

Comparison of methods for controlling acquiescence bias in balanced and unbalanced scales

Gustavo Henrique Martins¹

Felipe Valentini¹

Bruno Bonfá-Aranjo²

Adriana Satiko Ferraz³

Fernanda Otoni⁴

André Pereira Gonçalves⁵

¹University of São Francisco, Campinas, São Paulo, Brazil

²The University of Western Ontario, London, Ontario, Canada

³University Center of Brasília, Brasília – Distrito Federal, Brazil

⁴University Tuiuti do Paraná, Curitiba, Paraná, Brazil

⁵Federal University of Bahia, Vitória da Conquista, Bahia, Brazil

Abstract

Controlling acquiescence bias typically involves the application of positive and negative keyed items. However, little is known about the effect of balancing positive and negative items on bias control. The aim of this study was to compare three Confirmatory Factor Analysis models (without control, MIMIC, and Random Intercept) to recover the factor structure of unbalanced and balanced instruments, using simulated and real data (from an instrument that assesses Personality). By controlling for acquiescence, the results indicated that the performance of balanced scales was better than that of unbalanced scales, as well as in the absence of control for response bias, when considering balanced and unbalanced scales. Thus, this research suggests the possibility of controlling acquiescence through balanced instruments associated with the use of statistical methods in modeling. *Keywords:* Factor Analysis; Psychometrics; Psychological Tests; Response Bias; Response Style.

Comparação de métodos para o controle do viés da aquiescência em escalas balanceadas e desbalanceadas¹

Resumo

O controle do viés de aquiescência normalmente envolve a aplicação de itens positivos e negativos. Contudo, pouco se sabe sobre o efeito do balanceamento entre itens positivos e negativos sobre o controle do viés. O objetivo deste estudo foi comparar três modelos de Análise Fatorial Confirmatória (sem controle, MIMIC e Intercepto Randômico) para recuperar a estrutura fatorial de instrumentos desbalanceados e balanceados, a partir de dados simulados e reais (procedentes de um instrumento que avalia Personalidade). Mediante o controle da aquiescência, os resultados indicaram que a performance de escalas balanceadas foi melhor do que de escalas desbalanceadas, bem como na ausência de controle desse viés de resposta, ao considerar as escalas balanceadas e desbalanceadas. Dessa maneira, esta pesquisa aponta para a possibilidade de controle de aquiescência por meio de instrumentos balanceados associada ao uso dos métodos estatísticos na modelagem.

Palavras-chave: Análise Fatorial; Psicometria; Testes Psicológicos; Viés de resposta; Estilo de Resposta.

Comparación de métodos para controlar el sesgo de aquiescencia en escalas balanceadas y no balanceadas

Resumen

El control del sesgo de aquiescencia involucra la aplicación de ítems positivos y negativos. Sin embargo, el efecto del equilibrio entre ítems positivos y negativos en el control del sesgo sigue siendo una pregunta abierta. En este sentido, el objetivo de este estudio fue comparar tres modelos de Análisis Fatorial Confirmatorio (sin control, MIMIC e Intercepto Aleatorio) para recuperar la estructura factorial de instrumentos balanceados y desbalanceados, a partir de datos simulados y reales (a partir de un instrumento que evalúa personalidad). El control de este sesgo de respuesta indicó que el desempeño de escalas balanceadas fue mejor que el de escalas desbalanceadas, así como en la ausencia del control de la aquiescencia, al considerar escalas balanceadas y desbalanceadas. Por lo tanto, esta investigación sugiere la posibilidad de controlar este sesgo de respuesta por medio de instrumentos balanceados asociados con el uso de métodos estadísticos modelado.

Palabras clave: Análisis Fatorial; Psicometria; Testes Psicológicos; Sesgo de Respuesta; Estilo de Respuesta.

¹ Preprint of this manuscript in Portuguese: https://www.researchgate.net/publication/371898024_Comparacao_de_metodos_para_o_controle_do_vies_da_aquiescencia_em_escalas_balanceadas_e_desbalanceadas

Measurement instruments that use Likert-type scales are susceptible to agreement biases (or disagreement biases). For example, a person may strongly agree with the item “I am talkative” and, at the same time, strongly agree with the item “I am quiet”. This response bias (also known as response style) is called acquiescence, characterized by the respondent’s tendency to endorse items systematically, regardless of their content (Kam & Meyer, 2015; Paulhus, 1991; Wetzal et al., 2016).

The acquiescent response style can stem from the individual’s difficulty in understanding the content of the items and reflecting on how well the described characteristics align with their behaviors, thoughts, and feelings (He et al., 2014). Evidence suggests that the pattern of acquiescent response tends to be higher in childhood (i.e., 10 years) and decreases throughout adolescence until stabilizing in early adulthood (i.e., 18 and 19 years), suggesting influence by cognitive elements (Soto et al., 2008). In addition to these aspects, the display of acquiescence can also result from the construction of instruments with items that only present positive or negative meaning, which contributes to the acquiescent response pattern (Henninger, 2019; Plieninger, 2018; Podsakoff et al., 2003).

One of the problems associated with acquiescence is the presence of systematic error variance in responses and, consequently, in instrument scores (Danner et al., 2015; Kam & Meyer, 2015; Kuru & Pasek, 2016; Lechner & Rammstedt, 2015), which can compromise its psychometric characteristics (Valentini, 2017; Zanon et al., 2018). In the case of precision estimates, statistical inferences about the estimated parameters can reproduce sample means that do not correspond to the respondent’s latent trait. Furthermore, in investigations of validity evidence, acquiescence tends to inflate correlations among items worded in the same direction and suppress correlations among items worded in opposite directions (Danner et al., 2015; Kam & Meyer, 2015; Kuru & Pasek, 2016; Lechner & Rammstedt, 2015). This can have detrimental effects on psychological testing in various contexts of psychology because acquiescence tends to bias individuals’ scores, distancing them from their true score.

An attempt to avoid issues with acquiescence is to construct measurement instruments that present balanced scales, composed of items with both positive and negative directions; for example: “I am talkative” and “I am quiet” (Kam & Zhou, 2015; Lorenzo-Seva & Ferrando, 2009; Soto & John, 2017). In addition,

it is recommended that: a) one should avoid using the term “not” for item reversals as it tends to make reading and comprehension difficult (Barnette, 2000; Gehlbach & Artino Junior, 2018); b) the scoring key should be compatible with the construct/skill assessed by the instrument, and both positive and negative items should be answerable through them. The anchor terms of the Likert scale should be symmetrical (e.g., “1-Not at all”, “2-Slightly”, “3-Moderately”, “4-Very much”, and “5-Completely”; Sliter & Zickar, 2014; Wetzal et al., 2016).

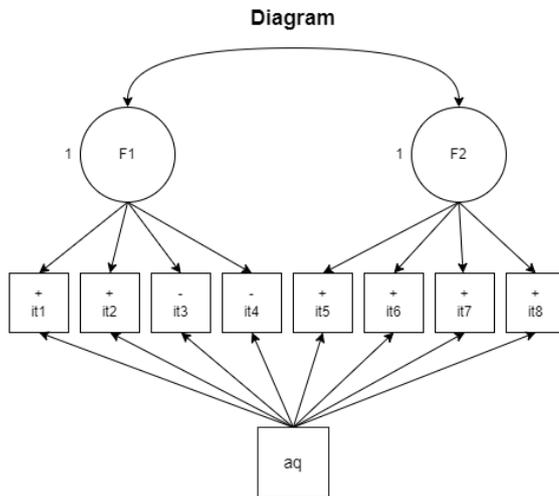
Another way to control acquiescence involves applying modeling to the data obtained from the instrument (Henninger, 2019; Plieninger, 2018; Podsakoff et al., 2003). In this study, we sought to demonstrate two strategies for modeling acquiescence - Multiple Indicators, Multiple Causes (MIMIC) and Random Intercept; as well as to evaluate the importance of balancing positive and negative items through simulated and real data.

Main models for controlling acquiescence: MIMIC and Random Intercept

Statistical methods such as MIMIC and Random Intercept can be adopted to control for acquiescence bias (Henninger, 2019; Plieninger, 2018; Podsakoff et al., 2003), through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). To use these control methods (MIMIC and Random Intercept), it is necessary to have items with opposite directions (positive and negative; Geiser et al., 2008), as there is a risk of overcontrol when applied to unbalanced scales, that is, scales consisting only of positively or negatively worded items (Valentini & Hauck Filho, 2020).

In MIMIC, in addition to the latent content factor (i.e., the construct being evaluated by the instrument), the items are also explained by (or regressed on) an observed external variable whose scores represent the acquiescence index (typically the mean of positive and negative items without reversal). In a balanced scale with a Likert-type response scoring key, for example, acquiescence is identified when an individual agrees with two opposite situations. Thus, in MIMIC, by assigning the acquiescence index as the cause of item responses, the factor loadings obtained on the content factor are controlled and, therefore, free from acquiescence bias (Wetzal & Carstensen, 2015). Figure 1 illustrates a CFA model with MIMIC for controlling acquiescence.

The diagram graphically presents the MIMIC model. In this case, two factors are displayed (represented by circles F1 and F2), both with two positive and



Note. The + and - signs refer to the direction of the items in the factors (F1 and F2).

Mplus

```
DEFINE:
Aq=MEAN(it1-it8);

MODEL:
F1 by it1-it4;
F2 by it5-it8;
it1-it8 ON Aq;
Aq with F1-F2@0;
```

R package "lavaan"

```
1 library(lavaan)
2 db$Aq <- apply(db[,c(1:8)], 1, mean)
3
4 model_MIMIC <- "
5 f1 =~ NA*it1 + it2 + it3 + it4
6 f1 =~ 1*f1
7 f2 =~ NA*it5 + it6 + it7 + it8
8 f2 =~ 1*f2
9
10 it1 + it2 + it3 + it4 ~ Aq
11 it5 + it6 + it7 + it8 ~ Aq
12 Aq =~ 0*f1 + 0*f2"
13
14 CFA_MIMIC <- sem(
15   model_MIMIC,
16   data = db,
17   ordered = TRUE
18 )
```

Figure 1. Diagram of a Confirmatory Factor Analysis with acquiescence control using MIMIC and MPlus and lavaan syntax

two negative items, along with an observed acquiescence score. In the MPlus software, using the DEFINE command, the mean of all items is calculated, which serves as the acquiescence score. In a balanced scale, the mean of the items, without reversing them, is the central value of the Likert scale. For example, for the item 'talkative', a participant marked response 5, and for the item 'quiet', the response was 1. Thus, the mean is 3 ($5 + 1 / 2$), the midpoint of the scale. However, any deviation from this mean indicates acquiescent or non-acquiescent responses. Next, using the MODEL command, in addition to the factor analysis, the items are regressed on the newly created acquiescence score, and the correlation between the acquiescence score and the factors is fixed at zero (Aq with F1-F2@0). In the R software, using the lavaan package, the first line of the syntax indicates how to estimate the mean of all items, followed by the estimation of the factor model.

In turn, Random Intercept models an uncorrelated latent acquiescence factor with the content factors (Maydeu-Olivares & Coffman, 2006). The item loadings on the acquiescence factor (Random Intercept) are fixed at +1 (both for positive and negative items). Thus, the model will estimate the acquiescence factor capturing only the strictly common and unidirectional variance

among positive and negative items (e.g., the tendency to agree with both positive and negative items). Figure 2 illustrates a CFA model with acquiescence control by Random Intercept. We emphasize that the loadings restriction is not a condition for the identification of the structural equation model, but it is used to capture the acquiescence variance. Considering that acquiescence reflects a tendency to positively endorse all items, positive or negative, the restriction imposed on the loadings at +1 aims to capture the common variance that is independent of item direction (i.e., acquiescence). Thus, releasing the loadings to be freely estimated may result in a model that captures other types of biases and even legitimate psychological content of the item, leading to overcontrol (i.e., overestimating the bias and removing variance from the psychological construct).

In both MPlus and R (lavaan) software, a common CFA is estimated, adding a line to model the latent variable of acquiescence. The correlation between acquiescence and factors is fixed at zero for model identification. Note that the variance of acquiescence is estimated (not fixed), allowing for comparison with the genuine variances of the factors.

In the present research, we used examples of balanced scales, that is, with an equal number of pairs of

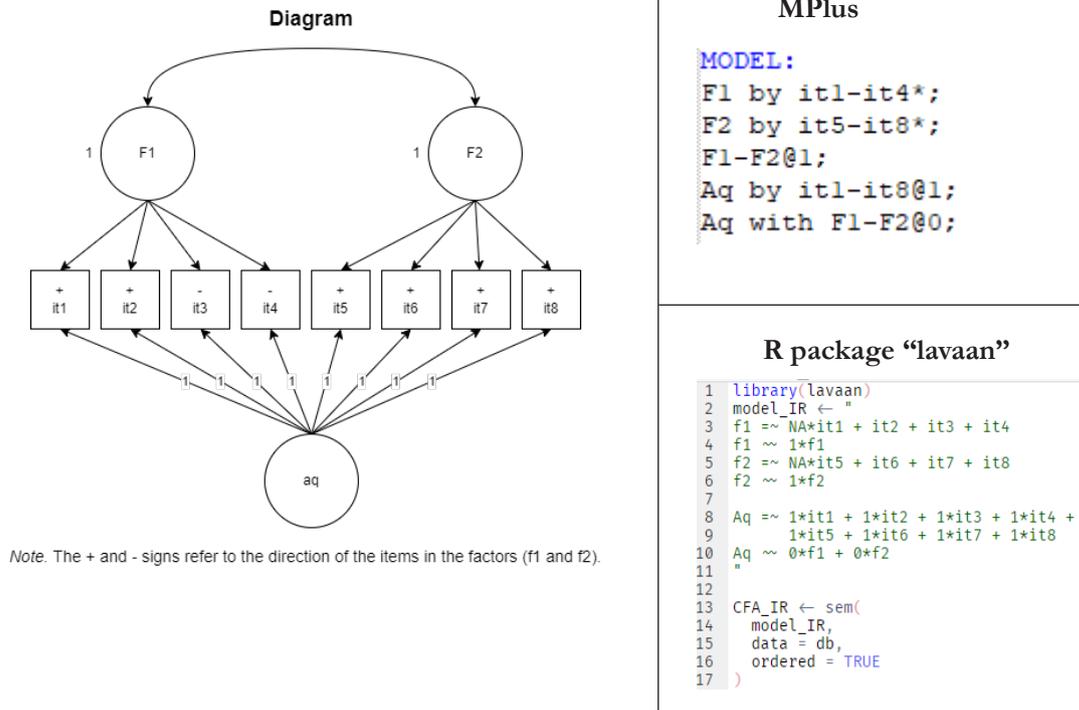


Figure 2. Diagram of a Confirmatory Factor Analysis with acquiescence control using Random Intercept and MPlus and lavaan syntax

opposing items (positive and negative) in each factor. However, MIMIC and Random Intercept methods hypothetically allow for imbalance between positive and negative items. In cases of imbalance, does the modeling performance remain adequate? With this in mind, our goal is to compare the performance of three CFA models (without control, MIMIC control method, and Random Intercept control method) for recovering the factorial structure of both balanced and imbalanced instruments. Our research involves two studies: the first with simulated data and the second with real data collected using items from the Big Five Inventory (BFI-2; Soto & John, 2017), based on both simulated and real data.

The BFI-2 (Soto & John, 2017) assesses personality from the perspective of the Big Five model, which, in broad terms, comprises this construct in five broad domains, namely: extraversion, agreeableness, conscientiousness, negative emotionality, and open-mindedness. This self-report instrument consists of items that describe the adjectives of each personality trait, structured in short and simple sentences, and is responded to using a Likert scale. The comprehensibility of the BFI-2 and its psychometric quality (i.e., precision estimates and evidence of validity) (John,

2021; Soto & John, 2017) determined its selection for the current study.

Simulation Study

Method

Simulation Conditions

We tested 18 conditions, varying the following criteria:

- a) Low/moderate factor loadings (0.30 to 0.50) and high factor loadings (0.60 to 0.80). We justified the criteria based on the commonly observed loadings in psychological instruments, as it is common practice to discard items with loadings below 0.30, and items with loadings above 0.80 are uncommon.
- b) Acquiescence variances (0.05, 0.10, and 0.20). Acquiescence variances between 0.05 and 0.10 are common in the literature (Soto et al., 2008), but we also included the condition with a variance of 0.20 to test the performance of acquiescence control methods in this less common but possible scenario.

- c) Sample sizes (200, 500, and 1000). We included these conditions to evaluate whether the acquiescence control methods are effective for different sample sizes.

Each condition represents an interaction of the three criteria. Regarding the thresholds, they were randomly drawn from a uniform distribution, with the same value for opposing pairs of items. The simulated correlations between factors ranged from 0.26 to 0.59, representing situations that are very common in the context of Psychology.

Simulation Procedures and Analysis

The data were simulated to reproduce an instrument with 30 items and five factors. This number of items and factors is common in other psychological instruments in the Brazilian context (Ambiel et al., 2022; Zanon et al., 2014; Siqueira, 2008). As such, four steps were followed: database simulation, item selection, modeling, and recovery and summary of modeled parameters.

In the first step, the databases were simulated. Although we aimed to reproduce an instrument with 30 items, in order to test different balancing conditions, it was necessary at this stage to simulate databases with 60 balanced items. These items represent psychometrically opposing pairs, that is, with the same magnitude of factor loading, threshold, and acquiescence variance between positive and negative items. We considered that items from unbalanced instruments are a sample from balanced items. Therefore, the simulation of extra items allowed the selection of samples of balanced and unbalanced items in the second step. If we had simulated separate databases for balanced and unbalanced conditions, we could have had an experimental design effect that would mainly affect the unbalanced items. We simulated databases to represent responses on a five-point Likert scale. The simulated factor structure consisted of five factors, with 12 items each.

In the second step, we selected 30 items to represent a balanced instrument (three positive and three negative items per factor) and 30 items to represent an unbalanced instrument (five positive and one negative item per factor). Thus, we had four common items per factor in both versions, with three positive and one negative item. Furthermore, we chose not to select items that were perfect opposing pairs, as psychometrically opposing pairs are rarer in real instruments.

In the third step, the modeling phase, we conducted three CFAs for each version (i.e., balanced and unbalanced). The difference between the three CFA models occurred in the control of acquiescence. In the first model, acquiescence was not controlled; in the second model, we controlled it using the MIMIC method; and in the third model, we controlled it using the Random Intercept method. All CFA models replicated the simulated structure consisting of five factors. Moreover, the Diagonally Weighted Least Squares (DWLS) estimator was used, considering the total weight matrix to correct the calculation of standard error, mean, and variance-adjusted test statistic, which is suitable for ordinal data (Li, 2016). For each analysis of each tested condition, we conducted 500 replications.

In the fourth step, to compare the performance of the tested CFAs, biases were calculated, which refer to the distance between the obtained factor loading in the analysis compared to the simulated factor loading (i.e., obtained factor loading - simulated factor loading / simulated factor loading). In this study, biases of up to 10% (i.e., 0.10) were considered acceptable. A positive bias indicates that the obtained factor loading had a higher magnitude than the simulated loading, while a negative bias means that the obtained loading was of a lower magnitude than the simulated loading. To calculate the bias of a set of positive and negative items, the average of the biases obtained for the positive and negative items, respectively, was calculated for a specific condition. This way, an overall measure of bias present in a set of items, considering both positive and negative items, can be obtained. The data were analyzed using the R software with the “lavaan” package (Rosseel, 2012). The script for the simulation conducted in this study is available at the following link: https://github.com/GustavHM/simulacao_acquiescencia

Results

Table 1 presents a summary of the biases of the factor loadings obtained in the tested conditions.

In general, in Table 1, it is possible to observe that the highest frequency of biases greater than 0.10 was observed in the unbalanced data, without control of acquiescence, and with low (i.e., 0.30 to 0.50) content factor loadings (i.e., loadings on the factor representing the construct intended to be measured by the instrument). It is also noted that, in most conditions, the biases indicated an overestimation of the loadings of the positive items (i.e., positive biases)

Table 1.

Summary of the mean of biases of the loadings obtained in the tested conditions in the simulation

Conditions			Unbalanced items						Balanced items					
			Without control		MIMIC		RI		Without control		MIMIC		RI	
Loadings	Acq Var	Sample	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Low	0.20	200	0.50	-1.60	0.12	-0.20		-0.15	0.11	-2.11				
Low	0.20	500	0.50	-1.61	0.13	-0.22			0.12	-2.10				
Low	0.20	1000	0.50	-1.61	0.13	-0.22			0.12	-2.10				
Low	0.10	200	0.29	-1.00	0.12	-0.20		-0.10	-0.48	-0.56				
Low	0.10	500	0.30	-1.01	0.12	-0.22			-0.59	-0.52				
Low	0.10	1000	0.29	-1.01	0.12	-0.22			-0.74	-0.44				
Low	0.05	200	0.17	-0.55		-0.14								
Low	0.05	500	0.17	-0.55		-0.16								
Low	0.05	1000	0.17	-0.56		-0.16								
High	0.05	200		-0.18										
High	0.05	500		-0.18										
High	0.05	1000		-0.18										
High	0.10	200	0.11	-0.35										
High	0.10	500	0.11	-0.35										
High	0.10	1000	0.11	-0.35										
High	0.20	200	0.20	-0.68										
High	0.20	500	0.20	-0.68										
High	0.20	1000	0.20	-0.68										

Note. Acq Var = Acquiescence Variance ; Low loadings = 0.30 a 0.50; High loadings = 0.60 a 0.80. Biases equal to or greater than 0.50 are in bold, and biases lower than 0.10 have been omitted.

and an underestimation of the loadings of the negative items (i.e., negative biases). To better visualize the results, we selected some conditions that showed moderate biases (≥ 0.50) and severe biases (≥ 1). In Table 2, we present the unbalanced data; and in Table 3, the balanced data. Considering that the biases remained practically the same regardless of the sample size, a sample of 1000 cases was selected to illustrate the biases obtained for different acquiescence variances and simulated loadings.

The results of the conditions presented in Table 2 show that acquiescence overestimated the loadings of the positive items. However, the main estimation errors were related to the loadings of the negative items, which in the three conditions presented had their factor

loadings strongly underestimated. In the first and fourth conditions, the negative items had negative loadings, but below 0.30. In the second condition, the negative items had loadings close to zero, and in the third condition, the negative items had their loadings reversed, presenting positive loadings on factors that were supposed to be negative. In summary, in these four cases, all five negative items had low factor loadings and were incongruent with the theoretically expected values. Table 3 presents the simulated and obtained factor loadings in two conditions for the balanced data.

It can be observed in Table 3 that, in the two presented conditions, the issues involved all items, both positive and negative. In the first condition, all loadings were underestimated to the point that all items had

Table 2.
Factor loadings obtained and simulated in the data with unbalanced items

Simulations with low loadings						Simulation with high loadings	
Acq Var 0.05		Acq Var 0.10		Acq Var 0.20		Acq Var 0.20	
Simulated loadings	Loadings without control	Loadings without control	Loadings without control	Loadings without control	Simulated loadings	Loadings without control	
0.43	0.48	0.52	0.59	0.59	0.73	0.85	
0.38	0.44	0.49	0.58	0.58	0.68	0.82	
0.42	0.48	0.52	0.59	0.59	0.72	0.85	
0.40	0.45	0.50	0.58	0.58	0.70	0.83	
0.33	0.41	0.46	0.56	0.56	0.63	0.78	
-0.31	-0.09	0.09	0.31	0.31	-0.61	-0.12	
0.33	0.41	0.46	0.55	0.55	0.63	0.78	
0.38	0.45	0.49	0.57	0.57	0.68	0.82	
0.30	0.38	0.45	0.54	0.54	0.60	0.76	
0.42	0.48	0.53	0.59	0.59	0.72	0.85	
0.45	0.51	0.54	0.60	0.60	0.75	0.87	
-0.45	-0.22	-0.03	0.22	0.22	-0.75	-0.27	
0.47	0.52	0.56	0.62	0.62	0.77	0.89	
0.39	0.45	0.51	0.59	0.59	0.69	0.83	
0.38	0.45	0.50	0.58	0.58	0.68	0.82	
0.41	0.47	0.52	0.59	0.59	0.71	0.84	
0.49	0.54	0.57	0.63	0.63	0.79	0.91	
-0.43	-0.21	-0.02	0.23	0.23	-0.73	-0.26	
0.41	0.47	0.51	0.58	0.58	0.71	0.84	
0.35	0.42	0.48	0.57	0.57	0.65	0.80	
0.34	0.42	0.47	0.56	0.56	0.64	0.79	
0.32	0.39	0.45	0.55	0.55	0.62	0.77	
0.46	0.52	0.55	0.61	0.61	0.76	0.88	
-0.49	-0.25	-0.06	0.20	0.20	-0.79	-0.31	
0.48	0.53	0.56	0.61	0.61	0.78	0.89	
0.33	0.41	0.47	0.56	0.56	0.63	0.78	
0.37	0.44	0.49	0.57	0.57	0.67	0.81	
0.45	0.51	0.54	0.60	0.60	0.75	0.87	
0.39	0.46	0.50	0.58	0.58	0.69	0.83	
-0.41	-0.18	0.01	0.26	0.26	-0.71	-0.21	
Bias		Bias		Bias		Bias	
Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
0.17	-0.56	0.29	-1.01	0.50	-1.61	0.20	-0.68

Note. Acq Var = Acquiescence Variance In all four conditions, a sample of 1000 cases was selected. The loadings of the negative items are in bold.

Table 3.
Factor loadings obtained and simulated in the data with balanced items

	Acq Var 0.10	Acq Var 0.20
Simulated Low loadings	Loadings without control	Loadings without control
0.43	0.10	0.43
0.38	0.08	0.43
0.42	0.10	0.43
-0.40	-0.20	0.44
-0.33	-0.18	0.44
-0.31	-0.18	0.44
0.33	0.09	0.43
0.38	0.10	0.42
0.30	0.07	0.43
-0.42	-0.24	0.46
-0.45	-0.25	0.46
-0.45	-0.25	0.46
0.47	0.14	0.42
0.39	0.11	0.42
0.38	0.11	0.42
-0.41	-0.24	0.45
-0.49	-0.27	0.46
-0.43	-0.24	0.45
0.41	0.12	0.41
0.35	0.09	0.42
0.34	0.09	0.42
-0.32	-0.19	0.44
-0.46	-0.26	0.45
-0.49	-0.26	0.45
0.48	0.13	0.42
0.33	0.08	0.42
0.37	0.09	0.42
-0.45	-0.24	0.45
-0.39	-0.22	0.45
-0.41	-0.23	0.45
	Bias	Bias
	Pos Neg	Pos Neg
	-0.74 -0.44	0.12 -2.10

Note. Acq Var = Acquiescence Variance in both conditions, a sample of 1000 cases was selected. The loadings of the negative items are in bold.

loadings below 0.30, especially the positive items. In the third condition, all items had loadings in the same direction and with the same magnitude, ranging from 0.41 to 0.46, suggesting that they were only representing the variance of acquiescence bias (i.e., 0.20) and not the content variance. Therefore, in both conditions, all items had issues due to very low and/or theoretically inconsistent factor loadings.

In addition to biases in the loadings, we observed some identification errors in the tested conditions for the 200 and 500-case samples. The cases with most identification problems, ranging from 14.40% to 19.60% (out of 500 replications), occurred in all three analyses (without control, MIMIC, and Random Intercept) with balanced items, low factor loadings, 200-case sample size, and regardless of the acquiescence variance. In addition to these, with high loadings, acquiescence variance of 0.20, and 200 cases, the analysis with Random Intercept showed identification errors in 12.40% of the replications conducted. In the remaining conditions, the identification problems were below 4% of the replications. We emphasize that the identification problems occurred due to implausible parameters, such as negative variance.

Illustration with real data

We sought to illustrate the use of Random Intercept and MIMIC for acquiescence control in a real database, in which a balanced personality instrument was used. Additionally, we tested an unbalanced version of item selection to evaluate the impact of this condition on acquiescence.

Method

Participants

The sample consisted of 888 adults from various regions of Brazil. Most participants were female (70.9%), and ages ranged from 18 to 73 years ($M = 23.34$; $SD = 7.65$).

Instruments

Big Five Inventory (BFI-2; Soto & John, 2017).

The instrument is composed of 60 balanced items that are equally divided into the five dimensions of the Big Five model: extraversion (e.g., Is outgoing, sociable - $\alpha = 0.87$), agreeableness (e.g., Is helpful and unselfish with others - $\alpha = 0.83$), conscientiousness (e.g., Is efficient, gets things done - $\alpha = 0.88$), negative emotionality (e.g., Is moody, has up and down mood swings

- $\alpha = 0.91$), and open-mindedness (e.g., Is curious about many different things - $\alpha = 0.84$). The scoring key is provided on a five-point Likert scale ranging from “1 - Disagree strongly. Does not describe me at all” to “5 - Agree strongly. Describes me perfectly”.

Procedures

The sample of this study is the by-product of two Master’s dissertations, whose projects were sub-mitted and approved by the Research Ethics Committee of University of São Francisco (CAAE: 01465718.3.0000.5514; CAAE: 08033419.9.0000.5514). Data was collected in-person ($n = 285$) and online ($n = 603$). The in-person data collection was conducted collectively with university students using a pencil-and-paper format. Conversely, the link for online data collection was made available through social media platforms.

Data Analysis

The CFAs were replicated without control, with MIMIC, and with Random Intercept using real data from the BFI-2. In all analyses, the DWLS estimator was used, considering the total weight matrix to correct the calculation of standard error, mean, and test statistic adjusted for variance. To replicate the analyses, 30 balanced and unbalanced items were selected from the 60 items of the BFI-2. It is worth noting that each factor of the BFI-2 has six positive items and six negative items. The item selection followed a similar systematic approach to that used in the selection of simulated items. Specifically: a) in the balanced version, the first three positive items and the last three negative items were selected, according to the order in which the items are presented in the complete instrument, and b) in the unbalanced version, the first five positive items and the last negative item of each factor were selected. With this systematic item selection approach, four identical

items were always selected per factor in both versions (i.e., balanced and unbalanced), with three positive and one negative item. The fit indices were interpreted as follows: Confirmatory Fit Index (CFI > 0.95), Tucker-Lewis Index (TLI > 0.95), Root Mean Square Error of Approximation (RMSEA < 0.06; Hu & Bentler, 1999). The data were analyzed using the R software with the “lavaan” package (Rosseel, 2012). The script for the simulation conducted in this study is available at the following link: https://github.com/GustavHM/simulacao_aquiescencia.

Results

The same analyses as the simulation (without control, MIMIC, and Random Intercept) were then tested with unbalanced and balanced items using real data from the BFI-2. Table 4 presents the fit indices obtained in the analyses.

Overall, the fit indices were all acceptable and similar in the three analyses (without control, MIMIC, and Random Intercept). There was an improvement in the fit indices for the balanced items in the analyses with MIMIC and Random Intercept compared to the analysis without control. Table 5 describes the results of the factor loadings obtained in the three analyses conducted with the unbalanced and balanced items of the BFI-2.

It can be observed that with the unbalanced items, the factor loadings of the negative items were lower in the analyses without control and with MIMIC compared to the analysis with Random Intercept. On the other hand, no substantial differences were found in the factor loadings of the balanced items in the three analyses. The variance of acquiescence in the analyses with Random Intercept was 0.05 (unbalanced) and 0.04 (balanced) in the BFI-2.

Table 4.

Fit indices of the analyses with the BFI-2

	$\chi^2 (df)$	CFI	TLI	RMSEA
Unbalanced without control	3240.49 (395)	0.86	0.84	0.09
Unbalanced MIMIC	2922.16 (395)	0.88	0.85	0.09
Unbalanced RI	2917.26 (394)	0.87	0.86	0.09
Balanced without control	2222.19 (395)	0.85	0.84	0.07
Balanced MIMIC	1860.23 (395)	0.89	0.87	0.07
Balanced RI	1866.66 (394)	0.88	0.87	0.07

Table 5.
Factor loadings in the analyses with the unbalanced and balanced items of the BFI-2

Items	Unbalanced			Items	Balanced		
	Without control	MIMIC	RI		Without control	MIMIC	RI
BFI1	0.77	0.77	0.73	BFI1	0.74	0.75	0.76
BFI6	0.40	0.41	0.35	BFI6	0.39	0.47	0.40
BFI21	0.52	0.53	0.48	BFI21	0.57	0.65	0.58
BFI41	0.79	0.77	0.78	BFI31-	-0.74	-0.71	-0.75
BFI46	0.58	0.58	0.54	BFI36-	-0.44	-0.40	-0.43
BFI51-	-0.52	-0.49	-0.58	BFI51-	-0.64	-0.57	-0.63
BFI2	0.82	0.82	0.78	BFI2	0.47	0.58	0.47
BFI7	0.74	0.73	0.71	BFI7	0.74	0.81	0.75
BFI27	0.41	0.41	0.38	BFI27	0.31	0.39	0.32
BFI32	0.87	0.85	0.84	BFI37-	-0.72	-0.61	-0.72
BFI52	0.76	0.76	0.72	BFI42-	-0.40	-0.26	-0.39
BFI47-	-0.36	-0.30	-0.45	BFI47-	-0.45	-0.30	-0.44
BFI13	0.61	0.64	0.57	BFI13	0.53	0.56	0.54
BFI18	0.75	0.72	0.72	BFI18	0.81	0.83	0.80
BFI33	0.83	0.80	0.80	BFI33	0.87	0.90	0.87
BFI38	0.61	0.61	0.57	BFI28-	-0.69	-0.65	-0.69
BFI43	0.68	0.71	0.62	BFI48-	-0.83	-0.81	-0.85
BFI58-	-0.54	-0.47	-0.65	BFI58-	-0.60	-0.52	-0.60
BFI14	0.73	0.61	0.71	BFI14	0.81	0.73	0.81
BFI19	0.55	0.46	0.53	BFI19	0.60	0.48	0.59
BFI39	0.91	0.87	0.89	BFI39	0.68	0.55	0.67
BFI54	0.93	0.88	0.90	BFI29-	-0.64	-0.70	-0.66
BFI59	0.71	0.59	0.69	BFI44-	-0.63	-0.70	-0.64
BFI49-	-0.37	-0.41	-0.39	BFI49-	-0.42	-0.48	-0.44
BFI10	0.57	0.57	0.53	BFI10	0.57	0.63	0.58
BFI15	0.46	0.46	0.40	BFI15	0.48	0.59	0.50
BFI20	0.80	0.80	0.77	BFI20	0.53	0.55	0.55
BFI35	0.83	0.83	0.80	BFI45-	-0.75	-0.69	-0.75
BFI40	0.48	0.47	0.41	BFI50-	-0.54	-0.50	-0.55
BFI55-	-0.38	-0.40	-0.51	BFI55-	-0.45	-0.41	-0.44

Note. The sets of six items represent, respectively, the following dimensions: Extraversion, Agreeableness, Conscientiousness, Negative Emotionality, and Open-mindedness. The symbol “-” in the description of some items indicates that they are negatively worded.

Discussion

The present study aimed to compare three modeling approaches (without control, MIMIC method, and Random Intercept) for instruments that may be influenced by acquiescence, using simulated data and a real dataset. Overall, the results pointed to the importance of modeling acquiescence in factor analyses of instruments composed of positive and negative items. Also, the Random Intercept model was slightly superior to MIMIC, and both models varied in performance under different conditions.

Regarding the simulation study, performance varied considerably in the interactions between balancing conditions, control method, and factor loadings sizes. Additionally, the size of acquiescence influenced the models' performance. However, the sample size did not have a significant impact.

For unbalanced scales, there was an overestimation of positive factor loadings (i.e., items in greater quantity) and, above all, an underestimation of negative factor loadings (i.e., items in smaller quantity). Underestimation was, therefore, associated with the loadings of negative items. A plausible explanation for this is that acquiescence can inflate correlations between items worded in the same direction and suppress correlations between items with opposite directions (Danner et al., 2015; Kam & Meyer, 2015; Kuru & Pasek, 2016; Lechner & Rammstedt, 2015). It is worth noting that simply reversing the imbalance, that is, having more negative items than positive ones, will only reverse the problem. Controlling for acquiescence in unbalanced scales is crucial to ensure that the items (positive or negative) that are in smaller quantity in the instrument do not have their factor loadings suppressed by acquiescence.

Conditions with lower content (psychological trait) factor loadings, in general, showed greater estimation bias, especially for unbalanced scales and without any acquiescence control method. Previous research has pointed to the influence of acquiescence on the internal structure of instruments (Kam & Meyer, 2015; Kuru & Pasek, 2016; Lechner & Rammstedt, 2015; Valentini, 2017). The present study advances this literature and provides evidence that some structures may be more affected than others. Items of lower quality (i.e., with a weaker relationship with the latent trait) are more susceptible to acquiescence bias, even when the variance of acquiescence itself is the same. This occurs because lower-quality items have less variance explained by the content factor and, as a result, a

significant proportion of the variance is attributed to acquiescent response bias.

Despite the strong recommendation for modeling acquiescence in instruments with low factor loadings, we emphasize the differences in performance. In these cases, the absence of control is strongly discouraged as it substantially biases the factor loadings, both for unbalanced and balanced scales. In some cases, the bias exceeded 200% on average. The conclusions of the MIMIC and Random Intercept models vary: Random Intercept appears to solve the problem of bias in estimating low factor loadings in both unbalanced and balanced scales, while control through MIMIC seems to solve it only in balanced scales. This may have occurred due to the calculation of the acquiescence indicator used in MIMIC. For unbalanced scales, we used a reduced number of opposite pairs for calculating the acquiescence indicator, which may have considerably reduced its accuracy.

The results for conditions with high factor loadings in the content are less heterogeneous. In this case, we observed low performance only in the absence of control in unbalanced scales. Random Intercepts and MIMIC did not show significant biases in any of the conditions where the factor loadings were high. This was expected because the higher the content loadings, the higher the precision of the factor score. Therefore, high factor loadings can serve as protection against acquiescence, provided that the response bias is modeled or the scale is balanced. This also provides additional support to our conclusion that some structures may be more or less affected by acquiescence.

The size of acquiescence was also relevant for the modeling performance. In general, high acquiescence (i.e., variance of 0.20), without control, led to greater problems in the content loadings, even for balanced scales. This result provides additional evidence to the literature on the strength of acquiescence's influence in worsening the factor structure (Kam & Meyer, 2015; Kuru & Pasek, 2016; Lechner & Rammstedt, 2015; Valentini, 2017).

On the other hand, the sample size does not seem to have a significant influence on the bias generated by acquiescence. It is understood, therefore, that if the model has sufficient statistical power, or a sufficient sample size, acquiescence does not seem to require a larger sample size. Despite this, problems with model identification were more common in small samples, becoming an issue not only for acquiescence control but also for verifying the internal structure of

the instrument. The sample size will depend on the complexity of the model being tested, so there is no standard minimum sample size (Muthén & Muthén, 2002). To determine the ideal sample size, it is recommended to conduct a simple simulation based on the models that will be tested.

Regarding the illustration with real data from the BFI-2, we noticed that the tested acquiescence controls slightly improved the fit of the models. Although the fit was below expectations, this result is not different from what has been found in the literature (Soto & John, 2017). This can be partially attributed to the complexity of factor models for personality. Items in this context are typically multidimensional (i.e., they have cross-loadings), and the estimation of overly restrictive models without cross-loadings significantly reduces the fit (Aichholzer, 2014).

Regarding the factor loadings of the BFI-2, the results with real data are similar to the ones found in the simulations. Apparently, some negative items had their loadings underestimated, especially when testing the “without control” model in the unbalanced composition of the scale. However, for the balanced composition of the scale, we did not observe significant differences in the content factor loadings after controlling for acquiescence, whether through MIMIC or Random Intercepts. We highlight that the sample consisted of adults, and the variance of acquiescence was low (i.e., 0.04 and 0.05). Soto et al. (2008) showed that the variance of acquiescence tends to be higher in childhood and adolescence (i.e., between 10 and 18 years old) and seems to stabilize in adulthood (i.e., 19 and 20 years old).

Final Considerations

This study provides evidence on the importance of item balancing and acquiescence control for factor model estimation. We emphasize that the performance of balanced scales was better and that it is necessary to model acquiescence (using methods like MIMIC and Random Intercepts).

It is worth noting an important point regarding some recommendations against the use of negative items in self-report instruments, arguing that they are psychometrically inferior (Checa & Espejo, 2018; Gehlbach & Artino Jr., 2018; Lai, 1994; Ray et al., 2015; Salazar, 2015; Sliter & Zickar, 2014). Based on this study’s results, we noticed that negative items did not demonstrate inferiority compared to positive items

after controlling for acquiescence bias, as other authors have also suggested (Marsh, 1996; Peterson & Peterson, 1976; Primi et al., 2019a; Tamir, 1993). It is also important to remember that this control can only be operationalized in self-report instruments with Likert-type response scales, using both positive and negative items. Therefore, by mistakenly eliminating negative items because they may seem psychometrically worse, researchers are also impeding the possibility of controlling for acquiescence, which may contribute to biased structure and scores.

As a limitation, we can highlight that we tested a single factor structure composed of five factors. Additionally, we only adopted one condition of item imbalance, which was the same for all factors (i.e., one negative item and five positive items). We also did not test acquiescence biases in models formed solely with positive items. Furthermore, we did not test potential differences in standard errors across conditions, which could indicate greater instability in parameter estimation for some conditions. Moreover, we did not consider the possibility of model misfit, which may limit the adherence of the simulations to reality. We acknowledge that all these unaddressed issues in this article could generate distinctions in acquiescence biases. Therefore, we suggest that future studies be conducted to test acquiescence bias in instruments that address these different conditions of structure and imbalance, including small samples (i.e., fewer than 150 cases). Finally, the sample of real data used in this study consisted of adults, who are reported in the literature to exhibit lower acquiescence compared to adolescents and children. Therefore, we suggest that future studies empirically test acquiescence biases in balanced and unbalanced instruments using samples of children or adolescents.

Despite the limitations, the data presented in the study are sufficient to support the recommendation of constructing instruments with both positive and negative items, preferably balanced. We emphasize that the literature provides little evidence regarding the effect of item balancing in the construction of self-report scales. Furthermore, we suggest using some form of acquiescence modeling (such as MIMIC or Random Intercept) when analyzing the factorial structure of the instrument to avoid errors in its interpretation. Table 6 can serve as a guide for researchers in making decisions regarding scale development and bias modeling, helping them determine the best model for the specific characteristics of the scale being developed.

Table 6.
Recommendations for analysis for the conditions tested in this article

Loadings	Conditions		Unbalanced			Balanced		
	Acq Var	Sample	Without control	MIMIC	RI	Without control	MIMIC	RI
Low	0.20	200	Severe Bias	Light Bias	Light Bias	Severe Bias		
Low	0.20	500	Severe Bias	Light Bias		Severe Bias		
Low	0.20	1000	Severe Bias	Light Bias		Severe Bias		
Low	0.10	200	Severe Bias	Light Bias	Light Bias	Moderate Bias		
Low	0.10	500	Severe Bias	Light Bias		Moderate Bias		
Low	0.10	1000	Severe Bias	Light Bias		Moderate Bias		
Low	0.05	200	Moderate Bias	Light Bias				
Low	0.05	500	Moderate Bias	Light Bias				
Low	0.05	1000	Moderate Bias	Light Bias				
High	0.05	200	Light Bias					
High	0.05	500	Light Bias					
High	0.05	1000	Light Bias					
High	0.10	200	Light Bias					
High	0.10	500	Light Bias					
High	0.10	1000	Light Bias					
High	0.20	200	Moderate Bias					
High	0.20	500	Moderate Bias					
High	0.20	1000	Moderate Bias					

Note. Low loadings vary between 0.30 to 0.50; High loadings vary between 0.60 to 0.80. Unbalanced refers to an instrument with five positive items and one negative item per factor. Balanced refers to an instrument with an equal number of positive and negative items per factor. In the case of this study, there were three positive and three negative items.

References

- Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1–4. <https://doi.org/10.1016/j.jrp.2014.07.001>
- Ambiel, R. A., Moreira, T. D. C., Barros, L. D. O., Martins, G. H., Salvador, A. P., & Wille, B. (2022). Measuring career adaptabilities in the Brazilian context: Development and validation of the CAAS+ C Brazilian Form. *International Journal for Educational and Vocational Guidance, 1*-17. <https://doi.org/10.1007/s10775-022-09523-5>
- Barnette, J.J. (2000). Effects of Stem and Likert Response Option Reversals on Survey Internal Consistency: If You Feel the Need, There is a Better Alternative to Using those Negatively Worded Stems. *Educational and Psychological Measurement, 60*(3), 361-370. <https://doi.org/10.1177/00131640021970592>

- Checa, I., & Espejo, B. (2018). Method Effects Associated with Reversed Items in the 29 Items Spanish Version of Ryff's Well-Being Scales. *Neuropsychiatry (London)*, 8(5), 1533-1540. https://www.researchgate.net/profile/Begona_Espejo/publication/327751696_Method_Effects_Associated_with_Reversed_Items_in_the_29_Items_Spanish_Version_of_Ryff%27s_Well-Being_Scales/links/5bbb69c2a6fdcc9552d992e4/Method-Effects-Associated-with-Reversed-Items-in-the-29-Items-Spanish-Version-of-Ryffs-Well-Being-Scales.pdf
- Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, 57, 119-130. <https://doi.org/10.1016/j.jrp.2015.05.004>
- Gehlbach, H., & Artino Junior, A. R. (2018). The survey checklist (manifesto). *Academic Medicine*, 93(3), 360-366. <https://doi.org/10.1097/ACM.0000000000002083>
- Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*, 13(1), 49-57. <https://doi.org/10.1037/1082-989X.13.1.49>
- He, J., Bartram, D., Inceoglu, I., van de Vijver, J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology*, 45(7), 1028-1045. <https://doi.org/10.1177/0022022114534773>
- Henninger, M. (2019). *Psychometric modeling as a tool to investigate heterogeneous response scale use* [Tese de doutorado, Universität Mannheim]. <https://madoc.bib.uni-mannheim.de/52490/>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- John, O. P. (2021). History, Measurement, and Conceptual Elaboration of the BigFive Trait Taxonomy: The Paradigm Matures. Em O. P. John, & R. W. Robins (Eds.), *Handbook of personality: Theory and Research*, (pp. 35-82). Guilford Press.
- Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, 18(3), 512-541. <https://doi.org/10.1177/1094428115571894>
- Kam, C. C. S., & Zhou, M. (2015). Does acquiescence affect individual items consistently?. *Educational and psychological measurement*, 75(5), 764-784. <https://doi.org/10.1177/0013164414560817>
- Kuru, O., & Pasek, J. (2016). Improving social media measurement in surveys: Avoiding acquiescence bias in Facebook research. *Computers in Human Behavior*, 57, 82-92. <https://doi.org/10.1016/j.chb.2015.12.008>
- Lai, J. C. (1994). Differential predictive power of the positively versus the negatively worded items of the Life Orientation Test. *Psychological Reports*, 75(3), 1507-1515. <https://doi.org/10.2466/pr0.1994.75.3f.1507>
- Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological assessment*, 27(4), 1301-1310. <https://doi.org/10.1037/pas0000151>
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, 48(3), 936-949. <https://doi.org/10.3758/s13428-015-0619-7>
- Lorenzo-Seva, U., & Ferrando, P. J. (2009). Acquiescent responding in partially balanced multidimensional scales. *British Journal of Mathematical and Statistical Psychology*, 62(2), 319-326. <https://doi.org/10.1348/000711007X265164>
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts?. *Journal of Personality and Social Psychology*, 70(4), 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological methods*, 11(4), 344-362. <https://doi.org/10.1037/1082-989X.11.4.344>
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599-620. https://doi.org/10.1207/S15328007SEM0904_8

- Paulhus, D. L. (1991). Measurement and control of response bias. Em J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes: Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press.
- Peterson, C. C., & Peterson, J. L. (1976). Linguistic determinants of the difficulty of true-false test items. *Educational and Psychological Measurement, 36*(1), 161-164. <https://doi.org/10.1177/001316447603600115>
- Plieninger, H. (2018). *Towards a deeper understanding of response styles through psychometrics* [Tese de doutorado, Universität Mannheim]. <https://ub-madoc.bib.uni-mannheim.de/44325>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Primi, R., De Fruyt, F., Santos, D., Antonoplis, S., & John, O. P. (2019a). True or False? Keying Direction and Acquiescence Influence the Validity of Socio-Emotional Skills Items in Predicting High School Achievement. *International Journal of Testing, 1*-25. <https://doi.org/10.1080/15305058.2019.1673398>
- Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2015). Positive and Negative Item Wording and Its Influence on the Assessment of Callous-Unemotional Traits. *Psychological Assessment, 28*(4), 394-404. <https://doi.org/10.1037/pas0000183>
- Rosseel, Y. (2012). Quantitative aspects of blood flow and oxygen uptake in the human forearm during rhythmic exercise. *Journal of Statistical Software, 48*(2), 1-93. <https://pubmed.ncbi.nlm.nih.gov/5329335/>
- Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema, 27*(2), 192-199. <https://doi.org/10.7334/psicothema2014.266>
- Siqueira, M. M. M. (2008). Satisfação no trabalho. Em M. M. M. Siqueira (Org), *Medidas do comportamento organizacional: ferramentas de diagnóstico e gestão* (pp. 265-274). Artmed.
- Sliter, K. A., & Zickar, M. J. (2014). An IRT examination of the psychometric functioning of negatively worded personality items. *Educational and Psychological Measurement, 74*(2), 214-226. <https://doi.org/10.1177/0013164413504584>
- Sliter, K. A., & Zickar, M. J. (2014). An IRT Examination of the Psychometric Functioning of Negatively Worded Personality Items. *Educational and Psychological Measurement, 74*(2), 214-226. <https://doi.org/10.1177/0013164413504584>
- 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*, 117-143. <https://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>
- Tamir, P. (1993). Positive and Negative Multiple Choice Items: How Different Are They?. *Studies in Educational Evaluation, 19*(3), 311-325. Tamir, P. (1993). Positive and negative multiple choice items: How different are they? *Studies in Educational Evaluation, 19*(3), 311-325. [https://doi.org/10.1016/S0191-491X\(05\)80013-6](https://doi.org/10.1016/S0191-491X(05)80013-6)
- Valentini, F. (2017). Editorial: Influência e controle da aquiescência na análise fatorial. *Avaliação Psicológica, 16*(2). <http://dx.doi.org/10.15689/ap.2017.1602.ed>
- Valentini, F., & Hauck Filho, N. (2020). O impacto da aquiescência na estimação de coeficientes de validade. *Avaliação Psicológica, 19*(1), 1-3. <http://dx.doi.org/10.15689/ap.2020.1901.ed>
- Wetzel, E., & Carstensen, C. H. (2015). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment, 33*(5), 352-364. <https://doi.org/10.1027/1015-5759/a000291>
- Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. Em F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 349-363). Oxford University Press.

Zanon, C., Bardagi, M. P., Layous, K., & Hutz, C. S. (2014). Validation of the Satisfaction with Life Scale to Brazilians: Evidences of measurement noninvariance across Brazil and US. *Social Indicators Research*, 119, 443-453. <https://doi.org/10.1007/s11205-013-0478-5>

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. <http://dx.doi.org/10.15689/ap.2018.1704.3.03>

Recebido em: 22/04/2021
Reformulado em: 31/05/2023
Aprovado em: 04/06/2023

About the authors:

Gustavo Henrique Martins is a psychologist, has Master's degree, and Ph.D. candidate in the Graduate School of Psychology Assessment at the University of São Francisco. He is currently the Research Manager at the Laboratory of Sciences for Education (eduLab21) at the Ayrton Senna Institute. He is also a Founding Partner of the CATvante: Computerized Testing Laboratory.

E-mail: gustavoh.martins95@gmail.com

ORCID: <https://orcid.org/0000-0002-5125-2553>

Felipe Valentini is a psychologist and has a Ph.D. in Psychology from the University of Brasília (UnB). He is currently a professor in the Postgraduate Program in Psychology, with a focus on Psychological Assessment, at the University of São Francisco.

E-mail: valentini.felipe@gmail.com

ORCID: <https://orcid.org/0000-0002-0198-0958>

Bruno Bonfá-Araujo is a psychologist, specialized in Neuroscience from the Federal University of São Paulo, and has Master's and Ph.D. degrees in Psychology from the University of São Francisco. He is currently a postdoctoral researcher at the University of Western Ontario in Psychology.

E-mail: brunobonffa@outlook.com

ORCID: <https://orcid.org/0000-0003-0702-9992>;

Adriana Satico Ferraz is a psychologist, with Master's and Ph.D. degrees in Psychology from the University of São Francisco (research area: Psychological and Educational Assessment). She is currently a Professor in the Postgraduate Program in Psychology at the University Center of Brasília (CEUB), in the line of re-search Health, Clinical, and Psychological Assessment.

E-mail: adrianasatico.as@gmail.com

ORCID: <https://orcid.org/0000-0002-9856-0094>

Fernanda Otoni is a psychologist, with Master's and Ph.D. degrees in Psychology from the University of São Francisco, specialized in Neuropsychological Assessment and Child Psychology from Anhembi Morumbi University. She is a Professor in the Postgraduate Program in Psychology Forensic at the University Tuiuti do Paraná (UTP), in the line of research Neuropsychology Forensic.

E-mail: fer_ottoni@hotmail.com

ORCID: <https://orcid.org/0000-0002-9347-7144>

André Pereira Gonçalves has a Master's and Ph.D. in Psychology from the Graduate School of Psychology Assessment at the University of São Francisco. He is a psychologist from UNIFENAS. Currently, he is a professor in the Psychology course at the Federal University of Bahia (UFBA), campus Vitória da Conquista - BA. His research focuses on pathological personality traits and mental health.

E-mail: andregoncalvespsi@gmail.com

ORCID: <https://orcid.org/0000-0002-2470-4040>

Contact:

RuaWaldemar César da Silveira, 105 - Jardim Cura D'ars
Campinas-SP, Brasil
CEP: 13045-510