

 <p>ESCOLA DE CIÊNCIAS DA SAÚDE E DA VIDA</p>	<p>PSICO</p> <p>Psico, Porto Alegre, v. 55, n. 1, p. 1-12, jan.-dez. 2024 e-ISSN: 1980-8623 ISSN-L: 0103-5371</p>
<p>http://dx.doi.org/10.15448/1980-8623.2024.141218</p>	

SEÇÃO: ARTIGO

Differential Item Functioning Analysis on the Big Five Inventory-2 with Brazilian Adults

Análise do funcionamento diferencial do Big Five Inventory-2 no Brasil

Análisis del funcionamiento diferencial en el Big Five Inventory-2 en Brasileños

Jeferson Gervasio Pires¹

orcid.org/0000-0001-8121-8922
jefersongp@gmail.com

Carlos Henrique Sancineto da Silva Nunes¹

orcid.org/0000-0002-7769-6937
carlos.nunes@ufsc.br

Maiana Farias Oliveira Nunes¹

orcid.org/0000-0003-4891-5982
maiana.nunes@ufsc.br

Received on: July 28th, 2021.

Approved on: April 23th, 2022.

Published on: Oct 11th, 2024.

Abstract: Many instruments are available for assessing the Big Five Factors of personality. However, possible limitations to these tools can be highlighted in the assessment of older adults. In this study, we verified the presence of response bias (DIF) in the Big Five Inventory-2 (BFI-2) due to the respondent's age. A total of 808 adults aged between 17 and 93 years old ($M=33.08$, $SD=19.46$) answered the BFI-2. Fifteen items of the BFI-2 were identified with DIF, nine regarding Conscientiousness, four regarding Agreeableness, and two of Neuroticism. We observed a balanced amount of DIF in the BFI-2, either favoring younger people or older people. The items functioned similarly for people under 40 years old and over 60 years old, suggesting justice in assessing respondents' personalities in different life cycle stages. Despite this, some items deserve care during interpretation when answered by older people, and others would benefit from future revision of their wording.

Keywords: personality, assessment, elderly

Resumo: Diversos instrumentos estão disponíveis para avaliação dos Cinco Grandes Fatores de personalidade. Porém, possíveis limitações dessas ferramentas podem ser destacadas na avaliação de pessoas idosas. Neste estudo, verificamos a presença de viés de resposta (DIF) no *Big Five Inventory-2* (BFI-2) em razão da idade do respondente. Participaram 808 adultos, com idades entre 17 e 93 anos ($M=33,08$, $SD=19,46$), que responderam o BFI-2. Foram identificados 15 itens com efeito de DIF: nove de Conscienciosidade, quatro de Amabilidade e dois de Neuroticismo. Observou-se uma quantidade equilibrada de DIF no BFI-2, ora favorecendo pessoas mais jovens, ora as mais velhas. Os itens funcionaram similarmente para pessoas com menos de 40 e com mais de 60 anos, sugerindo justiça na avaliação da personalidade de respondentes de diferentes estágios do ciclo vital. Apesar disso, alguns itens merecem cuidado na interpretação quando respondidos por pessoas idosas, e outros poderão ter sua redação revisada em pesquisas futuras.

Palavras-chave: personalidade, avaliação, idosos

Resumen: Hay varios instrumentos disponibles para la evaluación de los Cinco Grandes Rasgos de personalidad. Sin embargo, algunas limitaciones en estas herramientas pueden destacarse en la evaluación de los mayores. En este estudio, verificamos la presencia de sesgo de respuesta (DIF) en el Big Five Inventory-2 (BFI-2) debido a la edad del encuestado. Los participantes fueron 808 adultos, de entre 17 y 93 años ($M=33,08$, $SD=19,46$), que respondieron al BFI-2. Se identificaron 15 ítems con DIF, nueve se refieren al rasgo Conscienciosidade, cuatro al Amabilidad, y dos de Neuroticismo. Cinco ítems en favor de los encuestados mayores de 60 años, y cuatro ítems que favorecen a los menores de 40 años. Sin embargo, el BFI-2 demostró ser un instrumento justo para evaluar a las personas de diferentes grupos de edad, aunque algunos ítems merecen cuidado en la interpretación cuando contestados por personas mayores, y otros ítems podrán ser revisados en estudios futuros.

Palabras clave: personalidad, evaluación, mayores



Artigo está licenciado sob forma de uma licença
[Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

¹ Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil.

Many instruments are currently available for the assessment of the five-factor model of personality (FFM) in the general population, such as The Neo-Personality Inventory - Revised (NEO PI-R), the *Big Five Inventory-2* (BFI-2), and the Factorial Battery of Personality (FBP). However, because these instruments were developed with reference to the general population, typically adults, part of their items evaluate behaviors that involve the scope of work and studies. This aspect may limit these tools when evaluating older people. Age-neutral instruments are different from age-specific instruments (Rossi et al., 2014). The former were not designed for a specific population but aimed at the general population; for example, they are not intended for people aged 60 years and over, despite having included elderly people in their normative sample. On the other hand, instruments with a specific age group are designed with a focus on a certain age group, in which the attributes of the items retain some affinity with this target audience. Because the instruments for evaluating the FFM currently available are age-neutral, we can assume possible negligence in relation to the fact that some content may become more or less relevant with the advance of age.

McAdams and Olson (2010), for example, indicate that goals in life vary with advancing age. For the authors, individuals between the ages of 18 and 29 are more concerned with education, intimacy, career, and friendships. At the age of 30-50, people are focused on the future of their children, security about what they have acquired, involvement with society, and improving their community. In turn, from the age of 50, people tend to focus on retirement and leisure; giving meaning to life and accomplishing things that were not previously done. In view of these variations in life goals over time, it may be thought that the behavior described in the content of some items may prove more or less appropriate for some respondents depending on their age, indicating limitation in the assessment of some age group.

It is prudent to consider the possibility that some items of these instruments operationalize

content more related with the intermediate stage of development, focusing on young people and adults (up to 40 years old) in their studies or professional activities, and may be insufficiently comprehensive with people over 60 years old (Rossi et al., 2014). Examples of BFI-2 are: items 23 "*I have trouble beginning tasks*" and 38 "*I do the tasks well and without wasting time*". As examples of the NEO PI-R, there are items 77 "*I work slowly but persistently*" and 102 "*I usually I let others talk in meetings*". As an example of FBP, item 28 "*I can handle many tasks at once*" and item 44 "*I have difficulty adapting to work that involves a fixed routine*". However, given that many older people do not work or study, or have never done so, if they are submitted to the FFM assessment with these inventories, they will answer items that involve these contexts, which can be a problem to the assessment of their personality traits.

Despite this limitation, the instruments for the FFM assessment have enabled the development of psychological science to the extent that they are used to test hypotheses related to individual functioning in terms of personality traits and related variables (Soto, 2019). We highlight the evidence that personality traits develop throughout the life cycle (Nye & Roberts, 2019), adapting to the demands of the new stage that the individual accesses. It has been reported that people become more confident, calm, and mature over time (McAdams & Olson, 2010), less inclined to mood swings and negative emotions (Specht et al., 2011), more responsible, caring, and less impulsive, engaging less in risk situations (McAdams & Olson, 2010). In addition, the scores of these instruments have allowed us to associate between individual functioning and a series of relevant variables throughout the life cycle (Soto, 2019), such as cognitive functions and mortality (Yoneda et al., 2022). These evidence highlighted with the instruments that evaluate the FFM characterize the potential of these inventories in screening clinical demands, elaborating interventions that promote healthy development throughout the life cycle, and producing knowledge in Developmental and Personality Psychology.

Even with the advances in knowledge about personality that these instruments have enabled, it is necessary to recognize that construction studies and validity evidence of these inventories involve, in most cases, samples of adolescents and adults, respondents who are in the productive phase of life, a fact that could incur injustice in the assessment. One consideration in this regard refers to the possibility that its items can generate information that has different interpretations in the phases of the life cycle. For example, item 8 of BFI-2: *"I tend to be lazy"*. Whereas in relation to a young adult, we could infer that a high score would indicate an inadequate profile for some job vacancies; on the other hand, for an elderly person, a high score in this item could mean a "positive" characteristic, informing how much the subject seeks to "live in a more relaxed way". Following this reasoning, other BFI-2 items that could present this effect would be 18 *"I am organized, I like to keep things in order"*, and 23 *"I have trouble beginning tasks"*. With older people, especially those with motor difficulties, even everyday tasks can be very complex, such that people may have difficulties gathering the objects necessary to perform the task because they are scattered around the house, or because they do not remember where they are. The same can be said for item 77 of the Neo-PI-R, *"I work slowly but persistently"*, which can receive different interpretations, depending on the stage of development in which the respondent is found.

However, it is worth noting that these limitations of the instruments that evaluate FFM are not exactly negative. It is important that the instrument items effectively differentiate respondents in the intermediate stage of development because adults still constitute most of the national population (IBGE, 2018). Additionally, the practical utility of these instruments, in one way or another, is more relevant in the selection of personnel, research, and diagnosis, contexts that typically involve individuals in the intermediate stage of development. However, we cannot neglect other potential respondents to these inventories, such as the elderly, if we aim for justice in the assess-

ment processes.

Analyzing the form in which the items are written, we cannot fail to consider that some of them may receive greater endorsement from respondents up to 40 years of age, given that some of these items clearly address the scope of work and studies. Consequently, some of these items may prove to be obsolete for people aged 70-80-90 years, and the instruments may even become irrelevant when assessing people who are not studying or working, as is often the case with older people. This makes us hypothesize the presence of differential functioning of some items in favor of younger respondents, highlighting possible injustices in the assessment of respondents who are not in the school or work phase (productive stage) of life, an issue that deserves to be scientifically investigated.

The relevance of investigating this hypothesis is given, mainly, by the estimate of the Instituto Brasileiro de Geografia e Estatística (IBGE, Brazilian Institute of Geography and Statistics). In 2060, the Brazilian population will be composed mostly of people over 60 years old, that is, elderly people (IBGE, 2018). This prediction even invalidates the recently mentioned perspective that personality instruments should consider being able to differentiate only people in the productive stage of life. Also, in the coming years, the life expectancy of Brazilians will increase (IBGE, 2018) since, over time, older people will be healthier, have higher professional qualifications and schooling, better general living conditions, and more time available for their personal and professional achievements, or to rest. These predictions make us reflect that the promotion of health, work, and education of people over 60 years old will be an increasingly emerging theme in the coming years in Brazil and, to follow this trend, the instruments, theories, and techniques used by psychology professionals should be adapted.

As a consequence, the IBGE predictions denote the need to conduct research verifying the impact of sociodemographic variables on the scores of personality tests. The analysis of the differential item functioning (DIF) is one of the techniques of the Item

Response Theory most used to identify response biases in instruments for psychological and educational assessments. DIF analyses investigate signs of interaction between the patterns of responses given to the instrument and sample characteristics (Linacre, 2010), indicating how many groups of respondents with the same level of the latent trait measured in the test have different responses to the items (Linacre, 2010). The literature considers the fact that a good test is not one without DIF but one in which the item-DIFs are balanced, facilitating and/or hindering, to a similar magnitude, members of all tested groups (Linacre, 2010). This balancing is known as "justice" and occurs because one item-DIF cancels another (Linacre, 2010; Teresi, 2006).

In the case of identifying item-DIFs by comparing adult people (18 to 40 years) and elderly people (60 years onwards), we would indicate that the instrument does not work in an equivalent way and may be unfair to a specific group. In practice, the presence of the DIF means that even if the item has its usual difficulty for one group, it can be easier or more difficult than usual for people in the other group. Furthermore, item-DIFs may appear because of real differences in the manifestation of the latent trait measured between the groups. In these cases, they should not be considered problematic (Karami, 2012). Still, it is necessary to consider that having fair instruments with respondents with different characteristics is essential for studies and interventions involving personality, respecting the specificities of the stages of development and, above all, ensuring access to quality psychological services for people of all phases of the life cycle (APA, 2014).

These questions involving possible response biases in the items of the FFM should be tested by researchers in the area of psychological assessment in such a way that studying the differential functioning in the items of these personality instruments could test whether this hypothesis is defensible or not. When conducting a search in some international scientific bases, we identified only one study that addressed this question (Van den Broeck et al., 2012), which denotes both the lack of investment in research seeking to answer it and how little we know about the

impact of age on the response to personality tests.

Van den Broeck et al. (2012) compared the NEO PI-R responses of 411 adult people (18-40 years old) and 434 elderly people (65-92 years old). The results indicated the presence of 12 large item-DIFs, generating disadvantage for younger people. In summary, the authors identified four items-DIF of Conscientiousness, four of Agreeableness, two of Openness, and two of Extraversion. The researchers concluded that the NEO PI-R is a fair instrument since its item-DIFs are balanced. However, despite mentioning evidence of NEO PI-R age neutrality, it is not clear how participants aged 70-80-90, who do not study or work, reacted and responded to the items that evoke the terms *work* and *study*, as discussed above. Thus, even in the face of the evidence mentioned above, one cannot fail to consider the possibility of injustice in relation to these items of the NEO PI-R.

Given the need for a fair instrument for assessing the personality of respondents in the different stages of development, in the present study, we verified the presence of bias and response trends in the *Big Five Inventory-2* (BFI-2) according to the age of the respondent. The objective of this study is to test the hypothesis that some BFI-2 items are biased according to the age of the respondent, especially those related to productivity, work, and studies, favor respondents up to 40 years old since they usually work and/or study. In practice, this study may bring both initial evidence of the age neutrality of the BFI-2 and indicate how fair the inventory is to evaluate the FFM of respondents belonging to different stages of the life cycle.

Method

Participants

A total of 808 adults, aged between 17 and 93 years ($M=33.08$, $SD=19.46$, $Md=24$), of which 357 were men. Most of the sample coursed higher education ($n=395$), followed by those who had completed higher education ($n=186$), those with complete high school ($n=138$), with elementary education ($n=65$) and those with complete *stricto sensu* specialization ($n=24$). Table 1 shows other information regarding the sample.

Table 1 – Characteristics of the participating sample

Characteristics of participants		17-40 years (n=656)		60-93 years (n=152)	
		f	%	f	%
Schooling	Elementary	12	1.83	53	34.88
	High School	102	15.55	36	23.68
	Incomplete Higher Education	381	58.08	14	9.21
	Complete Higher Education	140	21.34	46	30.26
	Specialization	21	3.20	3	1.98
Gender	Male	325	49.54	32	21.05
	Female	331	50.46	120	78.95
Type of data collection	On site	386	58.84	139	91.45
	Online	270	41.16	13	8.55
Age	Mean	24.17		71.53	
	Standard deviation	5.66		7.33	
	Median	24		71	

Instruments

Sociodemographic and profile questionnaire. Instrument with questions to verify social and demographic variables: gender, age, marital status, and schooling.

Big Five Inventory 2 (BFI-2). The inventory of the Big Five Factors-2 is an instrument created to hierarchically assess personality. It is a reformulated version of the BFI, containing 60 self-report items (Soto & John, 2017). The BFI-2 allows a more detailed analysis of personality than that obtained by the original BFI since the BFI-2 has two strengths: 1) its internal structure is conceptually coherent and empirically robust, with three facets in each major personality factor, and 2) it seeks to minimize the influence of the acquiescence effect, which concerns the tendency of respondents to agree or disagree with the items, regardless of their content (Soto & John, 2017).

To answer the inventory, the respondent must indicate the agreement in relation to the statements presented on a five-point scale, which varies from 1 "It has nothing to do with me" to 5 "It has everything to do with me". The BFI-2 items are easy to understand and can be applied to individuals of a wide variety of ages and schooling. In addition, its factors have

obtained adequate accuracy, with Cronbach's alpha coefficients between =0.82 and =0.86 (Soto & John, 2017), the same pattern identified in the present study.

The version of BFI-2 used in the present study was translated and adapted from the English version (Soto & John, 2017) with the permission from the authors, by the team: Dr. Ricardo Primi (USF), Dr. Carlos H. S. Nunes (UFSC), Dr. Nelson Hauck Filho (USF), Dr. Mauricio Hass Bueno (UFPE), Dr. Cristian Zanon (UFRGS), Dr. Lucas de Francisco Carvalho (USF), and Dr. Rodolfo Ambiel (USF) of The Graduate Program in Psychology at University of San Francisco. It is worth mentioning that we conducted a study of initial evidence of validity of the BFI-2 based on its internal structure, in which adequate confirmatory indices were obtained (Pires et al., in press).

Participant recruitment and data collection

After approval of this research by the Ethics Committee (CAAE 00811018.5.0000.0121) of a University in southern Brazil, participants were recruited in different ways. We invited people from the researchers' contact network and contracted services (public and private) with groups aimed at people aged 60 and over, located in the

metropolitan region of a municipality in southern Brazil. The collections occurred individually or collectively through self-application or interview, especially for respondents with low schooling levels, or for those who reported some difficulty, such as low vision or insufficient level of reading. Part of the sample ($n=525$) answered the instruments in pencil and paper format, and the others ($n=283$) answered via an *online* form. All participants signed a free and informed consent form before answering the instruments.

Data analysis

Initially, the participants were divided into two groups in relation to their age: those aged 17 to 40 years were allocated in Group 1 and those aged 60 years or older were allocated in Group 2 (details are presented in Table 1), whose age cut is based on the ages compared in the study by Van den Broeck et al. (2012). Next, before conducting the DIF analysis, we performed a principal component analysis (PCA) of the residues using the Winsteps software (Linacre, 2010) to verify the existence of a pattern in the residues and ensure a unifactorial assumption (necessary for the DIF analysis). In this analysis, we tried to falsify the hypothesis that there are correlations between residues in such a way that they can constitute factors. When this happens, we have evidence that there is a second (or third) construct influencing the responses to items. In the case of eigenvalue, the first contrast is at the noise level if it is greater than 2.0, and the assumption of one-dimensionality is not guaranteed. This analysis aims to confirm the possibility of conducting analyses in the Rasch model (Linacre, 2010).

Subsequently, we conducted the DIF analysis also using the Winsteps software. We considered

the χ^2 Mantel-Haenszel to identify the presence of item-DIFs (Linacre, 2010). We used the criterion proposed by Linacre (2010) to interpret the size of the DIF effect on the items (the contrast) (Penfield, 2007), in which: contrasts ≤ 0.43 are considered negligent, between 0.43 and 0.64 are moderate, and those > 0.64 are considered large. It is worth highlighting two pieces of information: a) a good instrument does not present more than 25% of moderate or large item-DIFs in favor of only a part of the respondents (Penfield, 2007), and b) the DIF is only considered problematic when the source of variance is irrelevant to the construct measured in the test (Karami, 2012).

It should be noted that we analyzed the differential item functionality (DIF) in the items of the *Big Five Inventory-2* (BFI-2) considering only the age of the participants. It is important to clarify that even having other characteristics of the sample for possible verification of differential functioning (for example, by the type of response *online* or on site), we chose to analyze only the age variable due to the insufficient number of participants in certain demographic groups and because they escape the scope of the present investigation.

Results

When considering the 60 items of the BFI-2, 15 were identified with considerable DIF size (whether positive or negative DIF). Of these items, six obtained large DIF size and the others resulted in moderate size. Among the 15 items, nine were more difficult for the group of respondents aged 60 years or older, while six items were more difficult for people aged up to 40 years. Table 2 presents details in this regard.

Table 2 – Summary of the BFI-2 items that presented large and moderate DIFs by the age of the respondents (N=808)

Big Five Inventory-2 items	Personality factor	Personality facet	Contrast DIF	Mantel - Haenszel	
				χ^2	P
3. I tend to be disorganized	C	Organization	-0.51	32.61	0.000
8. I tend to be lazy	C	Productivity	-1.07	99.99	0.000
12. I often find fault with others	A	Trust	-0.51	45.50	0.000
13. I'm responsible, I do my part at work	C	Responsibility	0.58	25.33	0.000
18. I am organized, I like to keep things in order	C	Organization	0.60	47.07	0.000
22. I tend to challenge people	A	Respect	-0.47	35.49	0.000
23. I have trouble beginning tasks	C	Productivity	-0.59	47.29	0.000
24. I feel confident, satisfied with who I am	N	Depression	0.98	99.99	0.000
33. I keep things clean and tidy	C	Organization	0.79	82.88	0.000
38. I do the tasks well and without wasting time	C	Productivity	0.70	76.06	0.000
47. I can be different, cold and distant from others	A	Compassion	-0.87	80.20	0.000
48. I leave things messy, without cleaning	C	Organization	-0.52	36.74	0.000
53. I am persistent, I dedicate myself to the tasks until they are finished	C	Productivity	0.59	40.89	0.000
54. I often feel depressed, sad	N	Depression	-0.48	28.01	0.000
58. Sometimes I behave irresponsibly	A	Respect	-0.75	57.72	0.000

Note. A negative contrast indicates that the item is more difficult for the group ≥ 60 years; C= Conscientiousness, A= Agreeableness, N= Neuroticism. The items shown were firstly translated from English to Portuguese, then from Portuguese to English.

From the perspective of personality traits, the 15 item-DIFs identified belong to factors C, K, and N. No E and O items presented DIF. In terms of quantity, there was a predominance of bias in the C items since none of the 15 item-DIFs of the BFI-2 belong to this personality factor. Of these nine items, four are related to the Organization facet, four of Productivity, and one is related to Responsibility. Agreeableness was the second personality trait with the most item-DIFs ($f=4$), all of which were more difficult for people aged 60 years or older, with two items referring to the Respect facet, one of Trust, and one of Compassion. Two other items, belonging to the Neuroticism trait, resulted in DIF, one of which favored younger people and belongs to the Depression facet, and

the other, also referring to the Depression facet, resulted in favoring older people.

In general, it can be said that the item-DIFs identified in the BFI-2 appear relatively well distributed between the two age groups tested in the present study. Part ($f=9$) of these items favors respondents up to 40 years old and the other part ($f=6$) favors respondents aged 60 and over. A facet-level analysis (Table 2) allows the identification of the cancellation effect of the item-DIFs. For example, items 22 and 58 of the Respect facet (A) resulted in DIF. However, one of them favors older people and the other favors younger people. In turn, items 3, 8, 33, and 48, all referring to the Organization facet (C), also resulted in DIF, two favoring younger people, and two favoring older

people. Additionally, items 23 and 38, referring to the Productivity facet (C), favor elderly people and young adults, respectively. The same can be said about items 24 and 54, of the Depression facet (N), in which one item favors young adults and the other favors older people.

Discussion

The present study aimed to verify the presence of response bias in the BFI-2 items regarding the age of the respondents. As expected, we identified that some items presented bias regarding the age of the respondent, some of them proving more difficult for people over 60 years old. Despite this, it was also observed that other items were more difficult for respondents under 40 years of age. In addition to highlighting the balance in the quantities of item-DIFs in the BFI-2, that is, hindering and facilitating to people in the two age groups tested to a similar extent, this finding denotes justice, as one item-DIF cancels another (Teresi, 2006). This leads us to believe that the response bias identified in these items does not affect the assessment of the FFM performed with the BFI-2, regardless of the age of the respondent. Also, the total number of item-DIFs identified in the BFI-2 were below the cut of 25% of the total number of items in disfavor of a specific group suggested by Penfield (2007). These results suggest that the BFI-2 is a fair instrument (Linacre, 2010) for assessing the FFM of people at different stages of the life cycle.

Even knowing that the studies of construction and validation of personality tests involve, in most cases, samples of subjects who are in the productive stage of life, it was possible to identify that the BFI-2 items somehow overcome this problem described in our hypothesis, having their scores not negatively biased by the wording of the items. However, this result does not mean that some items of the BFI-2 should not be re-elaborated, thinking of elderly people, or even perhaps omitted from the instrument when this is the population that answers it, either in research or in clinical practice in Psychology. Considering this issue is also fundamental since we have evidence

that younger adults and older people process their experiences in the different domains of life distinctly (McAdams & Olson, 2010), and these particularities cannot be ignored in the process of building and validating psychological instruments. Additionally, it should be noted that the wording of the items that presented DIF in the present study does not mostly refer to the educational or work context, suggesting that other differences marked the differential functioning of the items.

From the perspective of large factors, we identified a predominance of bias in the responses of Conscientiousness and Openness items in relation to the age of the respondent. In addition to resulting as expected, this finding is in line with the study by Van den Broeck et al. (2012), who reported the highest number of item-DIFs in C and K. The Conscientiousness items of the BFI-2 identified with bias in disfavor of people aged 60 years and over address attributes of functioning indicative of responsibility, productivity, and organization (behaving irresponsibly, disorganization, difficulty to begin work and study tasks, and leave things messy). In a way, the fact that items with these contents prove more difficult for older people corroborates the literature of personality development in which, over time, people become more responsible, more satisfied with what they have, and less concerned with productivity (Marsh et al., 2013). In this sense, it may even be essential that older people begin to worry more about their own health and safety (and less about productivity), and this adaptation of interest is even a protective attribute of personality in aging (Graham & Lachman, 2012).

At the item level, we hypothesized that some BFI-2 C items would present DIF, namely: 8 "*I tend to be lazy*", 18 "*I am organized, I like to keep things in order*", 23 "*I have trouble beginning tasks*", and 38 "*I do the tasks well and without wasting time*". In this regard, the present research indicated that these items presented differential functioning, corroborating the hypothesis that it is necessary to exercise caution to interpret the responses of elderly people to these Conscientiousness items. In addition, it seems valid to highlight that, when

we compare the wording of the BFI-2 and NEO PI-R items, we identified that the second has a high amount of items containing the terms *work*, *study*, *meeting* whereas the BFI-2 features two or three items only.

Agreeableness was the second personality trait with the most response bias in relation to the age of the respondents, both in the current research and in the study by Van den Broeck et al. (2012). However, the direction of bias was opposite in both studies. In the present study, the item-DIFs of Agreeableness were more difficult for elderly people, while in the research conducted by Van den Broeck et al. (2012), the item-DIFs favored the elderly. Although this seems inconsistent, this result is adequate since the A item-DIFs identified in the BFI-2 are negatively worded and portray the attitudes of finding fault in others (negative trust), contesting (negative respect), being indifferent and cold (negative compassion), and irresponsible (negative respect), pointing to the opposite pole of Agreeableness. It seems sensible to expect that older people tend to be kinder than young adults, as shown by some normative studies of this trait (McAdams & Olson, 2010; Nye & Roberts, 2019; Specht et al., 2011). In all these cases, the source of variance is apparently relevant to the construct, in such a way we should not interpret these item-DIFs as problematic (Karami, 2012).

In turn, the Conscientiousness items of the BFI-2 identified with bias in disfavor of people under 40 years old present attributes related to responsibility with work, organization, performing tasks well, persistence, keeping things clean, and feeling more confident. The fact that as people age, they no longer have an obligation to worry about productivity (Marsh et al., 2013) does not mean that they should not be responsible with their life and health, areas that become the new objects of concern. This result makes sense with normative personality surveys that report that people tend to increase their Conscientiousness levels linearly as they age (Damian et al., 2018), becoming more responsible and less involved in life-threatening situations. Maximizing responsibility with one's own life possibly has an adaptive

effect since it guarantees more life time. Thus, these C item-DIFs should not be seen as problematic. Apparently, these item-DIFs only reflect the impact of the age variable on the manifestation of the latent trait being measured between the compared age groups (Karami, 2012).

Although the response biases identified in these BFI-2 items are not problematic, when it comes to the assessment of elderly people, their presence highlights the relevance in readjusting the wording of these items, seeking to incorporate some attribute of the most typical reality of this population to the theme of the item. Conducting interviews with older people, or even conducting focus groups, to rewrite these items could be a viable form to objectively contribute to the task of accommodating the characteristics of older people in the instruments (APA, 2014). Adjusting these items may be essential for future assessments since we will very possibly have an increase in people aged 60 years or more working in the coming decades especially with the Brazilian pension reform, which will need to be considered in relation to personality assessment in the coming years.

Among the limitations of the study, we indicate some aspects that do not allow the affirmation that the differential functioning of the items presented is exclusively due to the age group since there is an imbalance in the number of elderly people compared to the adult group, in the number of men and women considering the different ages, and the fact that the collection was performed *online* and on site (with and without the help of the researcher). As in the group over 60 years, the amount of men and of *online* collections are very low. Therefore, we chose not to analyze DIF through these variables, requiring future studies to investigate this question.

Another limitation of the present research is the fact that we did not investigate the cognitive functioning of the elderly participants, which may make it difficult to guarantee that these participants were adequately understanding the items and instructions, and that they were giving consistent answers. Despite this possibility, it

is relevant to mention that, at the time of data collection in the institutions for older people, some professionals of the institution informed the researchers who were the individuals whose protocol should be analyzed with caution, whose protocols were not included in the database. Thus, some monitoring was done through the report of professionals who worked directly with the elderly, of those participants who presented some more severe cognitive impairment and who, for this reason, would have more chance of not adequately understanding the items of the instrument. Even so, it is prudent to highlight that the lack of this control in a standardized way can compromise the results described here. Thus, we hope that new studies will include the application of some instrument for the cognitive assessment of participants before using personality instruments.

Final Considerations

Together, the findings indicate that the BFI-2 presents balanced DIF. In other words, the differences in the difficulty levels of the items is balanced in the two age groups tested in the present research, highlighting their age neutrality. The BFI-2 proved to be a fair instrument for assessing the FFM of respondents at different stages of human development because the item-DIFs cancel each other (Teresi, 2006). Despite this, it was possible to corroborate the hypothesis that some items of the BFI-2 demand attention from researchers and psychologists who perform personality assessments with the instrument.

There was a predominance of bias in the Conscientiousness and Agreeableness factors, whose items of the Brazilian version deserve to be revised in the future, namely: item 13 could be rewritten, removing the term *work* since many people over 60 may no longer work. Additionally, item 23 can be rewritten, better informing the content of the difficulty to begin tasks, if by *lack of momentum* (laziness) or by *physical and bodily issues*. Also, item 58 *Sometimes I behave irresponsibly* could be rewritten, seeking to address attributes of responsibility related to the age of

70-80 (e.g., failing to take medication, undergoing examination/treatment, or caring for the body). We also understand that these reviews could be made with focus groups composed of elderly people invited to semantically analyze the items, eventually proposing new writings.

Adapting these items for older people maximizes the potential of BFI-2 in studying the relationship between the profile of older people, who will have ever more life time in Brazil (IBGE, 2018), and individual patterns of functioning in adult life. This knowledge allows the elaboration of psychological interventions that favor the healthy development of people throughout the life cycle, a topic of very high social relevance, especially for the coming decades. Furthermore, the concern with the quality of the BFI-2 items is fundamental for the work of psychologists in the different contexts of practice to accommodate the specificities in the functioning of people belonging to the different stages of human development, as advocated by the APA (APA, 2014). Finally, we highlight the importance of the use of analytical strategies, such as the DIF in psychological instruments in use in Brazil to evaluate the possible effect of different demographic variables on the participants' response pattern since there are few similar studies available and this step can add quality to the set of psychometric evidence of the instruments, providing useful information for the practice of psychologists who performs personality assessments.

References

- American Psychological Association. (2014). Guidelines for Psychological Practice with Older Adults. *American Psychologist*, 69(1), 34-65. <https://doi.org/10.1037/a0035063>
- Chapman, B. P., Benedict, R. H., Lin, F., Roy, S., R., Federoff, H. J., & Mapstone, M. (2017). Personality and Performance in Specific Neurocognitive Domains Among Older Persons. *The American Journal of Geriatric Psychiatry*, 25(8), 900-908. <https://doi.org/10.1016/j.jagp.2017.03.006>
- Graham, E. K., & Lachman, M. E. (2012). Personality stability is associated with better cognitive performance in adulthood: Are the stable more able? *The Journals of Gerontology: Series B*, 67, 545-554. <https://doi.org/10.1093/geronb/gbr149>

Instituto Brasileiro de Geografia e Estatística. (2018). Projeções e estimativas da população do Brasil e das Unidades da Federação. Estudos e pesquisas. <https://www.ibge.gov.br/apps/populacao/projecao>

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy. In *Handbook of personality: Theory and research* (pp. 114-158). Guilford Press. <http://www.ocf.berkeley.edu/~johnlab/2008chapter.pdf>

Jokela, M., Pulkki-Råback, L., Elovainio, M., & Kivimäki, M. (2014). Personality traits as risk factors for stroke and coronary heart disease mortality: Pooled analysis of three cohort studies. *Journal of Behavioral Medicine*, 37(5), 881-889. <https://doi.org/10.1007/s10865-013-9548-z>

Karami, H. (2012). An Introduction to Differential Item Functioning. *The International Journal of Educational and Psychological Assessment*, 11(2), 59-75. http://www.academia.edu/1030760/An_introduction_to_Differential_Item_Functioning

Kaup, A. R., Harmell, A. L., Yaffe, K. (2019). Conscientiousness is associated with lower risk of dementia among Black and White older adults. *Neuroepidemiology*, 52, 86-92. <https://doi.org/10.1159/000492821>

Linacre, J. M. (2010). Winsteps (Version 3.70.02) [Computer Software]. Winsteps.com

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>

McAdams, D. P., & Olson, B. D. (2010). Personality development: Continuity and change over the life course. *Annual Review of Psychology*, 61, 517-542. <https://doi.org/10.1146/annurev.psych.093008.100507>

Nye, C. D., & Roberts, B. W. (2019). A Neo-Socioanalytic Model of Personality Development. *Work Across the Lifespan*, 47-49. <https://doi.org/10.1016/B978-0-12-812756-8.00003-7>

Pires, J. G., Nunes, C. H. S. S., & Nunes, M. F. O. (2019). Avaliação da personalidade e o Modelo dos cinco Grandes Fatores. In Baptista, M. N., Muniz, M., Reppold, C. T., Nunes, C. H. S. S., Carvalho, L. F., Primi, R., Noronha, A. P. P., Seabra, A. G., Wechsler, S. M., Hutz, C. S., & Pasquali, L., (Org). *Compêndio de Avaliação Psicológica* (pp. 494-506). Editora Vozes.

Pires, J. G., Nunes, C. H. S. S., Nunes, M. F. O., Primi, R. (in press). Preliminary validity for the Big Five Inventory-2 in Brazilian adults. *PsicoUSF*.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement*, 43(4), 295-312. <https://doi.org/10.1111/j.1745-3984.2006.00018.x>

Roberts, B. W., & Mroczek, D. (2008). Personality Trait Change in Adulthood. *Current Directions in Psychological Science*, 17(1): 31-35. <https://doi.org/10.1111/j.1467-8721.2008.00543.x>

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>

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). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-X. *Journal of Research in Personality*, 68, 69-81. <https://doi.org/10.1016/j.jrpp.2017.02.004>

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, 101(4), 862-882. <http://dx.doi.org/10.1037/a0024950>

Sutin, A. R., Stephan, Y., Damian, R. I., Luchetti, M., Strickhouser, J. E., & Terracciano, A. (2019). Five-factor model personality traits and verbal fluency in 10 cohorts. *Psychology and Aging*, 34(3), 362-373. <https://doi.org/10.1037/pag0000351>

Tautvydaitė, D., Antonietti, J. P., Henry, H., von Gunten, A., & Popp, J. (2017). Relations between personality changes and cerebrospinal fluid biomarkers of Alzheimer's disease pathology. *Journal of Psychiatric Research*, 90, 12-20. <https://doi.org/10.1016/j.jpsychires.2016.12.024>

Teresi, J. A. (2006). Different Approaches to Differential Item Functioning in Health Applications: Advantages, Disadvantages and Some Neglected Topics. *Medical Care*, 44(11), 152-170. <https://doi.org/10.1097/01.mlr.0000245142.74628.ab>

Van den Broeck, J., Rossi, G., Dierckx, E., & De Clercq, B. (2012). Age-neutrality of the NEO-PI-R: Potential Differential Item Functioning in Older Versus Younger Adults. *Journal of Psychopathology and Behavioral Assessment*, 34(3), 361-369. <https://doi.org/10.1007/s10862-012-9287-4>

Yoneda, T., Graham, E., Lozinski, T., Bennett, D. A., Mroczek, D., Piccinin, A. M., Hofer, S. M., & Muniz-Terrera, G. (2022). Personality Traits, Cognitive Status, and Mortality in Older Adults. *Journal of Personality and Social Psychology*, 124(2), 381-395. <https://doi.org/10.1037/pspp0000418>

Jeferson Gervasio Pires

Doutor em Psicologia pela Universidade Federal de Santa Catarina (UFSC), em Florianópolis, SC, Brasil; com estágio pós-doutoral em Psicologia pela UFSC, em Florianópolis, SC, Brasil. Professor no Curso de Psicologia da Faculdade Anhanguera de São José, SC, Brasil.

Carlos Henrique Sancineto da Silva Nunes

Doutor em Psicologia pela Universidade Federal do Rio Grande do Sul (UFRGS), em Porto Alegre, RS, Brasil; com pós-doutorado em Psicologia pela Universidade São Francisco, em Itatiba, SP, Brasil. Professor associado da Universidade Federal de Santa Catarina (UFSC), em Florianópolis, SC, Brasil.

Maiana Farias Oliveira Nunes

Doutora em Psicologia pela Universidade São Francisco (USF), em Itatiba, SP, Brasil; com pós-doutorado em Psicologia pela Universidade Federal do Rio Grande do Sul (UFRGS), em Porto Alegre, RS, Brasil. Professora adjunta da Universidade Federal de Santa Catarina (UFSC), em Florianópolis, SC, Brasil.

Endereço para correspondência

Carlos Henrique Sancineto da Silva Nunes
Universidade Federal de Santa Catarina (UFSC)
Departamento de Psicologia
Campus Universitário Reitor João David Ferreira Lima
Trindade, 88040-900
Florianópolis, SC, Brasil

Os textos deste artigo foram revisados pela SK Revisões Acadêmicas e submetidos para validação do(s) autor(es) antes da publicação.