Article

Dimensionality and Reliability Assessment of a Field Implementation of the Big Five in Mexican Children

Journal of Psychoeducational Assessment
2021, Vol. 0(0) 1–14

DOI: 10.1177/07342829211002581
journals.sagepub.com/home/jpa
SAGE

Yadira Peralta[1,2], Adriana Aguilar-Rodriguez[2,3,4], Osiel González Dávila[2,4,5], and Alfonso Miranda[1,2,6]

Abstract

According to the literature, the use of the Berkeley Puppet Interview (BPI) to measure Big Five personality traits in children provides reliable and valid scores. However, the implementation of the BPI could be costly, especially when working with large sample sizes. Big Five self-reports were collected from 1118 Mexican children aged 7–8 years using a modified version of the BPI protocol and a Spanish version of the Big Five questionnaire. The main objective of this study was to inquire whether some modifications in the application protocol of the BPI could still provide reliable personality scores for the population under study. We report item–rest correlation, Cronbach's alpha and omega as reliability measures, and confirmatory factor analytic models to investigate dimensionality. The results show that the personality trait scores are markedly reliable and that the dimensionality of the instrument holds for the Mexican sample.

Keywords

personality traits in children, Big Five in children, Field Puppet Interview, dimensionality and reliability, personality development

Introduction

The five-factor model of personality is nowadays one of the most popular tools used in applied work to explain interindividual differences in personality and to investigate how personality correlates with policy-relevant outcomes in many fields. The model proposes five comprehensive

[1]Department of Economics, Center for Research and Teaching in Economics (CIDE), Aguascalientes, Mexico
[2]Program for Longitudinal Studies, Experiments and Surveys, Center for Research and Teaching in Economics (CIDE), Aguascalientes, Mexico
[3]Center for Research in Geospacial Information Science (CentroGeo), Aguascalientes, Mexico
[4]Mexican National Council for Science and Technology (CONACYT), Mexico
[5]University of Monterrey, Mexico
[6]Institute for the Study of Labor, Germany

Corresponding Author:
Yadira Peralta, Department of Economics, Center for Research and Teaching in Economics, Circuito Tecnopolo Norte 117, Aguascalientes 20313, Mexico.
Email: yadira.peralta@cide.edu

dimensions: *extraversion*, *agreeableness*, *conscientiousness*, *neuroticism*, and *openness* to new experiences—for a discussion on the historical roots and structure of the concepts of personality based on the five-factor model and its limitations; see, for instance, Digman (1990), Goldberg (1990), McAdams (1992), McCrae and John (1992), and Wiggins (1996). Briefly, extraversion is related to emotional expressiveness and sociability; agreeableness refers to interpersonal traits such as trust, altruism, and kindness; conscientiousness is characterized by impulse control and goal-oriented behaviors; neuroticism describes negative attributes such as anxiety, irritability, sadness, and emotional instability; and openness portrays intellectual characteristics such as imagination, cognitive ability, and curiosity.

A number of studies have measured and analyzed the Big Five personality traits in elementary school children in cross-sectional and longitudinal designs (e.g., Abe, 2005; Asendorpf & Van Aken, 2003; Caspi et al., 2003; Ehrler et al., 1999; Hampson et al., 2016; Resing et al., 1999). In these studies, teachers, parents, or other adults usually report on children's personality. This practice was motivated by evidence of questionable estimates of temporal stability of children's personality when reported by children (e.g., Roberts & DelVecchio, 2000). Thus, relying on reports from adults seemed reasonable. However, there was still interest from researchers in asking children directly about their personality characteristics because of the possibility that adults may not be aware of a number of problems that children might face at a young age (Ringoot et al., 2013).

Aiming to investigate whether it was possible to have stable and valid personality scores reported from children directly, Ablow and Measelle (1993) developed the administration and scoring system manuals for the Berkeley Puppet Interview (BPI). The BPI allows a natural conversational exchange between a child and two puppets. The puppets make positive and negative statements about themselves regarding behaviors or attributes (e.g., "I am smart"; "I am not smart") and then ask the child to describe her/himself (e.g., "How about you?"), minimizing the socially appealing response. The interaction between the child and the two puppets is videotaped. Trained scorers watch each interview and score the answers in a 7-point Likert scale (Ablow & Measelle, 1993).

The use of puppets during the interview is crucial, and it expanded the work of a number of researchers who used puppets in their studies of young children (Eder, 1990; Irwin, 1985; Schaefer & O'Conner, 1983; Woltmann, 1952). In their seminal study, Measelle et al. (1998) discussed the psychometric properties of self-perception scales using the BPI. The authors found that children had a multidimensional self-concept that could be reliably measured. Later, Measelle et al. (2005) conducted a follow-up study where they investigated the validity and stability of self-reports of Big Five personality traits in 5–7 year old children using the BPI in conjunction with the Big Five questionnaire (BPI-BFQ hereafter). The children's perceptions of their personality were compared with adult informants' reports coming from the child's mother, father, and teachers. The authors found satisfactory psychometric properties of the instrument, and most personality dimensions were stable and coherent throughout the period of the study.

Although the BPI-BFQ has been shown to provide valid and reliable scores, its implementation may turn expensive. Performing the interview, videotaping, and scoring may become impractical in certain settings. Collecting behavioral data from children in developing countries where funding is scarce, and participants are from remote areas may become unfeasible. Still, it is precisely in developing countries where studies on children's personality are scanter, and evidence is urgently needed. Accordingly, the present study proposes a modified version of the BPI-BFQ where items from the Big Five questionnaire remain unchanged, and the scoring process is modified to obtain self-reported dichotomous data directly from children.

The main objective of the present study is to assess the dimensionality and reliability of a Spanish version of the Big Five questionnaire using our modified version of the BPI protocol. We performed item analysis and estimated confirmatory factor analytic models to investigate

dimensionality. The article is structured as follows. Next section introduces the data, followed by a description of subjects and procedures, and data analysis. Results are presented, and the final section concludes with discussion, limitations, and future directions.

Methods

Data and Sample Design

We used data from the first wave of the Aguascalientes Longitudinal Study of Child Development (EDNA-Spanish acronym) (Miranda et al., 2019, 2020). EDNA is a multi-thematic, multilevel, and multidisciplinary study of 1000 students in 100 schools that are a representative sample of the population of children who started first grade in August 2016 in a public primary school of the state of Aguascalientes, Mexico.[1] EDNA interviews mothers (or the primary caretaker) in their households, children in their school, and teachers in the premises of regional units of education. The sample has a classic probabilistic stratified two-stage cluster design, where schools are selected in the first stage and pupils are selected in the second stage. Children interviews started in February 2018 and concluded in June 2018, when children were in second grade. The analytical sample used for the present analyses is comprised of 1118 children with a mean of 7 years 3 months old (SD = .49); 51% were women.

EDNA's child questionnaire measures children's personality traits using 58 items from the BPI-BFQ, which was graciously provided by Jeffrey Measelle, at the University of Oregon, to whom we are grateful. Given the large scale of the sample and the logistic challenges that visiting over 100 schools and 1000 households impose, achieving laboratory interviewing conditions and videotaping all children's interviews was unfeasible. Thus, the scoring of the BPI as originally implemented by the BPI-BFQ was not viable. Hence, we established a simplified implementation that used the building blocks of the puppet interviewing techniques developed by Measelle et al. (2005), authors of the original BPI-BFQ scale, but that did not require strict laboratory conditions and/or interview videotaping, and that could be easily used with large samples. For simplicity, hereafter, we call such implementation the Field Puppet Interview-Big Five Questionnaire (FPI-BFQ). We also translated the BPI-BFQ items into Spanish. The translated version was piloted and adjusted accordingly before its field implementation. The Spanish version of the instrument is available by contacting the authors.

Subjects and Procedures

All interviews were conducted by professional and previously trained female interviewers. On the day of the interview, enumerators arrived at the school before the beginning of the morning shift (8:00 hrs.) or the afternoon shift (13:00 hrs.) to prepare the facilities. The interview room was in most cases the computer lab or the school library. The room was divided in two interviewing stations by a portable room divider. In each interviewing station, there were a table and two chairs facing each other. Attendance of each EDNA selected child was verified. Each child was then walked by his/her teacher to the interviewing room and walked back to the classroom once the interview was done. We followed international best practices, and informed consent from children's parents was obtained in order to apply the instrument. In addition, verbal assent from children at the time of the interview was also requested.

At the beginning of the interview, a tablet was given to the child along with instructions on how to use it (see Figure 1). For the FPI-BFQ application, the interviewer explained the activity and introduced two identical hand puppets (two gray and white cats, see Figures 2 and 3) to the child, called "Iggy" and "Ziggy." Following Measelle et al. (1998), the sex-neutral and phonetically similar names were chosen to avoid any possible bias induced by the puppets' name. The two

Figure 1. Field Puppet Interview-Big Five Questionnaire implementation: Tablet delivery to the participant.

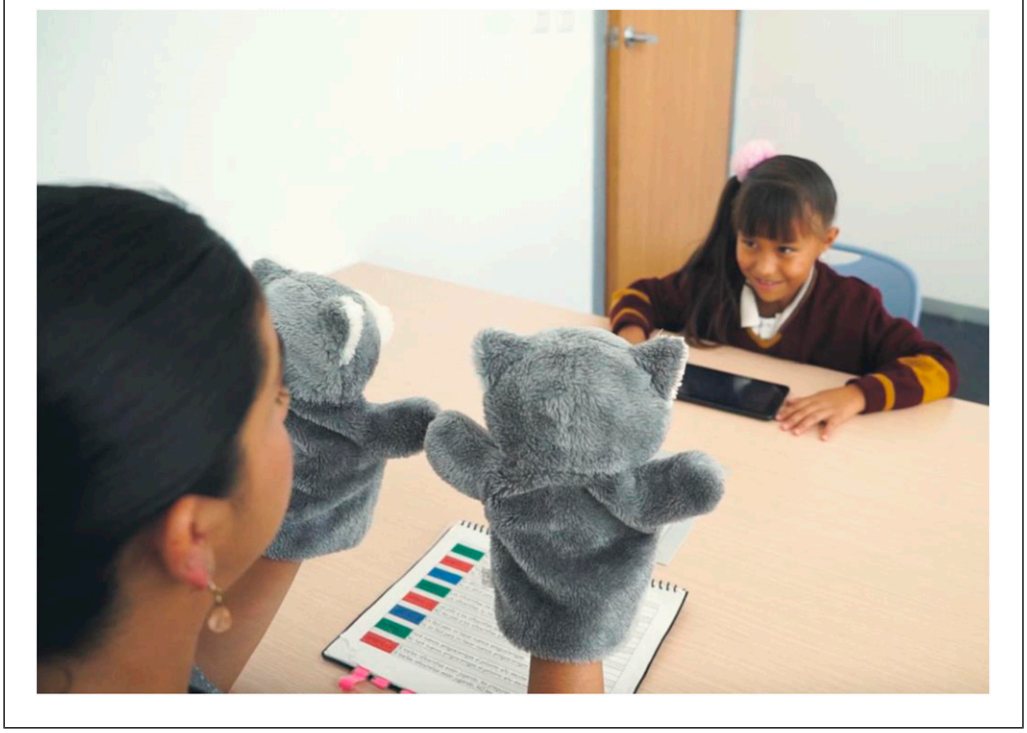

Figure 2. Field Puppet Interview-Big Five Questionnaire implementation: Introducing the puppets.

puppets made opposite statements about themselves and asked the child to identify her/himself with one or the other behavior. Instead of videotaping the interview and scoring the responses afterward, we implemented a dichotomous-choice response format. Children were asked to respond directly on the tablet, without taking the tablets on their hands, by touching with their finger the figure of the puppet that would match their response (see Figures 1–3).

Using an item as an example, we assessed whether the procedure was clear for the child (e.g., "Hi, I am Ziggy and I am in the second grade of primary school"; "Hi, I am Iggy and I am in the fourth grade of primary school; how about you?") and continued until the training question was answered correctly. Positive and negative responses alternate between the two puppets, in such way that "Iggy" or "Ziggy" were not associated with only positive or negative answers.

Data Analysis

Negatively worded items were reverse coded in order to appropriately calculate reliability indices and item analysis. The questionnaire includes the items described in Measelle et al. (2005), although authors have added more items to the inventory over time. Extraversion was measured by 10 items, agreeableness by 13, consciousness and neuroticism by 12, and openness by 13 items each. Table 1 presents the number of items used in this study and the number of items reported in Measelle et al. (2005). Cronbach's alpha measures reported in Measelle et al. (2005) and the ones computed using data from our study for each personality trait are also reported in Table 1. More

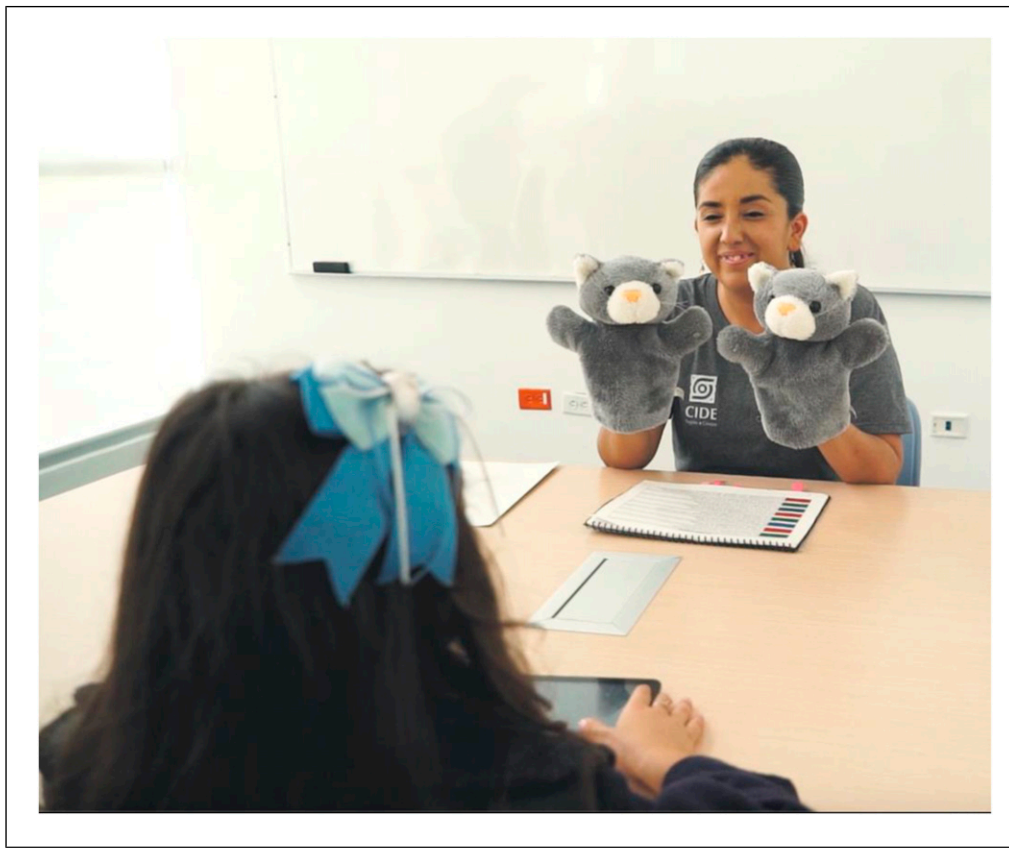

Figure 3. Field Puppet Interview-Big Five Questionnaire implementation: Selecting an option.

Table 1. Number of Items and Coefficient Alpha in EDNA and Measelle et al. (2005).

	Number of items		Coefficient alpha	
Subscale	EDNA	Measelle et al. (2005)	EDNA	Measelle et al. (2005)
Extraversion	10	10	.73	.67
Agreeableness	13	10	.82	.71
Conscientiousness	12	9	.71	.71
Neuroticism	12	7	.75	.70
Openness	13	5	.80	.66

items were used to measure each personality trait in our study, except for extraversion, where 10 items were used in both studies. Coefficient alpha was higher for each personality scale in our study than the measures reported by Measelle et al. (2005) where it ranged from .66 to .71, whereas in our study, considering all initial items, it ranged from .71 to .82. Descriptive statistics of all items under study are presented in Table S1 in the Supplementary Materials.

Reliability measures. There are several measures of reliability, but Cronbach's alpha is typically reported. Cronbach's alpha or coefficient alpha is an estimate of internal consistency. However, several limitations of coefficient alpha have been discussed in the literature (e.g., Dunn et al., 2014; McNeish, 2018; Sijtsma, 2009; Yang & Green, 2011; Zinbarg et al., 2005). Researchers argue that coefficient alpha is an unsuitable measure of internal consistency because it relies on assumptions that are commonly violated especially when working with psychological data (Dunn et al., 2014). To elaborate, one of the main problematic assumptions is the one of tau-equivalence, which in practical terms means that all items are related equally to the construct or that each item contributes equally to the total score (McNeish, 2018). Typically, items will relate differently to the construct being measured, some items will be more strongly related than others or will measure the construct more precisely. When assumptions are violated, coefficient alpha tends to underestimate the internal consistency reliability (Dunn et al., 2014; Sijtsma, 2009; Zinbarg et al., 2005). So, what is reported is typically a lower bound of reliability. Researchers have encouraged the use of other measures of reliability. One of such measures is called *omega*. Omega makes less restrictive assumptions than coefficient alpha, which makes it a more suitable measure of internal consistency reliability (Dunn et al., 2014). It assumes that not all the items contribute equally to the total score. Thus, we adopted this measure by using the omega function in the psych R package (Revelle, 2019).

For the purpose of comparing results with Measelle et al. (2005), we computed coefficient alpha reliability index, as shown in Table 1. However, we used omega as a measure of reliability of the scales in the FPI-BFQ and also omega if item-deleted along with other indices when making decisions on keeping items to validate the personality scales. In addition, both reliability indices, coefficient alpha and omega, are reported using the tetrachoric covariance matrix, which is appropriate when working with dichotomous items, as recommended in the literature (Gadermann et al., 2012).

Item–rest correlation. We used item–rest correlation as an index of item discrimination to help us assess the quality of each item within each of the subscales of the FPI-BFQ. As a general rule for a discrimination index, items highly discriminating (with values of .50 or above) are retained, items with values below .20 are eliminated, and items between these two bounds should be considered for modifications (Thorndike & Thorndike-Christ, 2010). In addition, other measures

were considered to make decisions on the items measuring each subscale. This process is explained below.

Factor structure. We fit a series of confirmatory factor analysis (CFA) models investigating the dimensionality of the FPI-BFQ. As an initial step toward assessing the dimensionality of the FPI-BFQ instrument, we fit the model suggested by Measelle et al. (2005). That is, we estimated a correlated (oblique) five-factor CFA model. In such model, the five factors (five subscales in the Big Five instrument) were allowed to correlate with each other. We then assessed whether the guidelines for adequate model-data fit were satisfied (Hu & Bentler, 1999). Guidelines were not satisfied as it will be shown in the results section. Thus, we validated the Big Five dimensions in an iterative manner. We analyzed the estimation results of the CFA model, the item–rest correlation, and the omega statistic if item-deleted. We decided to eliminate items that had the lowest factor loadings and item–rest correlation and that would also increase reliability measured by omega from the subsequent analyses. That is, we relied on three measures rather than a single one to decide which items were not going to be included in further analyses. Then, we fit the updated correlated five-factor CFA model to the remaining data. We repeated this process one more time and ended up with an estimated correlated five-factor CFA model that had good factor loadings. Once item elimination was no longer supported by the data, we examined modification indices of the final correlated five-factor CFA model to improve model-data fit. We only added residual covariance parameters among the observed variables within each subscale when suggested by modification indices. Thus, the latent structure of the model remained, and the parameters added to the model did not alter it theoretically (Kline, 2011).

The lavaan R package (Rosseel, 2012) was used to estimate all CFA models. Due to the nature of the data (dichotomous data), diagonally weighted least squares estimation procedures were used. Commonly applied guidelines for adequate model-data fit suggest root mean square error of approximation (RMSEA) $\leq .06$, standardized root mean square residual (SRMR) $\leq .08$, and comparative fit index (CFI) and Tucker-Lewis index (TLI) $\geq .95$ (Hu & Bentler, 1999). In addition, ideally, the chi-square statistics must be nonsignificant to support the adequacy of the model. Regarding factor loadings, many researchers define weak factor loadings as those with values of .4 (Briggs & MacCallum, 2003; Widaman, 1993; Ximénez, 2006). But in practice, researchers might consider loadings as low as .2 and .3 (Briggs & MacCallum, 2003).

Results

As a starting point, considering all available items for the analysis, internal consistency reliability measured by omega ranged from .75 to .87 (see Table 2). These results suggest that the description on the five personality dimensions was fairly consistent among Mexican kids between 7 and 8 years old using the FPI-BFQ. Moreover, these results were further improved when removing some items that did not present strong properties in terms of the factor structure as discussed below.

In order to support factor analyses, the Kaiser–Meyer–Olkin (KMO) index and Bartlett's sphericity test are reported on Table 3. The KMO index ranged from .68 to .79, and Bartlett's test was statistically significant. Thus, factor analytic methods are suitable. Moreover, the composite reliability (CR) and average variance extracted (AVE) indices were assessed. CR ranged from .74 to .86, and AVE ranged from .26 to .41. According to Fornell and Larcker (1981), AVE is more conservative than CR, and one can still assess convergent validity using CR only. Thus, results show that there is an adequate level of agreement among items measuring each subscale and that convergent validity is still adequate (Fornell & Larcker, 1981). However, the amount of measurement error is considerable.

Table 2. Number of Items and Omega Statistic for the Initial and Final Pools of Items in the Field Puppet Interview-Big Five Questionnaire Instrument.

	Initial stage		Final stage	
Subscale	Number of items	Omega	Number of items	Omega
Extraversion	10	.79	7	.87
Agreeableness	13	.87	12	.88
Conscientiousness	12	.78	9	.82
Neuroticism	12	.75	10	.82
Openness	13	.86	9	.91

Table 3. Indices to Test for the Adequacy of Factor Analyses.

Subscale	KMO index	Bartlett's sphericity test	CR index	AVE index
Extraversion	.70	$\chi^2(45) = 3053.69$, $p < .001$	.77	.33
Agreeableness	.79	$\chi^2(78) = 4925.94$, $p < .001$	.83	.30
Conscientiousness	.68	$\chi^2(66) = 2843.80$, $p < .001$	.74	.26
Neuroticism	.73	$\chi^2(66) = 3078.18$, $p < .001$	.78	.27
Openness	.72	$\chi^2(66) = 4594.79$, $p < .001$	.86	.41

KMO = Kaiser–Meyer–Olkin; CR = composite reliability; AVE = average variance extracted.

To investigate the dimensionality of the FPI-BFQ, we estimated a correlated five-factor CFA model considering all items available, as supported by Measelle et al. (2005). The fit indices are presented in Table 4 under the *initial model* legend. RMSEA was lower than .06, but SRMR, CFI, and TLI did not satisfy their suggested thresholds. Notice that the chi-square test was statistically significant, so the exact model-data fit hypothesis is rejected. However, there is extensive literature discussing how large sample sizes result in significant chi-square test (e.g., Cheung & Rensvold, 2002; Kline, 2011; Meade et al., 2008). These studies suggest that chi-square constitutes a highly sensitive statistical test to sample size, especially when sample sizes bigger than 300 are under study (Kline, 2011). Thus, we present the chi-square test results in Table 4, but we do not discuss them extensively.

Then, we examined the factor loadings under the *initial model*, and we detected some items with low factor loadings. We highlighted those items and also considering results of item–rest correlation and omega if item-deleted, we decided to remove items that presented low values in these three statistics.[2] We did not want to remove unnecessary items considering the factor structure examination solely. Thus, we only dropped an item that had low factor loading, low item–rest correlation and that will increase, or maintain omega if item-deleted. Following this process, we removed items 3 and 10 from the extraversion subscale; item 2 from agreeableness; items 4 and 11 from conscientiousness; items 5 and 7 from neuroticism; and items 10, 11, and 12 from openness. We estimated again the correlated five-factor CFA model considering the remaining items. The fit indices are presented in Table 4 under the *intermediate model* legend. Again, RMSEA was lower than .06, but SRMR, CFI, and TLI did not satisfy their suggested thresholds. We redid the process of removing items in this second iteration and eliminated two more items: item 1 from extraversion and item 5 from conscientiousness. In addition, we considered modification indices for this final model. The extra relationships added to the model were residual covariances among the observed variables within each subscale. In this sense, the

Table 4. Fit Indices of the Correlated Five-Factor CFA Models.

Statistic	Correlated five-factor models		
	Initial model	Intermediate model	Final model
χ^2 test	5381.37 (df = 1642) p = .000	3491.08 (df = 1117) p = .000	2489.92 (df = 988) p = .000
RMSEA [90%CI]	.045 [.044, .046]	.044 [.042, .045]	.037 [.035, .039]
SRMR	.10	.10	.09
CFI	.83	.88	.92
TLI	.82	.87	.91
GFI	.88	.92	.94
AGFI	.88	.91	.93
IFI	.83	.88	.92

CFA = Confirmatory Factor Analysis; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index; TLI = Ticker-Lewis Index; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; IFI = Incremental Fit Index.

latent structure of the model remains the same, and the modifications did not alter the model theoretically (Kline, 2011).

Fit indices of this model are presented in Table 4 under the *final model* legend. RMSEA was lower than .06, SRMR = .09, CFI = .92, and TLI = .91. The last three indices were close to satisfy the suggested thresholds by Hu and Bentler (1999). However, researchers have pointed out limitations in such thresholds and have cautioned using them as "golden rules" (Fan & Sivo, 2005; Kline, 2011; Marsh et al., 2004; Yuan, 2005). In that sense, the value of SRMR is on the frontier of the established cutoff of Hu and Bentler (1999), and the values of CFI and TLI are greater than .90, which has also been supported in the literature as an acceptable cutoff (as cited in Marsh et al., 2004, 2005). Thus, we argue our final model provides an adequate fit to the data in hand. In this final model, all factor loadings are acceptable (see Table 5). Likewise, item–rest correlation and omega if item-deleted showed adequate values under the final model (see Table S3 in the Supplementary Materials). All estimated correlations between the five factors were statistically significant, ranging from .47 to .81 in absolute value, and were all in the expected direction according to the literature (see Table S4 in the Supplementary Materials). Furthermore, considering the final CFA model, we tested for gender differences in the CFA structure. We found that metric measurement invariance cannot be assumed ($\chi^2(42) = 212.87$, $p < .001$). Thus, there are gender differences in the model structure as it has also been found in the literature (e.g., Arpaci & Kocadag Unver, 2020).

Last, considering the items in the final model, we computed again omega as a measure of internal consistency reliability. This is presented in Table 2 under the *final stage* legend. Omega ranges from .82 to .91, which indicates scores derived from the Spanish version of the FPI-BFQ are highly consistent.

Discussion, Future Directions, and Limitations

We proposed a modified version of the BPI-BFQ, which we called the FPI-BFQ, where self-reported dichotomous data were collected directly from children using the original items from the BPI-BFQ translated into Spanish. The focus was on assessing the dimensionality and reliability of the collected data, which comes from EDNA, a pioneer study in Mexico conducted on a prospective cohort of children. The results suggest that the FPI-BFQ has adequate psychometric properties. High indices of internal consistency, measured by omega, were found for the five

Table 5. Factor Loadings of the Correlated Five-Factor CFA Final Model.

Item	Estimate	Std. Error	Item	Estimate	Std. Error
Extraversion			Neuroticism		
Item2	.56	.02	Item1	.70	.03
Item4	.66	.03	Item2	.46	.02
Item5	.45	.03	Item3	.58	.03
Item6	.46	.02	Item4	.40	.02
Item7	.70	.03	Item6	.36	.02
Item8	.60	.03	Item8	.36	.02
Item9	.54	.02	Item9	.59	.03
Agreeableness			Item10	.59	.02
Item1	.25	.02	Item11	.64	.03
Item3	.45	.02	Item12	.37	.02
Item4	.73	.02	Openness		
Item5	.48	.02	Item1	.60	.02
Item6	.51	.02	Item2	.51	.02
Item7	.72	.02	Item3	.74	.02
Item8	.66	.02	Item4	.45	.02
Item9	.32	.02	Item5	.73	.03
Item10	.61	.02	Item6	.51	.02
Item11	.64	.02	Item7	.76	.02
Item12	.63	.02	Item8	.64	.02
Item13	.36	.02	Item9	.71	.02
Conscientiousness					
Item1	.39	.02			
Item2	.28	.02			
Item3	.36	.02			
Item6	.49	.03			
Item7	.52	.03			
Item8	.71	.03			
Item9	.58	.03			
Item10	.65	.03			
Item12	.42	.03			

Note. All estimates were significant at $p < .001$. CFA = Confirmatory Factor Analysis.

personality traits. Moreover, factor analytic models confirmed a five-dimensional correlated structure for the Mexican sample to which the FPI-BFQ was applied. Some of the items of the original BPI-BFQ did not work well with the population at hand. In total, 11 items were eliminated.

To the best of our knowledge, this is the first study that develops a modified scoring protocol from the original BPI. Moreover, it was applied to a large sample of children. We believe that the use of a simplified scoring process, as implemented in the FPI-BFQ instrument, diminishes financial and human costs of application and scoring. This may help expanding personality research when financial constraints are important as it is the case in developing countries where research on children's development is scarce and funding restrictions exist. Furthermore, a paper-and-pencil version of the instrument where children directly mark their answers in a piece of paper could be an acceptable form of assessment that would certainly need further investigation. This could be an option that may encourage research in scenarios for which having or using a tablet might be problematic.

Findings suggested that gender differences exist. That is, the factor analytic structure presents some differences by gender. Future research should investigate thoroughly on such differences. Likewise, a further step on the assessment of the item and scale properties of the FPI-BFQ would be to employ item response theory (IRT) techniques. An IRT approach could enrich our understanding of the scores. For instance, we could investigate whether the FPI-BFQ can measure the personality traits with the same precision across the unobservable continuum of the personality spectrum. Another future direction would be to investigate the relation in personality traits between children and mothers or primary caretakers. The EDNA study also collected information on personality from the mother or primary caretaker, so we have readily available data.

Last, this study is not without limitations. First, stability of personality traits in children is a core subject of study that was not explored in the present article. Further research should employ longitudinal data in order to explore this matter. Furthermore, this type of data could be collected as part of the same EDNA study. Second, other Spanish-speaking populations may behave differently than the analyzed Mexican sample. Thus, if the Spanish items of the FPI-BFQ were applied to other populations, more research would be needed to validate them. Specifically, think aloud protocols or having a panel of experts in order to make any necessary changes to the wording of the items would be highly advisable.

Summarizing, the findings from this article confirm the Big Five personality structure in collected data from Mexican children using the FPI-BFQ. They also suggest that the proposed change in the BPI scoring methodology provides reliable scores for the population under study. In addition, internal consistency is congruent with previous literature in personality development. These results add to the literature of children's self-reported personality traits by giving evidence on how young children between 7 and 8 years old are able to provide reliable information on their personality traits by themselves. Moreover, we hope the proposed scoring process in this instrument, allowing children to directly provide dichotomous responses through the use of a tablet, will help other researchers facing funding constraints and for which it is not possible to implement a more expensive protocol. Thus, the FPI-BFQ may help conduct research in scenarios where funding is limited and the implementation of the original BPI protocol impractical.

Acknowledgments

We would like to thank Jeffrey Measelle for kindly providing the items from the BPI-BFQ. We are grateful to our colleagues Hernán Bejarano, Daniel Zizumbo-Colunga, and Jaime Sainz who contributed to the design of the EDNA's questionnaires; and to Sandra Nallely Núñez Nájera who was the field work coordinator during the data collection stage. In addition, we are grateful to Sandra Nallely Núñez Nájera, Valeria Gracia Olvera, and Daniel Zizumbo-Colunga for their suggestion of assigning item colors for easing the reading of the FPI-BFQ instrument during the piloting stage.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was financially supported by the Cátedras CONACYT program [project No. 874], from The Hewlett Foundation [project No. 2013-8758], and from CIDE [project FAI 12171077].

ORCID iD

Yadira Peralta https://orcid.org/0000-0003-4823-6939

Supplemental Material

Supplemental material for this article is available online.

Notes

1. More information about the study can be found in the survey's documentation in Author (2019) and also in Author (2020).
2. Tables with information about factor loadings, item–rest correlation, and omega if item-deleted for the three CFA models are available as Tables S2 and S3 in the Supplementary Materials.

References

Abe, J. A. A. (2005). The predictive validity of the five-factor model of personality with preschool age children: A nine year follow-up study. *Journal of Research in Personality*, *39*(4), 423-442.

Ablow, J. C., & Measelle, J. R. (1993). *Berkeley puppet interview: Administration and scoring system manuals*. University of California.

Arpaci, I., & Kocadag Unver, T. (2020). Moderating role of gender in the relationship between big five personality traits and smartphone addiction. *Psychiatric Quarterly*, *91*(2), 577-585. doi:10.1007/s11126-020-09718-5

Asendorpf, J. B., & Van Aken, M. A. G. (2003). Validity of big five personality judgments in childhood: A 9 year longitudinal study. *European Journal of Personality*, *17*(1), 1-17. doi:10.1002/per.460

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, *38*(1), 25-56. doi:10.1207/S15327906MBR3801

Caspi, A., Harrington, H., Milne, B., Amell, J. W., Theodore, R. F., & Moffitt, T. E. (2003). Children's behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of Personality*, *71*(4), 495-514.

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

Digman, J. (1990). Personality structure: Emergence of the 5-factor model. *Annual Review of Psychology*, *41*(1), 417-440. doi:10.1146/annurev.psych.41.1.417

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399-412.

Eder, R. A. (1990). Uncovering young children' s psychological selves: Individual and developmental differences. *Child Development*, *61*(3), 849-863.

Ehrler, D. J., Evans, J. G., & McGhee, R. L. (1999). Extending big-five theory into childhood: A preliminary investigation into the relationship between big-five personality traits and behavior problems in children. *Psychology in the Schools*, *36*(6), 451-458.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*(3), 343-367.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation*, *17*(1), 3. Available at: https://scholarworks.umass.edu/pare/vol17/iss1/3/

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216-1229. doi:10.1117/12.642204

Hampson, S. E., Edmonds, G. W., Barckley, M., Goldberg, L. R., Joan, P., & Hillier, T. A. (2016). A Big five approach to self-regulation: Personality traits and health trajectories in the Hawaii longitudinal study of personality and health. *Psychology, Health & Medicine*, *21*(2), 152-162. doi:10.1080/13548506.2015.1061676.A

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), 1-55. doi:10.1080/10705519909540118

Irwin, E. C. (1985). Puppets in therapy: An assessment procedure. *American Journal of Psychotherapy*, *39*(3), 389-400.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). The Guilford Press.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Multivariate applications book series. Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275-340). Lawrence Erlbaum Associates Publishers.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320-341. doi:10.1207/s15328007sem1103_2

McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. *Journal of Personality*, *60*(2), 329-361.

McCrae, R. R., & John, O. P. (1992). The five-factor model: Issues and applications. *Journal of Personality*, *60*(2), 175-532. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1635040

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412-433.

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

Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's views of their academic, social, and emotional lives: An evaluation of the self-perception scales of the berkeley puppet interview. *Child Development*, *69*(6), 1556-1576. doi:10.1111/j.1467-8624.1998.tb06177.x

Measelle, J. R., Ablow, J. C., John, O. P., Cowan, P. A., & Cowan, C. P. (2005). Can children provide coherent, stable, and valid self-reports on the big five dimensions? A longitudinal study from ages 5 to 7. *Journal of Personality and Social Psychology*, *89*(1), 90-106. doi:10.1037/0022-3514.89.1.90

Miranda, A., González Dávila, O., Rodríguez Martínez, L. Y., & Sainz Santamaría, J. J. (2019). Marco teórico general del Estudio Longitudinal del Desarrollo de las Niñas y los Niños de Aguascalientes (EDNA). *Programme for Longitudinal Studies, Experiments or Surveys (PANEL),* Working Paper Series No. 001, Centro de Investigación y Docencia Económicas (CIDE), México.

Miranda, A., González Dávila, O., Aguilar, A., Alonso, A., Zizumbo-Colunga, D., Rodríguez-Martínez, Y., & Sainz-Santamaría, J. J. (2020). The aguascalientes longitudinal study of child development: Baseline and first results. Longitudinal and Life Course Studies. *11*(3), 409-423. doi:10.1332/175795920X15858040679570

Resing, W. C. M., Bleichrodt, N., & Dekker, P. H. (1999). Measuring personality traits in the classroom. *European Journal of Personality*, *13*(6), 493-509. doi:10.1002/(sici)1099-0984(199911/12)13:6<493::aid-per355>3.3.co;2-m

Revelle, W. (2019). *Psych: Procedures for psychological, psychometrich, and personality research*. Northwestern University. R package version 1.9.12, https://CRAN.R-project.org/package=psych

Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., Van der Ende, J., Raat, H., Jaddoe, V. W., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2013). Young children's selfreported emotional, behavioral and peer problems: The Berkeley puppet interview. *Psychological Assessment*, *25*, 1273-1285.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Schaefer, C., & O'Conner, K. (1983). *Handbook of play therapy.* Wiley.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120.

Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Prentice-Hall.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*(3), 263-311. doi:10.1207/s15327906mbr2803_1

Wiggins, J. S. (Ed.). (1996). *The five-factor model of personality: Theoretical perspectives*. The Guilford Press.

Woltmann, A. G. (1952). Puppet shows as a psychotherapeutic method. In L. Bender (Ed.), *Child psychiatric techniques*. C. C. Thomas.

Ximénez, C. (2006). A Monte Carlo study of recovery of weak factor loadings in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(4), 587-614. doi:10.1207/s15328007sem1304_5

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*(4), 377-392.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, *40*, 115-148.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123-133.