



Investigating method effects associated with the wording direction of items of the Social Physique Anxiety Scale

Wanderson Roberto da Silva^{1,2} · Giovanna Soler Donofre¹ · Angela Nogueira Neves³ · João Marôco⁴ · Patrícia Angélica Teixeira¹ · Juliana Alvares Duarte Bonini Campos¹

Received: 8 December 2021 / Accepted: 29 June 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Purpose The use of psychometric instruments to measure latent concepts is common. The development of these instruments usually involves mechanisms to reduce response bias, such as the inclusion of reversed items. The aim of this study was to investigate method effects related to the wording direction of the Social Physique Anxiety Scale (SPAS) items, a one-dimensional instrument that assesses individual's level of anxiety when others observe their body.

Methods In total, 152 Brazilian adults (65.8% female) answered 2 formats of the SPAS: the original with 12 items (7 regular and 5 reversed); and a new format with all items written in the same direction (i.e., regular). Both formats were filled out at different times and alternately. Differential item functioning analysis (DIF) and confirmatory factor analysis were conducted.

Results The original SPAS did not fit the data, but after allowing covariances between all reversed items, the fit improved. The wording effect was supported by the DIF, indicating a better fit to the data for the new format with all items worded in the same direction.

Conclusion The wording of the SPAS items had effect on the psychometric properties of instrument. When the wording of the reversed items was modified, the factor model fitted the data. Future studies should take these findings into account and evaluate the SPAS with all items worded in the same direction in different contexts.

Level of evidence Descriptive (cross-sectional) study, Level V.

Keywords Method effect · Item wording direction · Methodological artifact · Scale · Social physique anxiety

Introduction

Methodological issues and wording effects

The use of self-reported psychometric instruments is increasingly common in scientific and clinical settings. The items and response scales of these tools are usually developed based on theoretical and methodological frameworks [1, 2]. One of the strategies to reduce response bias, such as acquiescence—the tendency to respond positively to items irrespective of the content—is including items worded in opposite directions, but with equivalent content to evaluate the same construct [3, 4].

Positive statements or statements that directly evaluate the construct (e.g., happiness) are generally considered

regular items (e.g., “*I feel happy*”). Negative statements are considered reversed items, and can be written using a word with an opposite meaning (e.g., “*I feel sad*”) or adding a negative word or expression of the regular item (e.g., “*I don't feel happy*”). When regular and reversed items are used simultaneously to evaluate a construct, the responses for reversed items are reversed and combined with the responses of the regular items, except when reversed item is an opposite statement to a regular item (e.g., “*I don't feel sad*”) [5].

Some researchers suggest that combining regular and reversed items can help reduce response bias, especially in one-dimensional measures [6]. However, not all reversed statements have an exact reversed meaning; for example, stating that “*the weekend was good*” can be different from saying that “*the weekend was not bad*” [7]. In addition, regular items have been shown to provide more accurate answers [5] and reversed items are more challenging to be well understood [1, 2, 6, 8–10]. Therefore, reliable answers

✉ Juliana Alvares Duarte Bonini Campos
juliana.campos@unesp.br

Extended author information available on the last page of the article

to reversed items depend on the respondent's ability to carefully read and interpret the item and the response scale.

Studies have shown that the use of regular and reversed items in an instrument can produce different psychometric results compared to having all items in the same wording direction [11–13]. Thus, to achieve good validity and reliability estimates, instruments with both regular and reversed items may need different dimensions to separate positive and negative statements or have reversed items excluded [14]. In such cases, the existence of a methodological artifact associated with the wording of the items should be investigated to control or minimize response bias as this can affect the instrument's psychometric properties and make the interpretation of results unclear [10, 11, 15–18].

An instrument that has a method effect associated with the wording direction of the items is the Rosenberg Self-Esteem Scale (RSES). This scale was developed to investigate self-esteem with a one-dimensional model including regular and reversed items. In view of the poor-fit of the RSES original model in different contexts, researchers found that the poor psychometric properties were related to the wording of items [16, 18, 19]. Other instruments have shown a method effect associated with the wording of items, such as the Revised Life Orientation Test (LOT-R) [8], the Maslach Burnout Inventory—Student Survey (MBI-SS) [20], and the Social Physique Anxiety Scale (SPAS) [15, 17, 21].

The present study

The main purpose of this study was to investigate method effects related to the wording direction of the SPAS items. This scale was developed to assess social physical anxiety [22], which is an affective reaction of an individual when their body is judged, factually or hypothetically, by other people [23]. In general, individuals seek to make good impressions on others, especially when it comes to physical characteristics, because there is a “*beauty-is-good*” stereotype [24]. When feeling unable to obtain positive reactions, some people may react negatively, which might lead to low self-esteem [25] among other issues.

The SPAS was originally proposed by Hart, Leary, and Rejesk [22] as a one-dimensional scale, consisting of 12 items, 7 of which are negative statements (regular items: 3, 4, 6, 7, 9, 10, and 12) and 5 are positive ones (reversed items: 1, 2, 5, 8, and 11). This structure has been questioned due to the poor goodness-of-fit indices found in the literature. The first considerations on the wording effect of the SPAS were made based on second item, which is: “*I would never worry about wearing clothes that might make me look too thin or overweight*” [26, 27]. The studies suggested removing the expression “*would never*” to match the regular items. This change was adopted in subsequent studies [28–30] and

second item was maintained in the factorial model, as it presented a better factor loading.

The other four reversed items of the SPAS have also seem to hinder a good fit of the model. Therefore, some reduced or modified models of the SPAS, in which one or more reversed items are excluded, are being proposed as more suitable. As far as we know, at least 19 different SPAS factorial models are available (see Supplementary Material). Although there is a tendency to use one-dimensional models, the two-dimensional model containing two correlated factors (“*Expectations of Negative Physique Evaluation*” and “*Comfort with Physique Presentation*”) that separate regular and reversed items is also being used [27, 31, 32]. Interestingly, when second item is presented as a regular item, it loads on the “*Expectations of Negative Physique Evaluation*” factor [33]. However, as the two-factor model is formed by different concepts built on the wording direction of the items, it has been criticized [17, 32].

In an attempt to maintain the original proposal of the SPAS to evaluate a single construct and to overcome the problem related to the two-factor model, a higher order factor model (i.e., two first-order factors subordinate to one second-order factor) was proposed [27, 31, 33]. However, this second-order model has also been criticized, as the “*Comfort with Physique Presentation*” is not exactly a dimension of social physique anxiety, but a concept related to it [34]. Although a considerable number of studies have assessed the validity and reliability of the SPAS [15, 17, 21, 28, 32], a more detailed investigation on the influence of the wording of the items on psychometric properties of the instrument is scarce. As far as we know, no study has investigated the effect of modifying SPAS items' wording so that all are worded in the same direction.

Another important aspect related to the SPAS is its several cross-cultural adaptations. The tool has been adapted for use in China, Japan and Korea [35], Portugal [36], Sweden, Estonia, and Turkey [28], France [29], Spain [28, 37], and Brazil [38–40]. In some occasions, there is more than one version available to be used in the same culture. For example, two independent studies with Spanish speakers have carried out the cross-cultural adaptation of the SPAS for use in Spain [28, 37]. The same occurred with studies carried out in Portugal [36, 41] and Brazil [38–40]. These different adaptations may lead to disparity in SPAS results, since the idiomatic content of the items is not standardized. Therefore, developing a unified Portuguese-language version of the SPAS—based on existing ones—will be relevant for use in future protocols.

Objectives and hypothesis

The aim of this study was to investigate the wording direction of the SPAS items as a possible response bias. Data

were collected in Brazil by applying the scale in 2 formats: (A) the original with 12 items—7 written as regular statements and 5 written as reversed statements; and (B) the new proposal with all items in the same wording direction, i.e., as regular statements. Our hypothesis was that the SPAS' one-dimensional model with all items written in the same direction (i.e., regular items) has better psychometric properties when compared to the original with items written in opposite directions (i.e., regular and reversed items). The secondary objective was to develop a unified Portuguese-language version of the SPAS based on the available versions in the literature.

Methods

Participants

Participants 18 to 40 years old were recruited from a public university in the state of São Paulo, Brazil, with a non-probability method, as data were collected in September and October 2021, when restrictions were imposed due to the novel coronavirus disease (COVID-19). This sample is in line with the original SPAS study, which was developed based on responses from individuals who were at a university [22]. The exclusion criteria were: being pregnant or breastfeeding, blindness, having been diagnosed by a clinician with a mental disorder in the last 6 months (self-reported), and low educational level (i.e., incomplete elementary [basic education] or secondary school). Information about age, education level, physical exercise, monthly family income, weight, and height were included in the questionnaire. Educational level was investigated using the Brazilian Criteria [42]. Self-reported body weight and height were used to calculate the body mass index (BMI) and for anthropometric nutritional status classification [43]. It should be mentioned that the characteristics of the sample (e.g., non-clinical, adults, both sexes) were chosen based on a previous cross-sectional study carried out with SPAS [32], which the hypothesis of the method effect associated with the wording direction of items was raised.

In total, 165 people were first included in the study, but 13 (7.9%) did not complete the second SPAS format and were excluded. Thus, the final sample was composed of 152 individuals. This sample size was adequate, as the minimum calculated was 134 individuals using a ratio of 5 participants per parameter (k) of the original SPAS model (k : 12 factor loadings + 12 residuals) and a dropout rate of 10% [44].

Most participants were women (62.5%), the average age was 27.7 (standard deviation [SD]=5.2) years, the majority (71.1%) reported having higher education and 28.3% of the rest of the sample reported having started higher education, and 0.6% indicated having completed the secondary school.

A total of 65.6% of the sample reported performing physical exercise, 13.8% had a monthly income higher than R\$ 11,262 (Brazilian reals), 14.5%, from R\$ 8,641 to R\$ 11,261, 60.5% from R\$ 2,005 to R\$ 8,640, 7.2% from R\$ 1,255 to R\$ 2,004 and 4.0% less than R\$ 1,254 (exchange rate in June 2022 was 1 US dollar to 5.23 Brazilian reals). Mean BMI of the sample was 25.0 ($SD=4.9$) kg/m² and normal weight was the most prevalent nutritional status category (56.6%), followed by overweight (27.0%), obesity (13.2%), and underweight (3.2%).

Measure

Respondents filled the SPAS items based on how the statement is characteristic of them using a five-point *Likert*-type response scale (1 = “not at all characteristic of me” to 5 = “extremely characteristic of me”). Although the factorial model originally proposed was one-dimensional with 12 items [22], a wide variety of alternative models is available in the literature (see Supplementary Material). Because of the possibility of a methodological artifact due to items wording, the present study investigated SPAS in its original model.

First, the available versions of the SPAS from Brazil [38–40] and from Portugal [36] were used for the development of a unified Portuguese-language version by collaborative researchers from the Body Image, Physical Exercise, and Psychometrics fields of study native to Brazil ($N=2$) and Portugal ($N=2$). The process involved using words and expressions that could be well understood in both countries while keeping the original content following international protocols [45, 46]. After building the unified Portuguese-language version of SPAS, the five reversed items were modified to be in a similar wording as the others, that is, as negative statements in relation to social physique anxiety. Thus, we obtained two formats of the scale: SPAS-A: the original with seven regular items and five reversed items; and SPAS-B: the new proposal with all items written in the same direction (i.e., as regular items). Table 1 shows the two formats of SPAS. Importantly, we requested and received the SPAS author' authorization to use it in this study.

Procedure

First, three researchers were trained to perform data collection to standardize procedures and avoid bias. The study was announced by email and in social media to students and staff of the university where the study was conducted. We received digital feedback from 171 people who were scheduled to come to the university, in groups of a maximum of 6 people. On the day of data collection, individuals who met the eligibility criteria ($n=165$) received the Free and Informed Consent Term with detailed information about

Table 1 English and unified Portuguese-language versions of the Social Physique Anxiety Scale (SPAS)

Item	English version	Item	Unified Portuguese-language version (SPAS-B)
	For each item, respondents indicate the “degree to which the statement is characteristic or true of you” on a 5-point scale (not at all, slightly, moderately, very, extremely characteristic)		<i>Opções de resposta:</i> (1) <i>nada característico ou verdadeiro para mim</i> (2) <i>ligeiramente característico ou verdadeiro para mim</i> (3) <i>moderadamente característico ou verdadeiro para mim</i> (4) <i>muito característico ou verdadeiro para mim</i> (5) <i>extremamente característico ou verdadeiro para mim</i>
1*	I am comfortable with the appearance of my physique/figure	1*	<i>Eu estou tranquilo com a aparência do meu corpo</i>
1†	I am not comfortable with the appearance of my physique/figure	1†	<i>Eu não estou tranquilo com a aparência do meu corpo</i>
2*	I would never worry about wearing clothes that might make me look too thin or overweight	2*	<i>Eu nunca iria me preocupar em vestir roupas que pudessem me fazer parecer muito magro ou acima do peso</i>
2†	I worry about wearing clothes that might make me look too thin or overweight	2†	<i>Eu me preocupo em vestir roupas que pudessem me fazer parecer muito magro ou acima do peso</i>
3	I wish I was not so uptight about my physique/figure	3	<i>Eu queria não ser tão tenso com relação ao meu corpo</i>
4	There are times when I am bothered by thoughts that other people are evaluating my weight or muscular development negatively	4	<i>Tem horas que eu fico chateado por pensar que outras pessoas estão avaliando meu peso ou meu desenvolvimento muscular negativamente</i>
5*	When I look in the mirror I feel good about my physique/figure	5*	<i>Eu me sinto bem quando vejo meu corpo no espelho</i>
5†	When I look in the mirror I do not feel good about my physique/figure	5†	<i>Eu não me sinto bem quando vejo meu corpo no espelho</i>
6	Unattractive features of my physique/figure make me nervous in certain social settings	6	<i>As características pouco atraentes do meu corpo me deixam nervoso em certos ambientes sociais</i>
7	In the presence of others, I feel apprehensive about my physique/figure	7	<i>Na presença dos outros, eu me sinto apreensivo quanto ao meu corpo</i>
8*	I am comfortable with how fit my body appears to others	8*	<i>Eu estou confortável em relação ao que os outros acham do meu corpo</i>
8†	I am uncomfortable with how fit my body appears to others	8†	<i>Eu não estou confortável em relação ao que os outros acham do meu corpo</i>
9	It would make me uncomfortable to know others were evaluating my physique/figure	9	<i>Eu ficaria desconfortável se soubesse que outras pessoas estão avaliando meu corpo</i>
10	When it comes to displaying my physique/figure to others, I am a shy person	10	<i>Quando vou exibir meu corpo para os outros, eu sou uma pessoa tímida</i>
11*	I usually feel relaxed when it is obvious that others are looking at my physique/figure	11*	<i>Eu normalmente me sinto relaxado quando percebo que os outros estão olhando meu corpo</i>
11†	I usually do not feel relaxed when it is obvious that others are looking at my physique/figure	11†	<i>Eu normalmente não me sinto relaxado quando percebo que os outros estão olhando meu corpo</i>
12	When in a bathing suit, I often feel nervous about the shape of my body	12	<i>Quando estou de roupa de banho, eu normalmente me sinto nervoso com a forma do meu corpo</i>

*Reversed item (original). SPAS-B: revised version with all items with wording in the same direction (i.e., regular items)

†Modified item (proposed in this study). The unified Portuguese-language version was developed from the content published in Brazil by Hart (2003), by Souza and Fernandes (2009), and by Campana (2011) and in Portugal by Malheiro and Gouveia (2001)

the study and all of them signed the document voluntarily agreeing to participate in the research that was approved by the Ethics Committee from the university where the study was conducted.

Then, the participants were randomly divided into two groups to complete the survey, using paper and pen, in a room with chairs distanced by 1.5 m. All participants first completed the personal characteristics form. The SPAS items were completed in two stages, in a crossover design. First, half of the participants received, at random (according to an alphanumeric code), the SPAS original format (SPAS-A) or the new one (SPAS-B). After a washout period of 7

days, the participants returned to the lab and filled the format that was not answered in the previous session. A total of 152 individuals completed the 2 stages.

Data analysis

The psychometric sensitivity of the SPAS items was verified for both formats of the scale (i.e., SPAS-A and SPAS-B) by calculating mean, median, mode, standard deviation, skewness, and kurtosis. For skewness and kurtosis, absolute values less than 2, indicated the absence of a severe violation of the assumption of normality of data distribution [44, 47].

Then, differential item functioning (DIF) was used to verify the measurement equivalence of the two formats of scale (i.e., SPAS-A and SPAS-B) by ordinal logistic regression based on the likelihood ratio χ^2 test using a significance level of 1%. The responses given by the participants to the reversed items were recoded to match the responses to the regular items. DIF is classified as uniform (i.e., if the effect is constant) or non-uniform (i.e., if the effect varies), but we performed an overall test of “total DIF effect” aiming to control for Type I error to identify both uniform and non-uniform effects. Items with a “total DIF effect” ($p < 0.01$) were considered non-equivalent, i.e., changing only one word/term were perceived by participants differently [48].

The psychometric properties of the SPAS’ one-dimensional model considering the two formats were investigated based on the proposal by Anastasi [3] and as described by Marôco [44]. These authors suggest evaluating the construct validity (i.e., factorial, convergent, and discriminant), as well as verifying the reliability of the data. Factorial validity was used to verify the wording effect of the SPAS items. The correlated traits–correlated uniqueness (CTCU) was used as described in previous studies [15, 19]. The CTCU treats the wording effect of the items as a methodological artifact controlling for the wording direction, since the errors of all items written in the same direction are correlated. Using the data set from the SPAS-A (original), four one-dimensional models were tested: (1) model without correlated uniqueness between items (SPAS-A); (2) model including correlated uniqueness between reversed items (SPAS-A⁺); (3) model including correlated uniqueness between regular items (SPAS-A⁻); (4) model including correlated uniqueness between both reversed and regular items (SPAS-A[±]). The presence of method effects was determined when the fit of the models to the data including correlated uniqueness (SPAS-A⁺, SPAS-A⁻, and SPAS-A[±]) was better compared to the model without correlated uniqueness (SPAS-A). A fifth one-dimensional model was tested using the dataset from the SPAS-B, without correlated uniqueness between items.

Confirmatory factor analysis (CFA) was carried out in the five models to test the fit to the data using the robust Weighted Least Squares Mean and Variance Adjusted (WLSMV) estimation method. The goodness-of-fit was analyzed by the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), and the Standardized Root Mean Square Residual (SRMR). The fit was considered good when CFI and TLI > 0.95 and SRMR < 0.08 [49, 50]. The factor loading (λ) for each item of the scale was calculated and values greater than 0.50 were considered adequate [47]. Convergent validity was investigated based on the proposal by Fornell and Larcker [51], which is supported by Hair et al. [52] and Marôco [44]. The average variance extracted (AVE) was calculated using the factor loadings of the items and values ≥ 0.50 were considered adequate. Reliability was verified from the omega (ω) [53] and ordinal

alpha (α) coefficients [54, 55]. The nonlinear structural equation modeling reliability coefficient (ρ_{NL}) was also calculated [56, 57]. For all coefficients, values greater than 0.70 were considered adequate [58].

Descriptive analyses were performed using SPSS Statistics for Windows, Version 22.0 (Armonk, NY: IBM Corp.). Psychometric analyses were performed in RStudio, Version 2022.02.0+443 (RStudio Team, 2022) with the lavaan [59], semTools [60], pbivnorm [61], psych [62], and lordif [48] packages.

Results

Table 2 shows the descriptive statistics of the SPAS responses of the two formats tested. No severe violation of data normality was found indicating the adequate psychometric sensitivity of the items. The results of the DIF analysis (see Table 2) indicated that the reversed items were not equivalent ($p < 0.01$) between the two formats. For the regular items, equivalence was observed ($p > 0.01$).

Table 3 shows the goodness-of-fit indices for each model tested considering the two formats applied of the SPAS. The original factor model (SPAS-A) presented indices lower than the cutoff points recommended for a good fit. Furthermore, the fit of the models tested with CTCU (i.e., SPAS-A⁺, SPAS-A⁻, and SPAS-A[±]) was adequate to the data, which shows the occurrence of method effect associated with items wording. When the SPAS factor model was tested with all items written in the same direction (i.e., new proposal—SPAS-B), an adequate fit was found (see Table 3), which indicates method effect associated with items wording. The factor loadings of the items were all above the recommended value ($\lambda > 0.50$) in both formats, but the SPAS-B had higher estimates (see Fig. 1).

The convergent validity and reliability estimates are shown in Table 3. The AVE values were adequate in both formats (values > 0.50), but a higher one was found for the SPAS-B (0.74) compared to SPAS-A (0.60). With regard to data reliability, all the calculated coefficients were adequate (values > 0.70), but higher values were identified for the SPAS-B (0.94–0.97) compared to SPAS-A (0.82–0.94). Importantly, a strong and significant correlation ($r = 0.96$, $p < 0.001$) was observed between the two formats of the SPAS, confirming the preservation of the theoretical construct from one version to the other (see Fig. 1).

Discussion

This study examined—using a crossover design—the potential effect of the wording direction of the SPAS items to produce bias. The fit of the SPAS-A, which included regular and reversed items, was poor, corroborating

Table 2 Descriptive statistics of items of the Social Physique Anxiety Scale (SPAS) and differential item functioning (DIF) analysis

Item	SPAS-A						SPAS-B						DIF [‡]
	Mean	Median	Mode	SD	Skewness	Kurtosis	Mean	Median	Mode	SD	Skewness	Kurtosis	
1 [†]	2.79	3.00	3.00	1.10	-0.12	-0.88	2.70	3.00	2.00	1.16	0.38	-0.58	<0.001*
2 [†]	2.59	3.00	3.00	1.18	0.16	-0.98	2.59	3.00	3.00	1.22	0.27	-0.87	<.001*
3	2.87	3.00	2.00	1.30	0.27	-1.09	2.72	2.00	2.00	1.38	0.31	-1.19	0.602
4	2.60	2.00	2.00	1.27	0.42	-0.89	2.47	2.00	1.00	1.26	0.45	-0.83	0.259
5 [†]	2.69	3.00	3.00	1.02	-0.13	-0.69	2.55	2.00	2.00	1.13	0.44	-0.57	<.001*
6	2.53	2.00	2.00	1.21	0.51	-0.70	2.51	2.00	2.00	1.25	0.52	-0.71	0.448
7	2.24	2.00	2.00	1.14	0.81	-0.08	2.20	2.00	2.00	1.14	0.78	-0.18	0.920
8 [†]	2.76	3.00	3.00	1.05	-0.20	-0.97	2.25	2.00	2.00	1.19	0.78	-0.28	<.001*
9	3.27	3.00	3.00	1.26	-0.10	-1.08	3.10	3.00	3.00	1.26	-0.01	-1.05	0.624
10	3.00	3.00	3.00	1.18	0.12	-0.82	2.93	3.00	2.00	1.22	0.08	-1.02	0.458
11 [†]	1.91	2.00	1.00	.94	0.63	-0.71	2.96	3.00	2.00	1.24	0.10	-1.01	<0.001*
12	2.83	3.00	2.00	1.25	0.25	-0.93	2.74	3.00	2.00	1.29	0.30	-0.93	0.480

SD standard deviation, SPAS-A items with original wording (i.e., seven regular items and five reversed items), SPAS-B all items with wording in the same direction (i.e., regular items)

[†]Reversed items in the SPAS-A version

[‡]*p* value for differential item functioning using for Chi-square test (SPAS-A vs. SPAS-B)

**p*<0.05

Table 3 Psychometric properties of the tested one-factor models of the Social Physique Anxiety Scale (SPAS)

Model	CFI	TLI	SRMR	λ	AVE	α	ω	ρ_{NL}
SPAS-A	0.94	0.93	0.08	0.59—0.86	0.60	0.94	0.93	0.82
SPAS-A ⁺	0.97	0.96	0.06	0.53—0.88	0.56	0.94	0.89	0.82
SPAS-A ⁻	0.98	0.97	0.04	0.58—0.87	0.51	0.94	0.92	0.82
SPAS-A [±]	1.00	1.00	0.00	0.59—0.91	0.51	0.94	0.80	0.82
SPAS-B	0.97	0.96	0.06	0.81—0.92	0.74	0.97	0.96	0.94

SPAS-A items with original wording (i.e., seven regular items and five reversed items), SPAS-A⁺ items with original wording including correlated uniqueness between all reversed items (1, 2, 5, 8, and 11), SPAS-A⁻ items with original wording including correlated uniqueness between all regular items (3, 4, 6, 7, 9, 10, and 12), SPAS-A[±] items with original wording including correlated uniqueness between both reversed and regular items, SPAS-B all items with wording in the same direction (i.e., regular items), CFI comparative fit index, TLI Tucker–Lewis index, SRMR standardized root mean square residual, λ factor loading, AVE average variance extracted, α ordinal alpha coefficient, ω omega coefficient, ρ_{NL} nonlinear structural equation modeling reliability coefficient

previous studies [15, 17, 27, 28, 31, 32, 35, 63]. For this reason, most studies make adaptations to the scale, such as items exclusion, insertion of covariance, and construction of dimensions. In this way, a question that arises is: *Why is the SPAS such an unstable instrument?*

Some studies point out that the weakness of the SPAS one-dimensional model may be related to items with regular and reversed wording to evaluate a single construct [15, 17, 33]. According to the results of our DIF analysis, the reversed items were not equivalent to their counterparts in the A and B formats, suggesting that the statements are interpreted differently by respondents. As the conceptual meaning of the reversed statements did not change, but

the responses were affected, the presence of a response bias is suspected.

Most of the SPAS factor models described in the literature have excluded at least one reversed item, usually items two or eleven, which are generally the most problematic ones, as they include opposite content in relation to others [28]. When maintained, second item had the wording often modified before the scale was applied, improving its performance [28–30, 33, 64]. Although the inclusion of items with regular and reversed wording is common in the development of an instrument [11], our findings suggested that this does not seem to be an effective method for the SPAS, as it could be the artifact that contributed to the factorial instability of

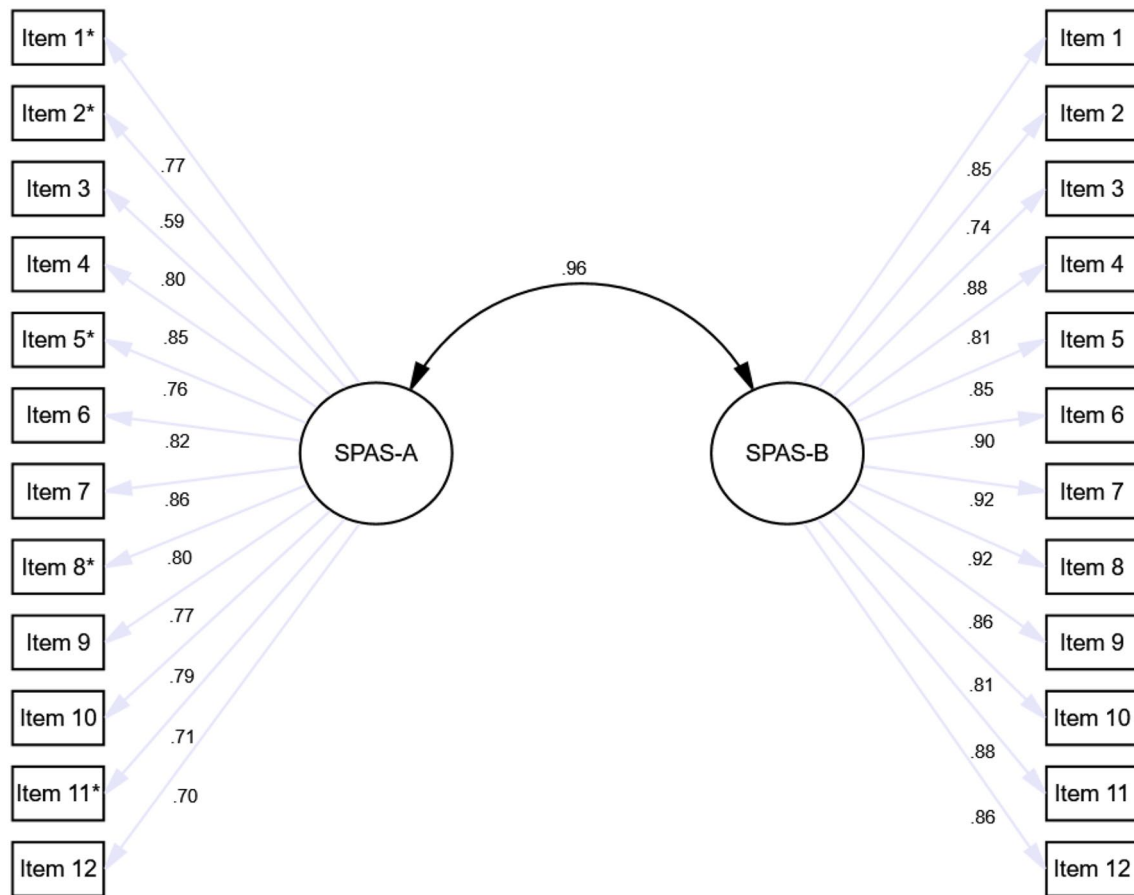


Fig. 1 Factor loadings of the items of the Social Physique Anxiety Scale (SPAS) considering the two formats. *SPAS-A* items with original wording (i.e., seven regular items and five reversed items), *SPAS-B* all items with wording in the same direction (i.e., regular items). *Reversed item

the scale. Therefore, the evaluation of method effects associated with wording of the SPAS items was the main objective of the present work.

Based on the CFA results, a method effect associated with items wording was found, which corroborates with previous research that analyzed models with correlated uniqueness between residual variances of items worded in the same direction (i.e., applying CTCU) and found adequate fit in different contexts [15, 17, 28]. However, new factor models were still proposed for the SPAS, some with different models for similar contexts. Therefore, another question that arose was: *If all the SPAS items were formulated in the same direction, would the instrument have better psychometric properties?* To assess the issue, a literature search was performed on how to appropriately modify the items of the SPAS.

According to Dalal and Carter [11], few studies present plausible justifications for the inclusion of positive and negative-formulated items to evaluate a single concept. The authors also mention that instruments that measure anxiety-related factors may require the inclusion of negatively written items in an attempt to arouse feelings of

“inadequacy” and identify risk behaviors. On the other hand, Roszkowski and Soven [6] and Kam and Fan [65] report that some people have greater difficulty in correctly answering negative items and thus, positive statements can work better. However, the subject is controversial, as demonstrated by the background information provided by researchers themselves. As Kam and Fan [65] suggest, investigating what makes negatively written items more difficult to understand may be more productive than debating whether or not to include this type of formulation in instruments.

As the reversed items of the SPAS (which are positively worded) are fewer than the regular ones, we chose to modify them to negatively worded sentences to have all items in the same direction, a technique used in other instruments with method effects [6, 8, 44]. This modification substantially improved the psychometric properties of the one-dimensional model of the scale in our sample. In addition, we found a strong correlation between SPAS-A (original with regular and reversed items) and SPAS-B (new proposal with all items written in the same direction) suggesting that

the modification did not interfere with the concept being measured.

Based on our findings, we believe that the best solution to assess social physique anxiety as a one-dimensional concept is to use the SPAS-B, which includes all items formulated in the same direction. However, as the results of the models tested with CTCU are not exclusive to identify the acquiescence effect, we are in agreement with the suggestion of Alessandri et al. [8] and Kam [66] that the decision of including or not items with negative and positive wording should be evaluated on a case-by-case basis. Roszkowski and Soven [6] suggest that when designing an instrument with regular and reversed items, strategies must be used to reduce response bias, such as balancing the number of items in each direction and including an alert note in the instructions about reversed items, which demand more attention. These and other strategies can be tested in future studies using SPAS containing reversed items and not just regular ones.

Finally, this study presents a unified Portuguese-language version of the SPAS that was well understood by the participants in both formats. This version can serve as a starting point for future studies in Portuguese-speaking countries. However, a pilot study should be carried out before applying this version in countries other than Brazil, to verify whether conceptual and cultural equivalences [45, 46] are adequate for the context, as countries that have the same official language might not share cultural aspects. As the modifications of the reversed items were minor, the English version of the SPAS with all the items worded in the same direction is also presented in this study. This can contribute to the conduct of future research in English speaking cultures.

Strength and limits

The use of SPAS with the items worded in the same direction can be useful for clinical practice, especially for psychologists and fitness coaches who can explore why certain people are anxious about their body appearance, while others are not. With these results in hand, professionals can develop strategies to reduce the related symptoms, such as discouraging frequent body evaluations and promoting discussions that contribute to the well-being of the population.

The sample characteristics and the data analysis used can be considered the main limitations of the study. We used a non-probability sampling method, recruiting mostly young adults, of upper middle class, and with high educational level, which jeopardize the generalizability of the results. We use self-reported weight and height to calculate BMI and this may not reflect the truth about anthropometric nutritional status, which is a limitation of the study. The SPAS with the items worded in the same direction must undergo evaluations before being applied in other contexts, such as in older populations and people with lower educational levels.

In addition, we used the CTCU and DIF analysis to investigate the method effect associated with the writing of the items, but other techniques are available, such as the correlated traits–correlated methods (CTCM).

Conclusion

Using a psychometrically sound measure of social physique anxiety is important for interventions that aim to prevent and treat the anxiety that some people experience when their physical appearance is negatively judged by others. In the present study, we found that the wording of the SPAS items had an effect on the data collected and the modification of the reversed statements significantly improved the estimates of validity and reliability of the one-dimensional model, indicating that the use of the SPAS with all items worded in the same direction is preferable. A unified Portuguese SPAS version is also presented, contributing to its future applications in Portuguese-speaking countries.

What is already known on this subject?

The inclusion of items with regular and reversed wording is common in the development of psychometric instruments; however, this does not seem to be an effective method for the Social Physique Anxiety Scale (SPAS). Different factorial models of this scale are presented in the literature, as poor psychometric properties of the one-dimensional model are commonly found and may be related to the reversed items.

What this study adds?

This study examined and found a potential effect of the wording direction of the reversed items of the Social Physique Anxiety Scale (SPAS) to produce bias. When the wording of these items was modified, the psychometric properties of the one-dimensional model improved substantially, indicating that formulating all items in the same direction can produce valid and reliable data to assess social physical anxiety.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40519-022-01439-x>.

Author contributions WRDS: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; roles/writing—original draft. GSD: data curation; formal analysis; roles/writing—original draft. ANN: investigation; methodology; writing—review and editing. JM: conceptualization; formal analysis; writing—review and editing. PAT: conceptualization; data curation; writing—review and editing. JADBC: conceptualization; funding acquisition; methodology; project administration; supervision; writing—review and editing.

Funding This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES)—Finance Code 001, and by the São Paulo Research Foundation (grant numbers 2017/20315-7, 2018/21467-8, 2019/18941-2, 2019/19590-9).

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of São Paulo State University, in São Paulo, Brazil (Ethics approval number: 88600318.3.0000.5416).

Consent to participate Informed consent was obtained from all the individual participants included in the study.

Consent for publication Not applicable.

References

- DeVellis RF (2016) Scale development: theory and applications, 5th edn. Sage, Thousand Oaks, CA
- Spector PE (1992) Summated rating scale construction: an introduction. Sage, Newbury Park, CA
- Anastasi A (1988) Psychological testing, 6th edn. Macmillan Publishing Company, New York
- Nunnally JC (1978) Psychometric theory, 2nd edn. McGraw-Hill, New York
- Schriesheim CA, Eisenbach RJ (1995) An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures. *J Manag* 21(6):1177–1193. [https://doi.org/10.1016/0149-2063\(95\)90028-4](https://doi.org/10.1016/0149-2063(95)90028-4)
- Roszkowski MJ, Soven M (2010) Shifting gears: consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assess Eval High Educ* 31(1):113–130. <https://doi.org/10.1080/02602930802618344>
- Colston HL (1999) “Not good” is “bad”, but “not bad” is not “good”: an analysis of three accounts of negation asymmetry. *Discourse Process* 28(3):237–256. <https://doi.org/10.1080/01638539909545083>
- Alessandri G, Vecchione M, Fagnani C, Bentler PM, Barbaranelli C, Medda E, Nisticò L, Stazi MA, Caprara GV (2010) Much more than model fitting? Evidence for the heritability of method effect associated with positively worded items of the life orientation test revised. *Struct Eq Mod* 7(4):642–653. <https://doi.org/10.1080/10705511.2010.510064>
- Billiet JB, McClendon MJ (2000) Modeling acquiescence in measurement models for two balanced sets of items. *Struct Eq Mod* 7(4):608–628. https://doi.org/10.1207/S15328007SEM0704_5
- van Sonderen E, Sanderman R, Coyne JC (2013) Ineffectiveness of reverse wording of questionnaire items: let’s learn from cows in the rain. *PLoS ONE* 8(7):e68967. <https://doi.org/10.1371/journal.pone.0068967>
- Dalal DK, Carter NT (2015) Negatively worded items negatively impact survey research. In: Lance CE, Vandenberg RJ (eds) More statistical and methodological myths and urban legends. Routledge Taylor and Francis Group, New York/London, pp 112–132
- Nunnally JC, Bernstein I (1994) Psychometric theory, 3rd edn. McGraw-Hill Humanities, New York
- Weems GH, Onwuegbuzie AJ, Lustig D (2003) Profiles of respondents who respond inconsistently to positively- and negatively-worded items on rating scales. *Assess Eval High Educ* 17(1):45–60. <https://doi.org/10.1080/14664200308668290>
- Barnette JJ (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. *Educ Psychol Meas* 60(3):361–370. <https://doi.org/10.1177/00131640021970592>
- DiStefano C, Motl RW (2006) Further investigating method effects associated with negatively worded items on self-report surveys. *Struct Eq Mod* 13(3):440–464. https://doi.org/10.1207/s15328007sem1303_6
- Marsh HW (1996) Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? *J Pers Soc Psychol* 70(4):810–900. <https://doi.org/10.1037//0022-3514.70.4.810>
- Motl RW, Conroy DE (2000) Validity and factorial invariance of the social physique anxiety scale. *Med Sci Sports Exerc* 32(5):1007–1017. <https://doi.org/10.1097/00005768-200005000-00020>
- Tomás JM, Oliver A (1999) Rosenberg’s self-esteem scale: two factors or method effects. *Struct Eq Mod* 6(1):84–98. <https://doi.org/10.1080/10705519909540120>
- Horan PM, DiStefano C, Motl RW (2003) Wording effects in self-esteem scales: methodological artifact or response style? *Struct Eq Mod* 10(3):435–455. https://doi.org/10.1207/S15328007SEM1003_6
- Maroco J, Maroco AL, Campos JADB (2014) Student’s academic efficacy or inefficacy? An example on how to evaluate the psychometric properties of a measuring instrument and evaluate the effects of item wording. *OJS*. <https://doi.org/10.4236/ojs.2014.46046>
- Motl RW, Conroy DE, Horan PM (2000) The social physique anxiety scale: an example of the potential consequence of negatively worded items in factorial validity studies. *J Appl Meas* 1(4):327–345
- Hart EA, Leary MR, Rejesk WJ (1989) The measurement of social physique anxiety. *J Sport Exerc Psychol* 11(1):94–104. <https://doi.org/10.1123/jsep.11.1.94>
- Davison TE (2012) Body image in social contexts. In: Cash T (ed) Encyclopedia of body image and human appearance. Elsevier, London, pp 243–249
- Griffin AM, Langlois JH (2006) Stereotype directionality and attractiveness stereotyping: is beauty good or is ugly bad? *Soc Cogn* 24(2):187–206. <https://doi.org/10.1521/soco.2006.24.2.187>
- Cash TF, Smolak L (2011) Body image: a handbook of science, practice, and prevention, 2nd edn. The Guilford Press, New York/London
- Crawford S, Eklund RC (1994) Social physique anxiety, reasons for exercise, and attitudes toward exercise settings. *JSEP* 16(1):70–82. <https://doi.org/10.1123/jsep.16.1.70>
- Eklund RC, Mack D, Hart EA (1996) Factorial validity of the social physique anxiety scale for females. *J Sport Exerc Psychol* 18(3):281–295. <https://doi.org/10.1123/jsep.18.3.281>
- Hagger MS, Asci FH, Lindwall M, Hein V, Mulazimoglu-Balli O, Tarrant M, Ruiz YP, Sell V (2007) Cross-cultural validity and measurement invariance of the social physique anxiety scale in five European nations. *Scand J Med Sci Sports* 17(6):703–719. <https://doi.org/10.1111/j.1600-0838.2006.00615.x>

29. Maiano C, Morin AJ, Eklund RC, Monthuy-Blanc J, Garbarino JM, Stephan Y (2010) Construct validity of the social physique anxiety scale in a French adolescent sample. *J Pers Assess* 92(1):53–62. <https://doi.org/10.1080/00223890903381809>
30. Neves AB, Neves AN, Zanetti MC, Brandão MRF, Ferreira L (2018) Validação Psicométrica da social physique anxiety scale [Psychometric Validity of Social Physique Anxiety Scale in Brazil]. *RIPED* 13(2):193–202
31. Petrie TA, Diehl N, Rogers RL, Johnson CL (1996) The social physique anxiety scale: reliability and construct validity. *J Sport Exerc Psychol* 18(4):420–425. <https://doi.org/10.1123/jsep.18.4.420>
32. Teixeira PA, Silva WR, Marôco J, Campos JADB (2021) Social physique anxiety scale: a psychometric investigation of the factorial model in Brazilian adults. *Arch Clin Psychiatry* 48(2):129–134. <https://doi.org/10.15761/0101-60830000000294>
33. Eklund RC, Kelley B, Wilson P (1997) The social physique anxiety scale: men, women, and the effects of modifying item 2. *J Sport Exerc Psychol* 19(2):188–196. <https://doi.org/10.1123/jsep.19.2.188>
34. Martin KA, Rejeski WJ, Leary MR, McAuley E, Bane S (1997) Is the social physique anxiety scale really multidimensional? Conceptual and statistical arguments for a unidimensional model. *J Sport Exerc Psychol* 19(4):359–367. <https://doi.org/10.1123/jsep.19.4.359>
35. Isogai H, Brewer BW, Cornelius AE, Komiya S, Tokunaga M, Tokushima S (2001) Cross-cultural validation of the social physique anxiety scale. *Int J Sport Psychol* 32(1):76–87
36. Malheiro AS, Gouveia MJ (2001) Ansiedade física social e comportamentos alimentares de risco em contexto desportivo. *Aná Psicológica* 19(1):143–155
37. Saenz-Alvarez P, Sicilia A, Gonzalez-Cutre D, Ferriz R (2013) Psychometric properties of the social physique anxiety scale (SPAS-7) in Spanish adolescents. *Span J Psychol* 16:E86. <https://doi.org/10.1017/sjp.2013.86>
38. Campana ANNB (2011) Relações entre as dimensões da imagem corporal: um estudo em homens brasileiros [Relationships among body image dimensions: a study in Brazilian males]. University of Campinas, Institutional Repository
39. Hart EA (2003) [Avaliando a Imagem Corporal]. In: Tritschler K (ed) [Medida e avaliação em educação física e esportes de Barrow & Mcgee]. 5 edn. Manole, Barueri/SP, p 478
40. Souza V, Fernandes S (2009) Adaptação da Social Physique Anxiety Scale (SPAS) ao contexto brasileiro [Adaptation of Social Physique Anxiety Scale (SPAS) to Brazilian context]. *Cien Cogn* 14(3):16–23
41. Calmeiro LMDS, Simões MCR, Matos MGD, Gamito P (2012) Factorial validity and group invariance of the Portuguese short version of the social physique anxiety scale in adolescents. *J Child Adoles Psych* 15:199–213
42. ABEP (2021) Brazilian criteria. Brazilian market research association. <http://www.abep.org/criterio-brasil>. Accessed June 2022
43. WHO (2000) Obesity: preventing and managing the global epidemic. World Health Organization. https://www.who.int/nutrition/publications/obesity/WHO_TRS_894/en/
44. Marôco J (2021) Análise de equações estruturais [Structural equation analysis]. 3 edn. ReportNumber, Ltd, Pêro Pinheiro
45. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 25(24):3186–3191. <https://doi.org/10.1097/00007632-200012150-00014>
46. Swami V, Barron D (2019) Translation and validation of body image instruments: challenges, good practice guidelines, and reporting recommendations for test adaptation. *Body Image* 31:204–220. <https://doi.org/10.1016/j.bodyim.2018.08.014>
47. Kline RB (1998) Principles and practice of structural equation modeling. The Guilford Press, New York
48. Choi SW, Gibbons LE, Crane PK (2011) lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *J Stat Softw* 39(8):1–30. <https://doi.org/10.18637/jss.v039.i08>
49. Cook KF, Kallen MA, Amtmann D (2009) Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res* 18(4):447–460. <https://doi.org/10.1007/s11136-009-9464-4>
50. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Mod* 6(1):1–55. <https://doi.org/10.1080/10705519909540118>
51. Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 18(1):39–50. <https://doi.org/10.2307/3151312>
52. Hair JF, Black WC, Babin B, Anderson RE (2019) Multivariate data analysis, 8th edn. Cengage Learning EMEA, United Kingdom
53. Dunn TJ, Baguley T, Brunsden V (2014) From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol* 105(3):399–412. <https://doi.org/10.1111/bjop.12046>
54. Zumbo BD, Gadermann AM, Zeisser C (2007) Ordinal versions of coefficients alpha and theta for Likert rating scales. *JMASM* 6(1):21–29. <https://doi.org/10.22237/jmasm/1177992180>
55. Gadermann AM, Guhn M, Zumbo BD (2012) Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Pract Assess Res Eval* 17(3):1–13
56. Kim S, Lu Z, Cohen AS (2020) Reliability for tests with items having different numbers of ordered categories. *Appl Psychol Meas* 44(2):137–149. <https://doi.org/10.1177/0146621619835498>
57. Yang Y, Green SB (2015) Evaluation of structural equation modeling estimates of reliability for scales with ordered categorical items. *Methodology* 11(1):23–34. <https://doi.org/10.1027/1614-2241/a000087>
58. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 6(4):284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
59. Rosseel Y (2012) Lavaan: an R package for structural equation modeling and more. Version 0.5–12 (BETA). *J Stat Softw* 48(2):1–36
60. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y, Miller P, Quick C, Garnier-Villareal M, Selig J, Boulton A, Preacher K, Coffman D, Rhemtulla M, Robitzsch A, Enders C, Arslan R, Clinton B, Panko P, Merkle E, Chesnut S, J. B, Rights JD, Longo Y, Mansolf M (2019) semTools: Useful tools for structural equation modeling. <https://cran.r-project.org/web/packages/semTools/index.html>. Accessed June 2021
61. Genz A, Kenkel B, Azzalini A (2015) Vectorized bivariate normal CDF. <https://github.com/brentonk/pbivnorm>. Accessed Oct 2021
62. Revelle W (2019) psych: Procedures for psychological, psychometric, and personality research. R package version 1.9.12. <https://CRAN.R-project.org/package=psych>. Accessed June 2021
63. McAuley E, Burman G (1993) The social physique anxiety scale: construct validity in adolescent females. *Med Sci Sports Exerc* 25(9):1049–1053. <https://doi.org/10.1080/00223890903381809>
64. Eklund RC, Crawford S (1994) Active women, social physique anxiety, and exercise. *J Sport Exerc Psychol* 16(4):431–448. <https://doi.org/10.1123/jsep.16.4.431>
65. Kam CCS, Fan X (2020) Investigating response heterogeneity in the context of positively and negatively worded items by using

factor mixture modeling. *Organ Res Methods* 22(2):322–341. <https://doi.org/10.1177/1094428118790371>

66. Kam CCS (2018) Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociol Method Res* 47(3):574–597. <https://doi.org/10.1177/0049124115626177>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Wanderson Roberto da Silva^{1,2}  · Giovanna Soler Donofre¹  · Angela Nogueira Neves³  · João Marôco⁴  ·
Patrícia Angélica Teixeira¹  · Juliana Alvares Duarte Bonini Campos¹ 

Wanderson Roberto da Silva
wandersonroberto22@gmail.com

Giovanna Soler Donofre
giovannadonofre@gmail.com

Angela Nogueira Neves
angela.esefex@yahoo.com.br; patricia.angelica@unesp.br

João Marôco
jpmaroco@ispa.pt

(UNESP), Rodovia Araraquara-Jaú, km 01, Araraquara,
SP 14800-903, Brazil

² Graduate Program in Nutrition and Longevity, Federal
University of Alfenas (UNIFAL-MG), Alfenas, MG, Brazil

³ Physical Education School of Brazilian Army, Rio de Janeiro,
RJ, Brazil

⁴ William James Center for Research (WJCR), ISPA – Instituto
Universitário, Lisbon, Portugal

¹ Graduate Program in Food, Nutrition, and Food
Engineering, Department of Biological Sciences, School
of Pharmaceutical Sciences, São Paulo State University