Preliminary validity for the Big Five Inventory-2 in Brazilian adults

Jeferson Gervasio Pires[1]
Carlos Henrique Sancineto da Silva Nunes[1]
Maiana Farias Oliveira Nunes[1]
Ricardo Primi[2]

[1]Federal University of Santa Catarina, Florianópolis, Brazil
[2]São Francisco University, Campinas, Brazil

Abstract
The Big Five Inventory-2 (BFI-2) assesses personality through the big five factor model model (FFM). This study aimed to verify whether the internal structure of the BFI-2 is corroborated in a Brazilian sample. Participants were 908 cisgender adults, aged between 17 and 93 years, 532 of whom were women. Confirmatory factor analysis and Exploratory structural equation modeling were performed with the BFI-2. The 15 facets of personality resulted in adequate fit and reliability, mainly for the respondents under 60 years of age. We conclude that the BFI-2 presents evidence of preliminary validity based on its internal structure, although further validity studies are required with the instrument, to testify its psychometric quality.
Keywords: Personality; Assessment; Validity; Confirmatory Factor Analysis; Internal Structure

Evidências Preliminares de Validade do *Big Five Inventory-2* em Adultos Brasileiros

Resumo
O *Big Five Inventory-2* (BFI-2) avalia a personalidade no modelo dos Cinco Grandes Fatores (CGF). O presente estudo objetivou verificar se a estrutura interna do BFI-2 é corroborada em uma amostra brasileira. Participaram 908 adultos cisgênero, com idades entre 17 e 93 anos, sendo 532 mulheres. Análise fatorial confirmatória e modelagem de equações estruturais exploratórias foram conduzidas no BFI-2. As 15 facetas de personalidade do BFI-2 resultaram em índices de ajuste adequados, principalmente para os respondentes com idade até 60 anos. Conclui-se que o BFI-2 apresenta evidência inicial de validade baseada em sua estrutura interna, apesar de que novos estudos de validade devem ser conduzidos com o instrumento, com o objetivo de atestar sua qualidade psicométrica.
Palavras-chave: personalidade; avaliação; validade; análise fatorial confirmatória; estrutura interna

Evidencias preliminares de validez del *Big Five Inventory-2* en adultos brasileños

Resumen
El *Big Five Inventory-2* (BFI-2) evalúa la personalidad según el modelo de los Cinco Grandes Factores (CGF). El presente estudio tuvo como objetivo verificar si la estructura interna del BFI-2 se corrobora en una muestra brasileña. Participaron 908 adultos cisgénero, de entre 17 y 93 años, de los cuales 532 eran mujeres. Se realizó un Análisis Factorial Confirmatorio y un Modelo de Ecuaciones Estructurales Exploratorio sobre el BFI-2. Las 15 facetas de personalidad del BFI-2 dieron lugar a índices de ajuste adecuados, especialmente para los encuestados de hasta 60 años. Se concluye que el BFI-2 presenta evidencias iniciales de validez basadas en su estructura interna, aunque se deben ejecutar más estudios de validez con el instrumento para atestiguar su calidad psicométrica.
Palabras clave: Personalidad; Evaluación; Validez; Análisis Factorial Confirmatorio; Estructura Interna

People tend to express consistent behaviors in the different roles they assume in their daily lives, for example, at work and in the family. These functioning patterns (cognitive, behavioral, and emotional), that remain stable in individuals and that can be found in different cultures, are described in the scientific literature as personality traits. One of the most scientifically accepted models for describing personality traits refers to the Five Factor Model (FFM) (John, Naumann, & Soto, 2008; Soto et al., 2008). This perspective comprehends the personality based on five broad traits, which result from the convergence of studies involving factorial analyses carried out independently and conducted in different cultures. This convergence advocates in favor of the hypothesis that these trais are universal (McCrae & Terracciano, 2005).

The *Openness to experience* (O) trait concerns exploratory behavior and the interest in new experiences. People with high scores are curious, imaginative and tend to have unconventional values. *Conscientiousness* (C) indicates people's ability to organize, produce and fulfill goals, being associated with issues such as persistence and motivation. Individuals with high scores are organized, persistent and methodical. *Extraversion* (E)

refers to how communicative, outgoing and assertive people are. High scores are found in sociable, talkative and dominant individuals. *Agreeableness* (A) refers to the depth and quality of interpersonal relationships, in that people with high levels of this trait tend to be generous, kind and compassionate, as well as trusting others. *Neuroticism* (N) indicates the level of people's emotional adjustment. High scores suggest a propensity to experience emotional volatility, depression, excessive worry and anxiety.

Several instruments to assess personality, based on the FFM, have been constructed in recent decades (Soto et al., 2008). These instruments have enabled the development of psychological science as they are used to test hypotheses related to psychological functioning in relation to the aspects covered by personality and related variables. Among the evidence, personality traits have been found to develop throughout the life cycle (Roberts, Martin, & Olaru, 2015), with some traits differentiated between men and women (Kajonius & Johnson, 2018; Marsh et al., 2010; Marsh, Nagengast, & Morin, 2013; Soto & John, 2017). Among these instruments to assess the FFM model, the Big Five Inventory (BFI) can be highlighted.

The BFI is a short inventory, composed of 44 self-report items, widely cited within international literature for use to assess personality traits. This inventory was constructed in 1999 (John & Soto, 2008) in order to measure the five broad personality factors and has been translated and adapted to different languages and cultures in recent years (Rammstedt & John, 2017). Because it is a short form, the BFI is one of the most used instruments for evaluating the FFM, having an internal structure replicable in different contexts (Schmitt et al., 2007; Soto, 2019).

Short versions of personality inventories are useful in contexts in which the application of various instruments occurs (Sleep, Lynam, & Miller, 2020), or in case where it is not possible to respond to many items, as in the case of some older adults. Thus, there has been an extensive effort by researchers to address this issue and develop short scales for personality assessment (Gouveia et al., 2021; Laros et al., 2018; Passos & Laros, 2015). These instruments are also relevant in research involving the differentiation of groups with large samples (as in large scale assessments), in which the objective is not to evaluate individuals, but groups of people. However, there is a significant cost to the use of brief measures in terms of their reduced reliability and validity (Sleep et al., 2020; Soto & John, 2018).

More recently, the BFI authors proposed an updated version of the inventory, composed of 60 self-report items, called the BFI-2 (Soto & John, 2017). Its items are distributed in a hierarchical structure, allowing the assessment not only of the five broad personality factors, but at the facets level, which would be equivalent to sub-factors. In the BFI-2, three facets were proposed for each broad personality factor. Measuring personality at the level of facets allows a detailed assessment of individual functioning, contributing with greater specificity in the professional practice and research in Psychology. Moreover, Soto e John (2017) also proposed two abbreviated forms of the BFI-2— the 30-item BFI-2-S, and the 15-item BFI-2-XS.

In addition, the BFI-2 was developed to allow the control of acquiescence, presenting, for this purpose, pairs of items with "opposite" polarity that evaluate the same attribute (Soto et al., 2008). Acquiescence refers to a response style in which individuals tend to compete or disagree with the statements presented in the test, regardless of the content of the item (Soto et al., 2008; Soto & John, 2017), whose effects are particularly pronounced in samples of children and adolescents, as well as adults with low levels of educational (Soto & John, 2018). This response pattern can distort the reliability, validity, and structure of a psychological measure at both the item and scale levels (Soto & John, 2018), by inflating the scale reliability, and its interitem correlations. To solve this issue, the autors constructed content-balanced scales for the BFI-2 at both the domain and facet levels (Soto & John, 2018). Authors understand that by balancing the number of true-keyed and false-keyed items on each scale would clearly distinguish meaningful personality information from acquiescence. This procedure also allows the control for acquiescence at the item level by centering each individual's item responses around their within-person mean (their mean response to the full item set, without reversing the false-keyed items) (Soto & John, 2018).

It has been documented in the literature that the FFM tend to be easily identified through exploratory factor analysis (EFA) procedures in the instruments constructed to evaluate this model; especially when using the EFA method with random intercept (Aichholzer, 2014; Soto & John, 2017) and varimax rotation. These analyses have concluded that the FFM are orthogonal, that is, they are not particularly correlated with each other, however, they can be subdivided into hierarchically smaller units, referred to as facets, which are correlated with each other.

This complex characteristic of the internal structure of the instruments that assess the FFM makes the performance of confirmatory factor analyses (CFA) in these measurement instruments imprecise (Marsh et al., 2010; 2013). It can be said that these analyses are destined to obtain poor fit indices (Gomes & Gjikuria, 2017; Joshanloo, 2018; Marsh et al., 2013), this being partly due to the assumptions of independence in the data for conducting CFA (Marsh et al., 2013). In theory, the more independent the components of a model are from each other, the better the fit the data modeled in the CFA will present. However, this characteristic is not guaranteed in the instruments that assess the FFM, since the items relate to other items, to facets and to factors, forming a complex and not very independent structure, making confirmation with proper fit indices difficult. Because of this, a strategy adopted in confirmatory studies of the internal structure of FFM instruments is the use of exploratory structural equation modeling (ESEM) procedures (Gomes & Gjikuria, 2017; Joshanloo, 2018; Marsh et al., 2010; 2013).

Confirmatory studies conducted recently with the BFI-2, have concluded that its internal structure is adequate in samples of Russian (Shchebetenko et al., 2019), German (Denissen et al., 2019; Rammstedt et al., 2020) and Slovakian (Halama et al., 2020) adults. Specifically, in the Russian sample, the FFM of the full version of the BFI-2 were extracted with adequate fit indices in the CFA, especially in the model that considered the 5 factors correlated and the model with 15 facets. In the German sample, the FFM of the version of the BFI-2 with 60 items could be extracted in the CFA, with its scores showing adequate reliability both at the level of factors and facets (Denissen et al., 2019). Also, in this population, the FFM of an alternative version of the BFI-2 (Soto & John, 2017), containing 30 items (BFI-2-S), was corroborated through the ESEM procedure (Rammstedt et al., 2020). In turn, in the Slovakian sample, the FFM presented adequate fit indices in the CFA, however, only when each factor was verified in isolation from the others. In these cases, each of the factors was composed of 3 facets, as in the original structural model of the BFI-2.

Together, this knowledge about the FFM obtained from the factor analyses of these instruments has led to the conclusion that their internal hierarchical structure, although complex, is congruent in different contexts and cultures. Despite this progress, a relevant topic to be studied in relation to these instruments concerns the possibility that their facets operationalize content more

focused on the intermediate stage of human development, in that its items are constructed focusing on young people and adults that are studying or professionals and may be insufficiently comprehensive for people aged over 60 (Rossi et al., 2014). As examples, the following items from the BFI-2 can be highlighted: "I have difficulty getting started on tasks", "I am a reliable worker/student" and "I do tasks well, without wasting time". Apparently, it could be that these items receive greater endorsement by younger respondents than by older adults. It can therefore be hypothesized that this could be an effect, for example, of retirement on individual functioning, since, in this stage of life, in most cases, people stop working/studying (Srivastava & Das, 2013). As there is favorable evidence in relation to this hypothesis, further studies need to be conducted with the BFI-2 focusing on these specificities.

In addition to an adequate internal structure in the FFM instruments, it has been observed that the five factors of personality predict a wide range of variables, such as health and academic and professional success (Soto, 2019). This makes the instruments for evaluating FFM useful in several contexts that require assessment of the personality. Considering the relevance of the instruments to assess personality through the FFM model, especially those that allow individual functioning at the level of facets to be assessed, in this study we aimed to: a) verify whether the internal structure of the BFI-2 is congruent in a sample of Brazilian respondents when compared to its original structure, at the level of factors and facets, b) estimate the reliability of the scores generated with the BFI-2 for its factors and facets, and c) identify whether the fit indices and the estimated reliability coefficients remain similar according to the respondent's age group.

Method

Participants

Participants were 908 cisgender adults, aged between 17 and 93 years (*M*=34.81 *DP*=19.12, *Md*=26), 532 of whom were women. The education level of the participants was varied, ranging from people with elementary education to respondents with master's and doctoral degrees. Further details about the sample are presented in Table 1.

Instruments

Sociodemographic questionnaire: Instrument for collecting social and demographic variables, including gender, age and education level.

Table 1.
Characteristics of the Sample

Sample		<=59 years old (*n*=756)		>=60 years old (*n*=152)		Total sample (N=908)	
		f	*%*	*f*	*%*	*f*	*%*
Educational level	Primary	16	2.12	33	34.65	49	7.60
	Secondary	117	15.48	36	23.68	153	16.85
	Tertiary	401	53.04	14	9.21	415	45.70
	Bachelor	134	17.72	46	30.26	180	19.82
	Masters and Doctoral	28	3.70	3	1.98	30	3.30
Gender	Male	344	45.50	32	21.05	376	41.41
	Female	412	54.50	120	78.95	532	58.59
Age	Mean	27.42		71.53		34.81	
	Standard deviation	10.13		7.33		19.12	
	Median	24		70		26	

Big Five Inventory 2 (BFI-2): Scale with 60 self-report items that assess personality traits through the FFM model. This scale allows a more detailed analysis of the personality, as it presents a hierarchical structure composed of 15 facets, organized in the five main factors (Soto & John, 2017). The respondent must indicate their agreement in relation to the statements presented, on a five-point scale, which ranges from 1 "disagree strongly" to 5 "agree strongly". The BFI-2 items are easy to understand and can be applied to individuals with a wide range of ages and levels of education. In the scientific literature, the BFI-2 factors present adequate Cronbach's alpha coefficients regarding estimated reliability, ranging from α=.82 and α=.86.

The BFI-2 has undergone translation and adaptation studies with Brazilian samples by researchers from the Psychological and Educational Evaluation Laboratory (LabAPE) of the Graduate Program in Psychology of Universidade São Francisco. The results of the BFI-2 adaptation study are expected to be published in the near future. This may explain the fact that the present study is a preliminary in terms of validation of the BFI-2 to Brazilian respondents.

Recruitment of participants and data collection

After approval of this study by an ethics committee, participants were recruited in different ways. People were invited from the researchers' contact networks. Services (public and private) with groups aimed at people over 60, located in the metropolitan region of Florianópolis, south of Brazil, were also contacted. The collections took place individually or collectively, through self-application or interview, with the latter strategy being especially used for respondents with low education or for those that reported some difficulty, such as poor vision or insufficient reading skills. This part of the sample (*n*=436) were all residents in the State of Santa Catarina (South region of Brazil) and answered the instruments using pencil and paper. The other part of participants (*n*=469) was invited to participate through social network; they were from different States and regions of Brazil, and answered the instruments through an online form. All the study participants signed the consent form prior to participate.

Participant data were excluded if they exhibited an invalid response style based on elevated scores in any fator of personality (n = 1), for exhibiting a singular response style (e.g., responding to survey with all 1, 3 or 5) on 85% or more of the items (n = 2). Most participants responded to all BFI-2 items (n = 906). In missing cases we imputed data with the mean score of the individual in that personality fator (n = 2). After invalid responders were removed, self-report data were available for 908 individuals. Given the fact that the BFI-2 wasn´t already validated with a Brazilian sample, the total amount of participants in the present study was calculated regarding the number of items in the inventory, in which are expected to have, at least, ten participants for each item (Damásio, 2012).

Data analysis

Initially, the sample was divided into two groups: a) those up to 59 years of age, and b) those aged 60 or over. This age criterion for division was chosen according to the World Health Organization (WHO, 2007), which considers that, in developing countries such as Brazil, those over the age of 60 are considered older adults. When performing this division in the sample, it was possible to identify whether the fit indices and reliability coefficients remained similar according to the respondent's age group. Subsequently, confirmatory factor analysis was carried out with these two age groups and with the total sample.

In view of the complexity in confirmatory studies involving instruments that assess the FFM (Alharyout & Abdullah, 2018; Marsh et al., 2010; 2013), both confirmatory factor analysis (CFA) and exploratory structural equation modeling (ESEM) analysis were conducted. Prior to this, however, the responses collected with the BFI-2 underwent acquiescence control (Soto et al., 2008; Zanon et al., 2018).

The maximum likelihood (ML) method, with the observed information matrix (OIM) technique, was chosen to estimate the models in the CFA (SEM). In these models, the BFI-2 items were considered to be endogenous variables, composing the exogenous variables, that is, each of the FFM. These analyses were performed using the Stata 14® (StataCorp, 2013) software. In turn, in the ESEM stage, the analysis was performed using the Mplus software (7.11). In this case, the maximum likelihood estimator and the oblique geomin rotation method were selected. All assumptions in data were tested, such as normality of distribution, for the use of ML estimator.

In the CFA, the fit was verified for the following models: a) the big five factors in the 60 items of the BFI-2 (original model), b) the 15 facets in the 60 items (original model_2), and c) the big five factors in the version of the BFI-2 with 30 items (Soto & John, 2017) (alternative model), for the two age groups and for the total sample. In the ESEM, the fit of the versions with five factors (original model) and with 15 facets (original model), both referring to the version with 60 items of the BFI-2, were verified, however, only the total sample was considered in this analysis. During the analyses, the fits of the models tested were compared, with and without the acquiescence control.

After the models were specified, their fit indices were estimated. For a good fit, the model needs to present the lowest possible χ^2 value, indicating less residual

in the model, as well as a significance greater than .05, although this index is less important in studies with small samples (Brown, 2015; Stata, 2013). The Root Mean Square Error of Approximation (RMSEA) and Standardized root mean squared residual (SRMR) are expected to have values below .05, although values less than .08 are acceptable. For the Comparative Fit index (CFI) and the Tucker-Lewis index (TLI), values above .80 are acceptable (Stata, 2013). Another indicator of fit for the confirmatory models can be obtained from the ratio between the χ^2 and the degrees of freedom of the models. In these cases, values below 5 indicate a good fit for the model (Kline, 2011).

It is important to stress the information that in the ESEM analysis only the total sample of respondents was considered, whereas in the AFC two subgroups (Age < 59 and Age >= 60) were used, as well as the total sample. However, considering the results reported in the scientific literature (Marsh et al., 2010; 2013), in which the fit indices in the CFA are expected to be poor for the respondents of all age groups, we did not perform the comparison of the ESEM and AFC models, for the total sample, in the present work. In contrast, it was hypothesized that in the ESEM the fit indices would be better and more relevant. Finally, after conducting the CFA and ESEM, the Z scores were calculated for each of the factors and facets of the BFI-2, considering the version of the instrument with 60 items, and Cronbach's alpha coefficients were estimated, also considering the two age groups previously mentioned. The BFI-2 factors and facets were expected to present a reliability equal to or greater than .60 (Nunes & Primi, 2010) for both age groups.

Results

The results of the ESEM and CFA are presented in Table 2. In general, the ESEM procedure resulted in better fit indices for the FFM than the CFA. This result was expected, despite what has already been highlighted in the scientific literature (Marsh et al., 2010; 2013; Rammstedt et al., 2020). Furthermore, the confirmation of the models based on 15 facets seems more evident in the internal structure than the traditional models composed of five broad factors.

Regarding the models tested in the CFA, all the fit indices remained relatively poor, especially from the perspective of the CFI and TLI, although the RMSEA and SRMR were acceptable. The models composed of the 15 facets presented better fit indices than the models

Table 2.
Models Fit for the BFI-2 (N=908)

Models – CFA		RMSEA	CFI	TLI	SRMR	χ^2 (degrees of freedom)	χ^2/df
<=59 y.o. (*n*=756)	5 factors - 60 items	.07	.61	.59	.09	8750.335 (1700), *p*<.005	5.14
	15 facets - 60 items	.05	.77	.75	.07	5628.706 (1605), *p*<.005	3.50
	5 factors (30 items) *	.08	.66	.62	.08	2478.534 (395), *p*<.005	6.27
>=60 y.o (*n*=152)	5 factors - 60 items	.07	.33	.30	.11	3276.285 (1700), *p*<.005	1.92
	15 facets - 60 items	**	**	**	**	**	
	5 factors (30 items) *	.08	.42	.36	.10	851.586 (395), *p*<.005	2.15
Total (N=908)	5 factors - 60 items	.07	.62	.61	.08	9341.214 (1700), *p*<.005	5.82
	15 facets - 60 items	.05	.75	.72	.07	6436.131 (1605), *p*<.005	4.01
	5 factors (30 items) *	.07	.65	.62	.08	2736.959 (395), *p*<.005	6.92

Models – ESEM		RMSEA (90% CI)	CFI	TLI	SRMR	χ^2 (degrees of freedom)	χ^2/df
Total (N=908)	5 factors - 60 items	.05 (.05 – .05)	.75	.70	.04	4101.326 (1480), *p*<.005	2.77
	15 facets - 60 items	.03 (.03 - .02)	.94	.90	.02	1297.459(975), *p*<.005	1.33

*Note. * 30 items-BFI-2-S (Soto & John, 2017). RMSEA= Root Mean Square Error of Approximation, CFI= Comparative Fit index, TLI= Tucker-Lewis index, SRMR= Standardized root mean squared residual, **=model not identified by Stata; y.o.=years old*

composed only of the big five factors, for the respondents under 60 years of age and for the total sample. In these models, the CFI and TLI were marginal in relation to the fit indicated as adequate in the literature, which would be above .80 (Brown, 2015; Stata, 2013).

Additionally, for the participants over 60 years of age, the fit indices were poorer than for the participants aged up to 59 years. For the group of people over 60 years of age, the model composed of 15 facets could not be identified by the statistical software. The CFA also showed that the alternative model of the BFI-2, proposed by Soto and John (2017), composed of 30 items, demonstrated a slightly better fit than the complete model of the BFI-2, with 60 items, for participants of all ages.

Regarding the models tested in the ESEM, the original version of the BFI-2, composed of five broad factors, resulted in fit indices below the adequate level, mainly concerning the CFI and TLI. However, in this model, it was found that the RMSEA and SRMR indices were excellent. In the ESEM, the model composed by the 15 facets of the BFI-2 resulted in all fit indices being excellent.

Overall, the χ^2 proved to be high for all the models tested, with the same being the case for the degrees of freedom. However, considering the ratio between the χ^2 and the degrees of freedom of the models, the previous result can be corroborated, with the models containing 15 facets resulting in adequate fits, as mentioned in the literature (Kline, 2011). The ratio between the χ^2 and the degrees of freedom of the ESEM models also indicated adequate fits.

Regarding the reliability of the BFI-2, estimated through Cronbach's alpha, the total sample obtained adequate coefficients for all the FFM, ranging from α=.82 to α=.90. More information is presented in Table 3.

Considering the reliability of the personality facets, the results for the total sample also remained adequate, between α=.50 and α=.90, even though some facets presented values below that recommended (Nunes &

Table 3.
Reliability coeficients (Alpha) for the BFI-2 factors and facets

Personality traits - BFI-2		17 – 59 years old (*n*=756)	60 – 93 years old (*n*=152)	Total Sample (N=908)
O	Aesthetic Sensitivity	α=.85	α=.66	α=.83
	Creative Imagination	α=.78	α=.52	α=.75
	Intelectual Curiosity	α=.53	α=.30	α=.50
	Open-Mindedness	α= .86	α= .68	α= .82
C	Productivity	α=.73	α=.48	α=.74
	Responsability	α=.67	α=.33	α=.65
	Organization	α=.86	α=.60	α=.83
	Conscientiousness	α= .85	α=.68	α=.86
E	Assertiveness	α=.75	α=.47	α=.71
	Energy Level	α=.73	α=.49	α=.70
	Sociability	α=.89	α=.72	α=.86
	Extraversion	α=.88	α=.76	α= .86
A	Compassion	α=.80	α=.56	α=.78
	Respectfulness	α=.76	α=.61	α=.74
	Trust	α=.69	α=.44	α= .67
	Agreeableness	α= .84	α= .69	α= .84
N	Anxiety	α=.64	α=.39	α=.61
	Depression	α=.85	α=.61	α=.84
	Emotional Volatility	α=.88	α=.57	α=.86
	Negative Emotionality	α= .91	α= .78	α= .90

Primi, 2010). In general, factors are expected to present higher reliability coefficients than facets, as facets have a restricted number of items.

This pattern could also be identified in the reliability of the factors and facets for the respondents up to 59 years of age. For these participants, the magnitude of the reliability of the FFM remained between α=.84 and α=.91, while for the facets the estimated reliability was between α=.53 and α=.86. In contrast, a clear reduction in the magnitude of the reliability of the broad BFI-2 factors was identified for the respondents over 60 years of age. For these participants, the estimated reliability of the factors remained between α=.68 and α=.76. The drop was most strongly observed at the level of the personality facets, in which the magnitudes of reliability remained between α=.30 and α=.72. Intellectual curiosity, Responsability and Anxiety were the facets in which the reliability was most affected.

Discussion

This study sought evidence that the internal structure of the Big Five Inventory (BFI-2) was congruent in a Brazilian sample. In addition, the reliability of the factors and facets of personality was also estimated, and whether the internal structure of the BFI-2 would show different fit indices in relation to the participant's age group was investigated.

Regarding the internal structure of the BFI-2, despite the expected poor fit indices obtained in the confirmatory models (Gomes & Gjikuria, 2017; Joshanloo, 2018; Marsh et al., 2013; 2013), the FFM could be identified in the total sample currently assessed. This result indicates that the internal structure of the BFI-2 was invariant in the Brazilian sample tested, in relation to the American (Soto & John, 2017), Russian (Shchebetenko et al., 2019), German (Denissen et al., 2019; Rammstedt et al., 2020) and Slovakian (Halama et al.,

2020) samples. This suggests that the BFI-2 adequately assesses the factors for which it was developed and, consequently, confirms the hypothesis of universality of the FFM (McCrae & Terracciano, 2005) to the extent that, in the Brazilian sample, the inventory obtained the same internal structure extracted in samples from other cultures.

The complex hierarchical structure of the FFM could be observed in the data, as expected, resulting in poor fit indices for the models tested (Gomes & Gjikuria, 2017; Joshanloo, 2018; Marsh et al., 2013), and in the respective χ^2 and degrees of freedom. It should be remembered that the degrees of freedom tend to be higher in models with more items, since the number of parameters in the model is higher. This, in a way, explains why the more parsimonious models obtained better fits, compared to the models with more items. However, the simpler models come with the cost of decreasing the reliability of the factors (Soto & John, 2017), in such a way that the researcher or professional that intends to evaluate the personality with some version of the BFI-2 should select between reliability or agility, choosing to use the most appropriate version of the inventory for the context.

Regarding the fit indices obtained in the confirmatory analyses, greater complexity was identified in the confirmation of the five broad personality factors of the BFI-2, than in its 15 facets. In addition to the fit indices, this information was corroborated in the relationship between χ^2 and the degrees of freedom of the models tested (Kline, 2011). This pattern in the findings is in line with the results obtained in other confirmatory studies with the BFI-2 (Denissen et al., 2019; Halama et al., 2020; Rammstedt et al., 2020; Shchebetenko et al., 2019; Soto & John, 2017). This was expected in the data involving personality, since the items, facets and factors are associated with each other, a characteristic that makes obtaining an appropriate fit for the FFM personality model complex (Gomes & Gjikuria, 2017; Joshanloo, 2018; Marsh et al., 2013).

Associated with this, the performance of the acquiescence control in the items of the BFI-2, prior to conducting the confirmatory procedures, contributed to the improvement of the fit indices in the models tested (Soto & John, 2017). This was confirmed by the clear reduction in the χ^2, degrees of freedom, RMSEA and SRMR when comparing the fit indices of the models tested in the present study, with and without acquiescence control. This finding is also in line with the indication for the use of techniques that involve

the random intercept (Aichholzer, 2014), aiming to obtain confirmatory models in the field of personality with more adequate fits. Accordingly, the control of acquiescence in items, prior to confirmatory analyses, is a valid strategy for confirming models of the BFI-2 (Rammsteadt et al., 2020; Soto & John, 2017).

The short version of the BFI-2, proposed by Soto and John (2017), consisting of 30 items, resulted in more adequate fit indices than the original model of the BFI-2, with 60 items. This result is in line with the finding by Rammsteadt et al. (2020), who identified the same pattern. This is partly because the number of model parameters is reduced and, consequently, the associations between items, facets and factors are weakened, which makes the model components less dependent.

In addition to an adequate internal structure, the BFI-2 showed that its scores tend to be stable considering people aged between 18 and 59 years, since the magnitude of the reliability of the broad factors and facets were excellent for the sample up to 59 years of age. This finding is broadly in line with that observed by Denissen et al. (2019), comprising initial evidence for the instrument's validity, based on its internal structure. However, it should be highlighted that both the internal structure and the reliability of the BFI-2 were strongly affected by the respondent's age. This indicates that some items may impact people differently at different stages of development, not working so well for people aged 60 or over, which is in line with the hypothesis that personality instruments, due to containing content more focused on the intermediate stage of development, end up being insufficiently comprehensive for people aged over 60 (Rossi et al., 2014), with some items of the BFI-2 tending to receive greater endorsement by younger respondents.

As the age of the respondent increased, there was a downward trend both in the confirmatory fit indices of the BFI-2 models and in the estimated reliability of their scores. Accordingly, although possible impacts of the age variable on personality scores can be identified, this result reflects the normative changes in personality that occur throughout the life cycle (Marsh et al., 2013; Roberts & Mroczek, 2008; Roberts et al., 2015; Soto & John, 2012). An alternative hypothesis for this result is that the reduction in the reliability of the factors occurred due to the low level of education of most of the participants over 60 years of age, which could have led to difficulties in understanding both the BFI-2 items and the activity as a whole. This hypothesis, therefore, needs to be investigated in the future.

Although, apparently, the BFI-2 has shown a better fit for people up to 60 years of age, these characteristics of the instrument should not be seen as completely negative. It is important that the items of the personality tests can effectively differentiate respondents in the adult stage of development. Firstly, because adults are the majority of the national population (Brazilian Institute of Geography and Statistics, 2018) and secondly, because the practical utility of these instruments is, in one way or another, relevant for personnel selection, research and diagnosis, contexts that typically involve individuals in the intermediary stage of development.

However, instruments are needed that can be used equally with different populations so that studies and interventions involving personality traits can take place adequately, considering the characteristics of all stages of development. Therefore, further analysis of the differential item functioning (DIF) of the BFI-2 is needed in order to identify whether the items favor and disadvantage, to the same extent, respondents in different stages of development guaranteeing equality in the evaluation. If applicable, adjusting the wording of some items of the BFI-2 for respondents over 60 years of age should be considered in the future, or even the creation of other instruments directed toward this target public.

Despite the findings obtained in the confirmatory analysis of the present study, which indicate that the BFI-2 presented evidence of validity based on its internal structure, it is essential that analyses of the internal structure of the BFI-2 be performed with respondents of different age groups. It is also suggested that a future ESEM multigroup procedure is carried out, not only with the total sample, as occurred in the present study. Furthermore, it is recommended that future studies should seek other sources of evidence of validity, such as through the association with external variables. The congruence in the associations between the BFI-2 scores and the scores in instruments that assess resilience, anxiety, mindfulness, and post-traumatic stress could be investigated.

Future studies could also examine the stability of the BFI-2 scores in the long term, conducting test-retest association studies. Moreover, a limitation of the paper is that only Cronbach's alpha was used as a measure of reliability. The alpha depends on the number of items and does not take factor loadings into account. Therefore, it is suggested that future accuracy studies include other coefficients, such as Guttmam's lambda 2. Also, convergent validity correlating the scores of the BFI-2 with scores of the Neo Pi-R would be of

importance. Only by guaranteeing the quality of psychological instruments will it be possible to effectively advance the scientific and instrumental development of Personality Psychology. Another possible future study with the BFI-2 is to compare scores, in the broad personality factors and in the facets, between male and female respondents. Some authors (Kajonius & Johnson, 2018; Marsh et al., 2010; 2013; Soto & John, 2017) have highlighted differences in this aspect. If differences are identified in these analyses, then the development of separate normative tables for the interpretation of scores would become vital for the adequate use of the instrument.

Final considerations

The internal structure of the BFI-2 proved to be congruent in two relatively distinct samples of Brazilian adults, especially at the level of the facets, while maintaining the reliability in the broad factors and the personality facets. The internal hierarchical structure of the FFM, despite being complex, is congruent in different cultures, which makes the use of these tools relevant in different contexts and with diverse populations. Despite this, it is necessary to clarify both the possible divergences between the structure of the FFM in older adults and young adults, as well as the relevance of the BFI-2 items for people aged over 60 years. Accordingly, a limitation of the present study refers to the low number of people over 60 in the sample, with less than 200 cases for analysis. This characteristic weakens the fit indices of the models currently tested involving this population.

Another interesting aspect is that short versions of personality inventories are useful in contexts in which different instruments are applied, or in situations where to it is not possible to respond to many items, as in the case of some older adults. The instruments that evaluate the FFM currently available to psychologists are composed of many items, for example, the Revised Neo Personality Inventory, with 240 items, and the Factorial Personality Battery (FPB), with 126 items; although the current effort by researchers to develop short scales for personality assessment (Gouveia et al., 2021; Laros et al., 2018; Passos & Laros, 2015). Therefore, considering the number of items and facets, the BFI-2 is different from the Neo PI-R and the FPB, as it allows the evaluation of 15 facets of personality through only 60 items, even though some of its facets present reliability below the minimum recommended.

Considering all, we conclude that the BFI-2 presents initial evidence of validity based on its internal structure for assessing personality of Brazilian respondents, however, further validity studies are required with the instrument.

References

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality.* 53: 1–4. http://dx.doi.org/10.1016/j.jrp.2014.07.001

Alharyout, A. O., & Abdullah, A. H. (2018). Confirmatory factor analysis of big five personality factors (CFA-BFPF). *British Journal of Education, Learning and Development Psychology.* 1(1): 29-36.

Brazilian Institute of Geography and Statistics. (2018). Projections and Estimates of the Brazilian Population. Rio de Janeiro: Studies and research. Retrieved from: https://www.ibge.gov.br/apps/populacao/projecao/ .

Brown, T. (2015). *Confirmatory factor analysis for applied research.* (2nd ed.). New York: The Guilford Press.

Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica,* 11(2), pp. 213-228

Denissen, J. J. A., Geenen, R., Soto, C. J., John, O. P., & van Aken, M. A. G. (2019). The Big Five Inventory–2: Replication of Psychometric Properties in a Dutch Adaptation and First Evidence for the Discriminant Predictive Validity of the Facet Scales. *Journal of Personality Assessment,* 1–16. doi:10.1080/00223891.2018.1539004

Gomes, C. M. A., & Gjikuria, E. (2017). Comparing the ESEM and CFA approaches to analyze the Big Five factors. *Avaliação Psicológica.* 16(3): 261-267. DOI: http://dx.doi.org/10.15689/ap.2017.1603.12118

Gouveia, V. V, de Carvalho Rodrigues Araújo, R., Vasconcelos de Oliveira, I. C., Pereira Gonçalves, M., Milfont, T., Lins de Holanda Coelho, G., Santos, W., de Medeiros, E. D., Silva Soares, A. K., Pereira Monteiro, R., Moura de Andrade, J., Medeiros Cavalcanti, T., da Silva Nascimento, B., & Gouveia, R. (2021). A Short Version of the Big Five Inventory (BFI-20): Evidence on Construct Validity. *Revista Interamericana De Psicología/Interamerican Journal of Psychology,* 55(1), e1312. https://doi.org/10.30849/ripijp.v55i1.1312

Halama, P., Kohút, M., Soto, C. J., & John, O P. (2020). Slovak Adaptation of the Big Five Inventory (BFI-2): Psychometric Properties and Initial Validation. *Studia Psychologica,* 62(1). 74-87 https://doi.org/10.31577/sp.2020.01.792

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy. Em *Handbook of personality: Theory and research* (pp. 114-158). Guilford Press New York, NY.

Joshanloo, M. (2018). Evaluating the Factor Structure of the MIDI Personality Scale Using Exploratory Structural Equation Modeling. *Japanese Psychological Report.* 60(3): 162-169. https://doi.org/10.1111/jpr.12186

Kajonius, P. & Johnson, J. (2018). Sex differences in 30 facets of the five factor model of personality in the large public (N=320,128). *Personality and Individual Differences.* 129: 126-130. https://doi.org/10.1016/j.paid.2018.03.026

Kline, R. B. (2011). Principles and Practice of Structural Equation Modeling. The Guilford Press. 3rd edition. New York.

Laros, J. A., Peres, A. J. S., Andrade, J. M. et al. (2018) Validity evidence of two short scales measuring the Big Five personality factors. *Psicologia: Reflexão e Crít*ica, *31.* https://doi.org/10.1186/s41155-018-0111-2

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*(3), 471-491. http://dx.doi.org/10.1037/a0019227

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

McCrae, R. R., & Terracciano, A., & 78 Members of the 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*(3), 547-561. http://dx.doi.org/10.1037/0022-3514.88.3.547

Nunes, C. H. S. S., & Primi, R. (2010). Aspectos técnicos e conceituais da ficha de avaliação dos testes psicológicos. Em Conselho Federal de Psicologia - CFP (Org.): *Avaliação psicológica: diretrizes na regulamentação da profissão* (pp. 101-128). Brasília: CFP.

Rammstedt, B., Danner, D., Soto, C. J, & John, O. P. (2020). 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. 149-161. https://doi.org/10.1027/1015-5759/a000481

Passos, M. F. D., & Laros, J. A. (2015). Construção de uma escala reduzida de Cinco Grandes Fatores de personalidade. *Avaliação Psicológica*, *14*(1), 115-123. Recuperado em 15 de junho de 2021, de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712015000100014&lng=pt&tlng=pt.

Rammstedt, B., & John O.P. (2017) Big Five Inventory. In: Zeigler-Hill V., Shackelford T. (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham https://doi.org/10.1007/978-3-319-28099-8

Roberts, R. D., Martin, J. E., & Olaru, G. (2015). A Rosetta Stone for Noncognitive Skills Understanding, Assessing, and Enhancing Noncognitive Skills in Primary and Secondary Education. ASIA Societies.

Rossi, G., Van den Broeck, J., Dierckx, E., Segal, D. L., & Van Alphen, S. P. J. (2014). Personality assessment among older adults: the value of personality questionnaires unraveled. *Aging & Mental Health*, 18(8): 936–940. https://doi.org/10.1080/13607863.2014.924089

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*, 173–212. https://doi.org/10.1177/0022022106297299

Shchebetenko, S., Kalugin, A., K. Mishkevich, A. M., Soto, C. J., & John, O. P. (2020) Measurement Invariance and Sex and Age Differences of the Big Five Inventory–2: Evidence From the Russian Version. *Assessment* 27:3, 472-486. https://doi.org/10.1177/1073191119860901

Sleep, C. E., Lynam, D. R., & Miller, J. D. (2020). A Comparison of the Validity of Very Brief Measures of the Big Five/Five-Factor Model of Personality. *Assessment*, 28(3). http://dx.doi.org/10.1177/1073191120939160

Soto, C. J. (2019). How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, 30(5), 711–727. https://doi.org/10.1177/0956797619831612

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1): 117-143. http://dx.doi.org/10.1037/pspp0000096

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. https://doi.org/10.1037/0022-3514.94.4.718

Srivastava K, & Das R C. (2013). Personality pathways of successful ageing. Industrial *Psychiatry Journal. 22*(1):1-3. doi: 10.4103/0972-6748.123584

StataCorp. (2013). Stata: Release 13. Statistical Software. College Station, TX: StataCorp LP.

Zanon, C., Lessa, J. P. A., & Dellazzana-Zanon, L. L. (2018). Aquiescência em autorrelatos de personalidade: uma comparação de métodos. *Avaliação Psicológica*. 17(4), 428-438. https://dx.doi.org/10.15689/ap.2018.1704.3.03

World Health Organization (WHO) (2007). Global report on falls prevention in older age. Switzerland. Recuperado em 12 de setembro de 2021, de https://www.who.int/ageing/publications/Falls_prevention7March.pdf

About the authors:

Jeferson Gervasio Pires é Psicólogo. Mestre e Doutor em Psicologia pela Universidade Federal de Santa Catarina. Atualmente, em estágio pós-doutoral no Programa de Pós-Graduação em Psicologia da Universidade Federal de Santa Catarina- Florianópolis, Santa Catarina, Brasil.
ORCID: https://orcid.org/0000-0001-8121-8922
E-mail: jefersongp@gmail.com

Carlos Henrique Sancineto da Silva Nunes é Psicólogo, Mestre e Doutor em Psicologia pela Universidade Federal do Rio Grande do Sul. Tem atuado na criação de testes para avaliação da personalidade e adaptação de testes cognitivos. É professor Associado da Universidade Federal de Santa Catarina- Florianópolis, Santa Catarina. Brasil.
ORCID: https://orcid.org/0000-0002-7769-6937
E-mail: carloshnunes@mac.com

Maiana Farias Oliveira Nunes é Psicóloga. Mestra e Doutora em Avaliação Psicológica pela Universidade São Francisco. Tem experiência na área de Psicologia, com ênfase em Orientação Profissional, atuando principalmente nos seguintes temas: avaliação psicológica, orientação profissional, testes psicológicos, interesses profissionais e personalidade. É professora adjunta da Universidade Federal de Santa Catarina- Florianópolis, Santa Catarina- Brasil.
ORCID: https://orcid.org/0000-0003-4891-5982
E-mail: maiananunes@mac.com

Ricardo Primi é Psicólogo. Doutor em Psicologia Escolar e do Desenvolvimento Humano pela Universidade de São Paulo com parte desenvolvida na Yale University (EUA) sob orientação de Robert J. Sternberg. Coordenador do Laboratório de Avaliação Psicológica e Educacional (LabAPE). É Professor Associado do Programa de Pós-Graduação em Psicologia da Universidade São Francisco- Campinas, São Paulo- Brasil.
ORCID: https://orcid.org/0000-0003-4227-6745
E-mail: rprimi@mac.com

Contact:

Jeferson Gervasio Pires
Campus Reitor João David Ferreira Lima – Trindade. Prédio novo CFH, bloco E, 3º andar
Florianópolis-SC, Brasil
Caixa Postal 476
CEP: 88040-900