of item wording (i.e., positively or negatively worded items). CFA models including five substantive factors and a series of method factors modelling shared covariance based on item wording, improved fit and coherence. This investigation suggests that unless method factors are explicitly modelled the tested Spanish translation may not be suitable for use in certain Spanish-speaking countries or samples composed of non-university educated participants. More broadly, the study has implications for many translated scales, especially when used without thorough psychometric evaluation.

Received 18 January 2019; Revised 6 April 2020; Accepted 07 April 2020

Keywords: IPIP, FFM, psychometric, method artefacts, Spanish translation

Assessments of personality most commonly use tools developed from within a Big Five or Five-Factor Model approach and assess the broad domains of Neuroticism, Extraversion, Openness/Intellect, Agreeableness, and Conscientiousness. The five factors assessed by these tools do differ but are generally regarded to refer to the same broad psychological constructs (cf. Block, 1995; Digman, 1990). Five-factor approaches remain the dominant framework for trait description, and the associated tools are the most widely applied across multiple fields of study. One of the most important elements of supporting evidence in favor of five-factor models is that they have shown a degree of cross-cultural stability (McCrae & Costa, 1997; McCrae et al., 2005), suggesting that they represent something of a universal taxonomy of broad personality factors.

As a result, five factor assessment tools have been translated into an array of languages, often using items

from The International Personality Item Pool (IPIP; Goldberg, 1999) as a starting point. The IPIP provides open access personality scales designed as proxies for many constructs including proprietary five factor inventories. Building on the benefits of free use, which has accelerated research beyond what would be possible using only proprietary tools, the IPIP has been used in a range of different cultures and translated to over 25 different languages (Goldberg, 1999; Goldberg et al., 2006).

However, translated IPIP scales are typically subject to reduced psychometric scrutiny compared to their English-language counterparts (Mlačić & Goldberg, 2007). Thus, it can be difficult for researchers to choose an appropriate translation for their study, especially when multiple versions exist. The lack of psychometric scrutiny is particularly problematic because translation

Correspondence concerning this article should be addressed to David J. Hughes. The University of Manchester. Alliance Manchester Business School. M13 9PL Manchester (UK).

E-mail: david.hughes-4@manchester.ac.uk

How to cite this article:

Hughes, David J., de Olazabal, D. P., Kratsiotis, Ioannis K., Twumasi, R., & Booth, T. (2020). An examination of the spanish translation of the 50-item international personality item pool big-five inventory in a spanish speaking peruvian sample. *The Spanish Journal of Psychology*, 23. e18. Doi:10.1017/SJP.2020.11

is an inherently complex process. Translators must ensure that translated items accurately assess the same construct (i.e., respondents draw upon the same class of memories and experiences when responding to the items; see Hughes, 2018) whilst contending with unique cultural, environmental, and grammatical differences. However, if translated items do not operate in an equivalent manner (i.e., words or phrases have different connotations, leading participants to draw upon different memories/processes; Boroditsky, 2001) then item responses are no longer equivalent and any scale score created from them changes in meaning. Often this lack of equivalence is reflected in the structure of the item responses (i.e., the factor structure will not replicate, Hughes, 2018).

Accordingly, we sought to investigate the psychometric properties of a Spanish translation of the 50 item IPIP Big-five inventory (henceforth referred to as the IPIP-50-S) within a Spanish speaking Peruvian sample. To our knowledge only two studies have previously investigated the psychometric properties of the scale: One within a sample of Argentinian teenagers (Cupani, 2009) and one within a mixed but predominantly student Argentinian sample (Gross et al., 2012). Both studies noted some problems concerning the factor structure including low loading items ($< .4$), large numbers of non-trivial cross-loadings, and some items having their largest loading on their non-target factor (Cupani, 2009; Gross et al., 2012). However, neither study was able to fully diagnose the causes of problems. The generalizability of these findings may also be somewhat limited because the samples consisted predominantly of Argentinian students. Therefore, further investigation of the performance of the translated measure in other Spanish speaking samples is of interest.

Accordingly, the major focus of the current study is on the identification of the appropriate factor structure for the translated items. Here we will consider both a priori confirmatory factor models, for a complete five-factor model and for each domain individually, as well as exploratory models where there is evidence of misfit. Specifically, a number of studies show that CFA models of personality data produce inadequate model fit according to conventional criteria (Booth & Hughes, 2014; Hopwood & Donnellan, 2010). This, it has been argued, is due to the complexity of personality items for which the responses may be influenced by multiple traits, and thus the independent cluster modelling assumption in typical CFA applications may be too restrictive (Marsh et al., 2010). As such, we will apply exploratory structural equation modelling (ESEM) in the presence of misfit to identify the sources of misfit and the alternative optimal factor structure. Typically, ESEM approaches improve personality model fit but

they remain some way from being adequately fitting models (Booth & Hughes, 2014).

Model misfit typically arises due to unmodeled sources of shared variation among indicators. Other possible sources of such variation in personality assessments stem from measurement errors commonly referred to as response biases and measurement artefacts (Podsakoff et al., 2003; Podsakoff et al., 2012). Thus, the third element of our analysis will be to explore the existence of such measurement artefacts. Previous research exploring scale translations has noted country-specific effects of extreme, acquiescent, and socially desirable responding (Diamantopoulos et al., 2006; Johnson et al., 2005). Indeed, previous research examining English-Spanish translations has suggested that the two most crucial item characteristics that influence cross-language equivalence are item complexity (length and language difficulty) and social desirability (Valentine, 2013). Thus, if CFA and ESEM models do fail to fit, we will explore the data for evidence of systematic measurement artefacts and seek to model them to improve the psychometric properties of the scale.

Method

Participants

Participants were 778 employees from fourteen stores of a supermarket retail company in Lima, Peru (379 male; 369 female; 30 missing values). Participants were selected at random from a list of all employees at each store who had worked at the company for over one month. Between 33 and 97 participants were collected from each store. All participants were Peruvian, aged from 18 to 60 years old ($M = 24.67$; $SD = 6.38$), and employed as customer service assistants. Participants' job tenure ranged from 1 to 228 months ($M = 16$; $SD = 22.37$). All participants had completed secondary education (from 13 to 17 years) in Peruvian state schools.

Procedure

Permission to recruit participants was provided by the Human Resources department of the company who also assisted with data collection. To ensure consistency across test administrators, a member of the research team provided Human Resource assistants with instructions on the delivery of the survey. Questionnaires were completed in paper-pencil format, and later transferred to an electronic database by the research team. Testing was conducted in the workplace and in order to maintain the confidentiality/anonymity of participants, no identifying information was taken; instead all participants received a unique identifier meaning that data was fully anonymous.

Ethics

The study was given ethical approval by the Psychology Research Ethics Committee, Department of Psychology, University of Edinburgh. Surveys were completely anonymized at point of input into the electronic database. The original surveys were not shared with the hosting institution.

Measures

The survey consisted of two sections, a series of questions on co-worker satisfaction, and a personality inventory. For the purpose of the current study, only the personality items are analyzed.

The IPIP-50-S was used to measure the Big Five personality domains of Neuroticism, Extraversion, Intellect, Agreeableness, and Conscientiousness. Participants had to rate themselves on a 5-point Likert-type scale ranging from 1 (*very inaccurate*) to 5 (*very accurate*), according to how accurately each statement describes them. The IPIP-5-S comprised 50 items, 10 per personality domain. Examples items are “Am interested in people” (agreeableness), “Am the life of the party” (extraversion), “Pay attention to details” (conscientiousness), “Am relaxed most of the time” (neuroticism) and “Have a vivid imagination” (intellect). All items, in English, and their mean and standard deviation are reported in Table 1. The specific translation used is available online¹ and also in Supplementary Material.

Analysis Strategy

Estimation and Evaluation: All models were estimated using weighted-least-squares means and variances (WLSMV) estimation in Mplus 7.4 (Muthén & Muthén, 1998–2017). Code for all analyses is available². Models were evaluated based on the magnitude of the factor loadings and on model fit. We followed typically applied criteria whereby CFI and TLI ranging from .90 to > .95 and RMSEA < .06 were deemed indicative of good model fit (Hu & Bentler, 1999; Schermelleh-Engel et al., 2003). As we implement WLSMV estimation in Mplus, we also report WRMR, however it is noted that to date, little simulation evidence is available to suggest indicative cut-off values.

Measurement models: We initially fit a confirmatory factor model for an independent cluster five-factor model, allowing each of the trait factors to correlate. The model was identified by fixing the first factor loading on each latent factor to 1.0. If the model failed to reach minimum standards for model fit, as is common in the extant literature, we planned to apply three sets of

models to identify misfit. First, single factor CFA models for each trait in order to identify possible correlated residuals. Second, an exploratory structural equation model (ESEM) with five correlated factors, modelling item cross-loadings and allowing for structural complexity. Third, we would consider the possibility of method factors in the data, and estimate five factor CFA models with latent factors included to account for variance due to different artefacts (see Podsakoff et al., 2012 for discussion of different approaches). Specifically, we estimated models including a general acquiescence factor (Figure 1, Panel A), positive and negative valence factors (Figure 1, Panel B), and finally a model with all three potential sources of method effect included (Figure 1, Panel C).

Results

Measurement models for the IPIP-50-S

The five-factor independent clusters CFA model converged, but the factor covariance matrix was non-positive definite due to factor correlations greater than 1.0. Given this, we considered this solution inappropriate.

Next, we examined each of the five factors independently. Four of the five single-factor CFA solutions showed poor fit, the one exception being Neuroticism (see Supplementary Tables S2 for model fit). Within these models, 12 of the 50 items did not load greater than .30 on their hypothesized factor, indicating that the items do not cohere as expected or produce a psychometrically strong scale. Perhaps more importantly, Neuroticism and Extraversion items, despite containing both positively (e.g., Don't mind being the center of attention) and negatively (e.g., Don't like to draw attention to myself) worded items, all loaded positively onto the single factor (see Supplementary Tables S3 to S7 for factor loadings).

To explore the data further, we first fit a five factor ESEM. Model fit for the ESEM model was reasonable, $\chi^2 = 2027,881(985)$, $p < .001$; CFI = .95; TLI = .94; RMSEA = .037; WRMR 1.029. The full factor loading matrix for the ESEM solution is provided in Table 2.

Consideration of the item loadings in Table 2 suggested that the solution was not conceptually similar to the a priori five-factor model. Factor 2 contained salient loadings (> .30) for a majority of the negatively worded items across traits, including loadings from eight of the 10 Neuroticism items. Similarly, Factor 3 contained salient loadings from all positively worded items from Conscientiousness and Intellect, and four positively worded items from both Extraversion and Agreeableness. Thus, these two factors seemed identifiable as method factors defined by item valence. Of the

¹<https://ipip.ori.org/SpanishBig-FiveFactorMarkers.htm>

²<https://osf.io/6dxbm/>

Table 1. Item Descriptive Statistics for the IPIP-S

Items	N	Mean	SD
Get stressed out easily (+N)	776	3.28	1.03
Am relaxed most of the time (-N)	777	3.04	1.17
Worry about things (+N)	774	3.65	1.08
Seldom feel blue (-N)	777	3.53	1.01
Am easily disturbed (+N)	775	2.14	1.21
Get upset easily (+N)	776	3.70	0.92
Change my mood a lot (+N)	778	3.03	1.11
Have frequent mood swings (+N)	776	2.31	1.13
Get irritated easily (+N)	768	2.62	1.10
Often feel blue (+N)	777	3.78	0.87
Am the life of the party (+E)	776	3.36	0.98
Don't talk a lot (-E)	776	2.95	1.07
Feel comfortable around people (+E)	774	3.28	1.19
Keep in the background (-E)	774	3.70	1.10
Start conversations (+E)	775	3.93	0.81
Have little to say (-E)	771	2.49	1.18
Talk to a lot of different people at parties (+E)	775	2.92	1.14
Don't like to draw attention to myself (-E)	774	2.50	1.14
Don't mind being the center of attention (+E)	774	3.70	0.85
Am quiet around strangers (-E)	774	2.40	1.23
Feel little concern for others (-A)	773	2.98	1.14
Am interested in people (+A)	767	3.11	1.16
Insult people (-A)	778	3.93	0.90
Sympathize with others' feelings (+A)	775	2.26	1.22
Am not interested in other people's problems (-A)	771	2.20	1.24
Have a soft heart (+A)	776	4.13	0.89
Am not really interested in others (-A)	775	2.91	1.16
Take time out for others (+A)	775	2.86	1.14
Feel others' emotions (+A)	771	2.74	1.08
Make people feel at ease (+A)	777	3.90	0.82
Have a rich vocabulary (+I)	772	3.42	1.03
Have difficulty understanding abstract ideas (-I)	776	3.18	1.11
Have a vivid imagination (+I)	772	2.49	1.22
Am not interested in abstract ideas (-I)	777	3.98	0.84
Have excellent ideas (+I)	777	3.90	0.88
Do not have a good imagination (-I)	772	3.75	0.87
Am quick to understand things (+I)	776	2.66	1.18
Use difficult words (+I)	777	2.31	1.18
Spend time reflecting on things (+I)	775	3.73	0.85
Am full of ideas (+I)	776	3.92	0.82
Am always prepared (+C)	775	3.88	0.92
Leave my belongings around (-C)	774	3.29	1.12
Pay attention to details (+C)	775	1.76	1.12
Make a mess of things (-C)	776	3.52	0.94
Get chores done right away (+C)	775	3.95	0.89
Often forget to put things back in their proper place (-C)	773	2.09	1.20
Like order (+C)	775	3.85	0.90
Shirk my duties (-C)	774	2.52	1.17
Follow a schedule (+C)	775	3.91	0.84
Am exacting in my work (+C)	776	2.76	1.05

Note. N = Neuroticism; E = Extraversion, O = Openness-to-experience; A = Agreeableness; C = Conscientiousness.

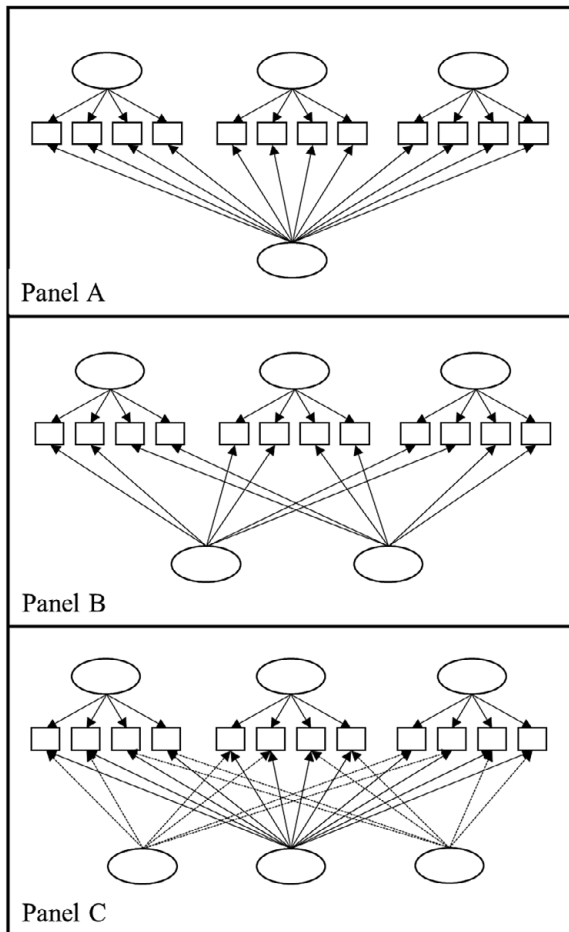


Figure 1. Diagrammatic Representation of Models Estimated to Investigate Method Artefacts. *Note.* A general acquiescence factor (Panel A), positive and negative valence factors (Panel B), model with all three potential sources of method effect included (Panel C). In all Panels, example personality factors are depicted above the factor indicators and method factors depicted below the factor indicators.

remaining factors, and based on the items with salient loadings, Factors 1 and 4 could be labelled Neuroticism and Agreeableness respectively. Factor 5 could not be readily labelled. To explore the data further, we also estimated ESEM models using CF-Parsimax Oblique, Oblimin Oblique, and Target rotation. The pattern of the results did not change. We have included the pattern matrices from these additional analyses in supplementary materials, Tables S10–S12.

Method Artefacts in the IPIP-50-S

Based on the indications from both the extant literature and the pattern of item loadings in Table 2, we explicitly modelled a series of method factors. Table 3 contains the model fit indices for models including positive and negative valence method factors (M1), a general

acquiescence method factor (M2), and a model with positive, negative, and general acquiescence factors (M3). In all models, factor variances were fixed at 1 to identify to models, and WLSMV estimation was used.

Model fit across all models was acceptable to good. Unsurprisingly, the model containing all three method artefact latent variables showed the best model fit. Fit of this model was comparable to the ESEM model but was more parsimonious. In addition, the factor loadings from all models were more consistent with what would have been expected a priori. In M1 (see supplementary Table S8 for factor loadings), positively and negatively worded items loaded consistently on their respective valence factors. However, eleven items had loadings below .30 on their substantive factors. A similar pattern was true for model M2. All items had positive loadings on the general method factor and appropriate directionality of loading on their substantive factors. Again, the same eleven items failed to load on their a priori substantive factors above .30. However, in both M1 and M2, the factor correlations were much greater than would be expected, with absolute r ranging from .48 to .85 for M1, and .50 to .86 for M2.

Table 4 shows the full factor loading matrix for M3. Two primary observations can be made from Table 4. First, whilst the inter-factor correlations for M3 were in line with most five factor research in magnitude ($+/- .10$ to $.42$), the direction these correlations are not as would be anticipated. Consideration of the direction of the factor loadings, and thus the definition of the factors, does not clarify the pattern of correlations. Second, a majority of the variance in the items is typically accounted for by the methodological factors rather than their substantive factor.

Discussion

Our goal was to evaluate the psychometric properties of the IPIP-50-S within a sample of Peruvian customer service employees. No previous studies had examined this scale in Peru or in a fully non-student sample. As expected, a CFA of the a priori model did not fit the data and with the exception of Neuroticism, the factors did not fit even when modeled independently. An ESEM model did improve the overall fit but the solution remained sub-optimal with numerous large cross-loadings and some items failing to load on the expected factor. These results are consistent with past research on five factor inventories (Booth & Hughes, 2014) and suggest that the IPIP-50-S is not well suited to research with Peruvian adults with a non-university level of education.

Further exploration of the possible sources of misfit were elucidating. Specifically, the ESEM pattern matrix suggested two factors that were consistently loaded by

Table 2. Factor Loading Matrix for the Five Factor ESEM

Items	F1	F2	F3	F4	F5
Get stressed out easily (+N)	0.29	0.31	-0.04	0.11	0.10
Am relaxed most of the time (-N)	-0.24	0.43	0.15	0.12	-0.07
Worry about things (+N)	-0.10	0.08	0.43	0.22	0.39
Seldom feel blue (-N)	0.05	0.18	0.10	0.25	-0.11
Am easily disturbed (+N)	0.41	0.45	0.01	-0.02	0.09
Get upset easily (+N)	0.40	0.54	-0.02	0.03	-0.02
Change my mood a lot (+N)	0.35	0.46	0.07	-0.02	0.10
Have frequent mood swings (+N)	0.47	0.51	0.00	0.03	-0.06
Get irritated easily (+N)	0.48	0.46	-0.05	0.10	-0.03
Often feel blue (+N)	0.33	0.49	-0.07	0.13	0.08
Am the life of the party (+E)	-0.23	0.08	0.37	0.14	-0.12
Don't talk a lot (-E)	0.05	0.48	-0.04	-0.04	0.31
Feel comfortable around people (+E)	-0.19	-0.05	0.40	0.20	0.07
Keep in the background (-E)	0.02	0.42	0.12	0.03	0.28
Start conversations (+E)	-0.13	0.03	0.45	0.29	-0.22
Have little to say (-E)	0.05	0.55	0.06	-0.02	0.24
Talk to a lot of different people at parties (+E)	-0.05	0.16	0.41	0.05	-0.22
Don't like to draw attention to myself (-E)	0.33	0.07	0.20	0.08	0.00
Don't mind being the center of attention (+E)	0.33	0.08	0.21	0.20	-0.10
Am quiet around strangers (-E)	0.26	0.20	-0.02	0.08	0.20
Feel little concern for others (-A)	-0.01	0.38	0.24	-0.08	0.15
Am interested in people (+A)	-0.10	-0.06	0.43	0.30	0.13
Insult people (-A)	0.13	0.70	0.03	-0.05	-0.14
Sympathize with others' feelings (+A)	0.01	0.07	0.08	0.65	0.28
Am not interested in other people's problems (-A)	-0.04	0.75	0.09	-0.32	0.01
Have a soft heart (+A)	0.08	0.05	0.08	0.54	0.21
Am not really interested in others (-A)	0.02	0.78	0.02	-0.33	0.02
Take time out for others (+A)	0.08	-0.03	0.46	0.32	0.01
Feel others' emotions (+A)	0.14	-0.06	0.33	0.53	0.01
Make people feel at ease (+A)	-0.13	-0.04	0.61	0.08	0.06
Have a rich vocabulary (+I)	-0.11	0.09	0.50	0.07	-0.04
Have difficulty understanding abstract ideas (-I)	-0.10	0.65	-0.08	0.06	0.13
Have a vivid imagination (+I)	-0.06	-0.02	0.52	0.11	-0.02
Am not interested in abstract ideas (-I)	-0.04	0.63	0.05	-0.04	0.00
Have excellent ideas (+I)	0.08	0.01	0.78	-0.11	-0.12
Do not have a good imagination (-I)	-0.05	0.72	-0.08	-0.06	0.07
Am quick to understand things (+I)	0.17	-0.17	0.68	-0.10	0.04
Use difficult words (+I)	0.29	0.32	0.35	-0.03	-0.23
Spend time reflecting on things (+I)	0.16	0.00	0.44	0.11	0.14
Am full of ideas (+I)	0.04	-0.02	0.73	-0.06	-0.11
Am always prepared (+C)	-0.07	0.04	0.57	-0.05	0.23
Leave my belongings around (-C)	-0.01	0.58	-0.05	0.18	-0.24
Pay attention to details (+C)	-0.17	-0.02	0.50	0.09	0.30
Make a mess of things (-C)	-0.03	0.61	-0.09	0.15	-0.29
Get chores done right away (+C)	0.00	-0.09	0.50	0.05	0.33
Often forget to put things back in their proper place (-C)	0.09	0.56	-0.13	0.23	-0.12
Like order (+C)	0.13	-0.15	0.42	-0.01	0.44
Shirk my duties (-C)	0.08	0.74	0.02	-0.05	-0.16
Follow a schedule (+C)	0.02	-0.07	0.52	0.03	0.07
Am exacting in my work (+C)	0.02	0.08	0.65	-0.10	0.15
Factor correlations	F1	F2	F3	F4	F5
	F1	-			
	F2	.29	-		
	F3	-0.08	-0.03	-	
	F4	.07	.24	.27	-
	F5	.04	-0.13	.27	.05

Note: Loadings in bold show those above 0.30. N = Neuroticism; E = Extraversion, O = Openness-to-experience; A = Agreeableness; C = Conscientiousness.

Table 3. Model Fit Statistics for the Method Artefact Measurement Models

	χ^2	<i>df</i>	CFI	TLI	RMSEA	WRMR
M1: Five-factor + positive & negative valence	2958.883*	1114	.92	.91	.046	1.441
M2: Five-factor + general method	2968.610*	1115	.92	.91	.046	1.444
M3: Five-factor + general, positive & negative valence	2265.253*	1064	.95	.94	.038	1.174

Note. * $p < .001$

Table 4. Factor Loading Matrix for the Five Factor CFA with a General Method Factor, and Positive and Negative Valence Factors

Items	Method	Positive	Negative	N	E	I	A	C
Get stressed out easily (+N)	-0.15		.40	0.32				
Am relaxed most of the time (-N)	0.01	-.48		-0.25				
Worry about things (+N)	0.65		.26	0.15				
Seldom feel blue (-N)	0.01	-.38		-0.03				
Am easily disturbed (+N)	-0.23		.53	0.38				
Get upset easily (+N)	-0.32		.62	0.34				
Change my mood a lot (+N)	-0.16		.54	0.34				
Have frequent mood swings (+N)	-0.33		.63	0.37				
Get irritated easily (+N)	-0.32		.60	0.41				
Often feel blue (+N)	-0.25		.59	0.30				
Am the life of the party (+E)	0.33	-.20			0.38			
Don't talk a lot (-E)	-0.08		.42		-0.31			
Feel comfortable around people (+E)	0.53	-.13			0.12			
Keep in the background (-E)	0.11		.45		-0.18			
Start conversations (+E)	0.39	-.32			0.40			
Have little to say (-E)	-0.05		.53		-0.20			
Talk to a lot of different people at parties (+E)	0.20	-.33			0.31			
Don't like to draw attention to myself (-E)	0.12		.30		-0.26			
Don't mind being the center of attention (+E)	0.08	-.41			-0.22			
Am quiet around strangers (-E)	-0.01		.32		-0.39			
Feel little concern for others (-A)	0.13		.41			-0.11		
Am interested in people (+A)	0.60	-.20				0.14		
Insult people (-A)	-0.33		.69			0.01		
Sympathize with others' feelings (+A)	0.36	-.39				0.56		
Am not interested in other people's problems (-A)	-0.28		.59			-0.30		
Have a soft heart (+A)	0.28	-.36				0.49		
Am not really interested in others (-A)	-0.37		.61			-0.29		
Take time out for others (+A)	0.51	-.34				0.16		
Feel others' emotions (+A)	0.44	-.38				0.37		
Make people feel at ease (+A)	0.66	-.19				-0.09		
Have a rich vocabulary (+I)	0.42	-.26					0.15	
Have difficulty understanding abstract ideas (-I)	-0.18		.56				-0.18	
Have a vivid imagination (+I)	0.47	-.19					0.26	
Am not interested in abstract ideas (-I)	-0.17		.56				-0.09	
Have excellent ideas (+I)	0.52	-.27					0.55	

Table 4. Continued.

Items	Method	Positive	Negative	N	E	I	A	C
Do not have a good imagination (-I)	-0.28		.59				-0.17	
Am quick to understand things (+I)	0.58	-.11					0.38	
Use difficult words (+I)	-0.09	-.58					0.32	
Spend time reflecting on things (+I)	0.44	-.31					0.00	
Am full of ideas (+I)	0.55	-.23					0.30	
Am always prepared (+C)	0.59	-.20						-0.11
Leave my belongings around (-C)	-0.26		.59					0.14
Pay attention to details (+C)	0.65	-.13						0.01
Make a mess of things (-C)	-0.34		.59					0.56
Get chores done right away (+C)	0.62	-.14						-0.30
Often forget to put things back in their proper place (-C)	-0.29		.62					0.49
Like order (+C)	0.56	-.08						-0.29
Shirk my duties (-C)	-0.36		.70					0.16
Follow a schedule (+C)	0.53	-.16						0.37
Am exacting in my work (+C)	0.56	-.29						-0.09
Factor correlations	Method	Positive	Negative	N	E	I	A	C
	Method	-						
	Positive	-	-					
	Negative	-	-.81	-				
	N	-	-	-				
	E	-	-	-	-.31	-		
	I	-	-	-	.16	-		
	A	-	-	-	.42	-.10	-.20	-
	C	-	-	-	-.34	.25	-.25	.32

Note. N = Neuroticism; E = Extraversion, O = Openness-to-experience; A = Agreeableness; C = Conscientiousness

either positively or negatively worded items, suggesting that the variance attributable to these item valence was substantial (Suárez-Alvarez et al., 2018). Once these two method factors were explicitly modeled, a CFA of all five factors demonstrated good levels of model fit, certainly comparable to other five factor inventories (Booth & Hughes, 2014). However, eleven items still failed to load substantially ($> .3$) on their hypothesized factor, with substantial loadings on respective method factors. Nevertheless, the current results suggest that when method factors are ignored, the IPIP-50-S is inappropriate for use within Peruvian samples. However, once the effect of acquiescence due to item valence has been modelled, the structure of the IPIP-50-S is closer to the a priori structure dictated by the English-language version (Goldberg, 1992). These findings are consistent with similar patterns in other questionnaires that use positive and negatively worded items. For example, Suárez-Alvarez et al. (2018) examined a self-efficacy scale, within a Spanish-speaking sample, and found that combinations of positive and negative items reduced

test reliability, undermined unidimensionality, and produced scale means that differed significantly from means derived from versions with all positive or negative items.

One striking observation is the magnitude of the method effects observed within this sample. We believe there are likely two main reasons for the substantial method effects. First, it is possible that diversity in lexical and syntactical structures across different Spanish-speaking nations meant that some items failed to translate in an equivalent manner, which exacerbated general method effects (Cupani & Lorenzo-Seva, 2016). Second, unlike previous studies to investigate this inventory, our sample was educated to secondary level, not university level (e.g., Cupani, 2009; Gross et al., 2012). Previous research has demonstrated that method artefacts, such as acquiescence, are exacerbated in samples with lower levels of educational attainment (Rammstedt et al., 2010, 2017).

Nevertheless, the modeling approach employed largely controlled for these substantial effects, and thus,

our results align with previous research demonstrating that once socially desirable or acquiescent responding is modelled, five factor inventories are somewhat structurally stable across cultures and educational levels (Rammstedt et al., 2010; 2013; Suárez-Alvarez et al., 2018).

To the authors knowledge, this is the first published attempt to examine the psychometric properties and appropriateness of the IPIP-50-S for use within a non-university educated sample, here a Peruvian sample. From the findings, it is recommended that caution be exercised in using the IPIP-50-S in such samples, without explicit actions taken to account for the influence of item valence and socially desirable responding. However, use of alternative measures may be preferable. For example, Cupani and Lorenzo-Seva (2016) proposed a variant of the Spanish IPIP designed to mitigate the effects of acquiescent responding. The data for the current study was collected prior to publication of this measure; however, future research might focus on the properties of this inventory across countries and educational levels.

In closing, we note the importance of psychometric evaluations of freely available translated inventories, like those provided by the IPIP, and would strongly advocate for continued efforts to link published and unpublished evaluations. Such a resource would allow researchers interested in cross-cultural research to identify whether translations provide accurate measurement in their target population and thus whether they are appropriate for the intended purposes (Hughes, 2018).

Supplementary Materials

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/SJP.2020.11>.

References

- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117, 187–215. <https://doi.org/10.1037/0033-2909.117.2.187>
- Booth, T., & Hughes, D. J. (2014). Exploratory structural equation modeling of personality data. *Assessment*, 21, 260–271. <https://doi.org/10.1177/1073191114528029>
- Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1–22. <https://doi.org/10.1006/cogp.2001.0748>
- Cupani, M. (2009). El cuestionario de personalidad IPIP-FFM: Resultados preliminares de una adaptación en una muestra de preadolescentes argentinos [The IPIP- FFM Questionnaire of Personality: Preliminary results for the adaptation in a sample of young Argentinean adolescents]. *Perspectivas en Psicología*, 6, 51–58.
- Cupani, M., & Lorenzo-Seva, U. (2016). The development of an alternative IPIP inventory measuring the Big-Five factor markers in an Argentine sample. *Personality and Individual Differences*, 91, 40–46. <https://doi.org/10.1016/j.paid.2015.11.051>
- Diamantopoulos, A., Reynolds, N. L., & Simintiras, A. C. (2006). The impact of response styles on the stability of cross-national comparisons. *Journal of Business Research*, 59, 925–935. <https://doi.org/10.1016/j.jbusres.2006.03.001>
- Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, 41, 417–440. <https://doi.org/10.1146/annurev.ps.41.020190.002221>
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26–42. <http://doi.org/10.1037/1040-3590.4.1.26>
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7–28). Tilburg University Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Gross, M., Zalazar-Jaime, M., Piccolo, N., & Cupani, M. (2012, October). Nuevos estudios de validación del Cuestionario de Personalidad IPIP-FFM [New validation studies of the IPIP-FFM Personality Questionnaire] [Conference Paper]. X Congreso Latinoamericano de Sociedades de Estadística. Córdoba, Argentina. https://www.researchgate.net/publication/274718129_Nuevos_Estudios_de_Validacion_Del_Cuestionario_De_Personalidad_IPIP-FFM
- Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14, 332–346. <https://doi.org/10.1177/1088868310361240>
- Hu, L., & 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, 1–55. <https://doi.org/10.1080/10705519909540118>
- Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. In P. Irving, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary approach to survey, scale and test development*. Wiley.
- Johnson, T., Kulesa, P., Cho, Y. L., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 Countries. *Journal of Cross-Cultural Psychology*, 36, 264–277. <https://doi.org/10.1177/0022022104272905>
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big-Five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22, 471–491. <https://doi.org/10.1037/a0019227>
- McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509–516. <https://doi.org/10.1037/0003-066X.52.5.509>
- McCrae, R. R., Terracciano, A., & Personality Profiles of Cultures Project (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures.

- Journal of Personality and Social Psychology*, 88, 547–561. <https://doi.org/10.1037/0022-3514.88.3.547>
- Mlačić, B., & Goldberg, L. R.** (2007). An analysis of a cross-cultural personality inventory: The IPIP Big-Five Factor markers in Croatia. *Journal of Personality Assessment*, 88, 168–177. <http://doi.org/10.1080/00223890701267993>
- Muthén, L. K., & Muthén, B. O.** (1998–2017). *Mplus user's guide*. 8th Ed. Muthén & Muthén.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P.** (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P.** (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Rammstedt, B., Danner, D., & Bosnjak, M.** (2017). Acquiescence response styles: A multilevel model explaining individual-level and country-level differences. *Personality and Individual Differences*, 107, 190–194. <https://doi.org/10.1016/j.paid.2016.11.038>
- 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, 53–61. <https://doi.org/10.1016/j.jrp.2009.10.005>
- Rammstedt, B., Kemper, C. J., & Borg, I.** (2013). Correcting Big Five Personality Measurements for Acquiescence: An 18-country cross-cultural study. *European Journal of Personality*, 27, 71–81. <https://doi.org/10.1002/per.1894>
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J.** (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149–158. <http://doi.org/10.7334/psicothema2018.33>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H.** (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23–74.
- Valentine, A.** (2013). *Is translation enough? A study of the item characteristics which influence equivalence between English and Spanish versions of a selection test* (Publication No. 3568766) [Doctoral dissertation, State University of New York at Albany]. ProQuest Dissertations and Theses Global.