Rosenberg Self-Esteem Scale: Method Effect and Gender Invariance

Tiago Jessé Souza de Lima[1]
Luana Elayne Cunha de Souza[1]

[1]Universidade de Fortaleza, Fortaleza, CE

Abstract
This study aimed to evaluate the dimensionality of the Rosenberg Self-Esteem Scale (RSES), by testing the adjustment of eight factorial models: a one-factor and two-factor model and six single-factor models controlling for the method effect associated with the wording of negative and positive items, through the correlated traits-correlated uniqueness (CTCU) and correlated traits-correlated methods (CTCM) approaches. We also tested measurement invariance across gender. A total of 689 participants took part in the study, with ages between 18 and 70 years (*M* = 25.5; *SD* = 8.06), mainly females (77.1%), who answered the RSES and sociodemographic questions. The results showed that single-factor models controlling for the effect of negative items alone or positive and negative items together best fit the data. The results also indicated that the RSES is invariant across gender, presenting the same theoretical structure and psychological meaning for men and women.
Keywords: self-esteem; scale; validity; method effect; invariance

Escala de Autoestima de Rosenberg: Efeito do Método e Invariância de Gênero

Resumo
O presente estudo objetivou avaliar a dimensionalidade da Escala de Autoestima de Rosenberg (EAR), testando o ajuste de oito modelos fatoriais: unifatorial, bifatorial, e seis modelos unifatoriais controlando-se o efeito do método associado à redação dos itens negativos e positivos através das estratégias *correlated traits-correlated uniqueness* (CTCU) e *correlated traits-correlated methods* (CTCM). Ademais, avaliou-se se a EAR é invariante quanto ao gênero. Compuseram a amostra 689 participantes com idades variando entre 18 e 70 anos (*M* = 25,5; *SD* = 8,06), sendo a maioria do sexo feminino (77,1%), os quais responderam a EAR e perguntas sociodemográficas. Os resultados indicaram que modelos unifatoriais, controlando-se o efeito do método associado aos itens negativos ou aos itens negativos e positivos conjuntamente, são os mais ajustados aos dados. Observou-se também que a EAR é invariante em relação ao gênero, apresentando a mesma estrutura teórica e significado psicológico para homens e mulheres.
Palavras-chave: autoestima; escala; validade; efeito do método; invariância

Escala de Autoestima de Rosenberg: Efecto del Método e Equivalencia a Través de Género

Resumen
Este estudio objetivó evaluar la dimensionalidad de la Escala de Autoestima de Rosenberg (EAR), probando el ajuste de ocho modelos factoriales: unifactorial, bifactorial, y seis modelos unifactoriales controlando el efecto del método de redacción de los ítems negativos y positivos a través de las estrategias *correlated traits-correlated uniqueness* (CTCU) y *correlated traits-correlated methods* (CTCM). Además, se evaluó si la EAR es equivalente a través de género. La muestra se conformó por 689 personas, con edades variando entre 18 y 70 años (*M*=25,5, *DE*=8,06), la mayoría mujeres (77,1%), que respondieron a la EAR y las preguntas socio-demográficas. Los resultados indicaron que los modelos unifactoriales, controlando el efecto del método asociado a los ítems negativos o a los ítems negativos y positivos conjuntamente, son los más ajustados a los datos. Se observó también que la EAR es equivalente a través de género, presentando la misma estructura teórica y significado psicológico para hombres y mujeres.
Palabras clave: autoestima; escala; validez; efecto del método; invariancia

Self-esteem can be defined as the evaluation that individuals make of themselves, anchored on personal feelings and beliefs about their skills, intelligence, social relations and future expectations, expressed in a positive attitude (of approval) or a negative one (of depreciation), indicating to what measure individuals believe they are capable, relevant, successful and deserving (Rosenberg, 1965). According to Hutz and Zanon (2011), high self-esteem apparently benefits people, since when they feel good about themselves, they are better able to deal effectively with challenges and negative feedback and generally believe that others value and respect them. In contrast, individuals with low self-esteem perceive the world through a negative filter, and their general self-aversion extends to the perceptions of everything in their surroundings. Therefore, self-esteem is associated with important aspects during people's lives. Studies have shown that high levels of self-esteem, for example, are correlated with subjective well-being (Kong, Zhao & You, 2013), satisfaction with interpersonal relationships (Erol & Orth, 2016), and satisfaction with life (Moksnes & Espnes, 2013). On the

other hand, low self-esteem is correlated with, among other feelings, depression (Rieger, Gollner, Trautwin & Roberts, 2016) and suicidal ideation (Kleiman & Riskind, 2013).

Researchers have proposed a wide range of self-esteem measures, including self-reported scales (Robins, Hendin & Trzesniewski, 2001; Rosenberg, 1965), indirect measures, such as evaluation of the preference for names and initials (Gebauer, Riketta, Broemer & Maio, 2008), and implicit association tests (Falk, Heine, Takemura, Zhang & Hsu, 2015). Among these, one of the most used measures continues to be the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965). This scale is composed of ten items, five worded positively and five worded negatively, addressing overall self-esteem as a set of feelings and thoughts of individuals about their own value, competence and adequacy, reflecting a positive or negative attitude about themselves. According to Sbicigo, Bandeira and Dell'Aglio (2010), the RSES was proposed as a one-dimensional measure, in which self-esteem is classified in three levels: low, characterized as feelings of incompetence, inadequacy and inability to face life's challenges; medium, characterized by fluctuation between feelings of approval and rejection; and high, consisting of self-judgment of value, confidence and competence.

Many studies have been conducted to investigate the factorial structure of the RSES, including in Brazil. The results obtained, however, have not been uniform. Studies performed in different countries (Supple, Su, Plunkett, Peterson & Bush, 2012), especially adopting an exploratory factorial approach, have indicated that the RSES is better represented by a two-factor structure, where one factor assesses positive self-esteem (composed of five positively worded items) and the other evaluates negative self-esteem (composed of five negatively worded items). On the other hand, various studies have also indicated a single-factor solution for the RSES (DiStefano & Motl, 2009; Michaelides et al., 2016). Of the validation studies conducted in Brazil, two have proposed a bifactorial structure (Avancini, Assis, Santos & Oliveira, 2007; Meurer, Luft, Benedetti & Mazo, 2012; Sbicigo et al., 2010) while one indicated a unifactorial structure (Hutz & Zanon, 2011).

The authors that advocate a two-factor structure for the RSES argue that positive and negative self-esteem are different concepts, which have presented different patterns of association with other theoretically related constructs (Supple et al., 2012). For example, Owens (1994) observed that positive self-esteem was

significantly related to school grades but not related to depressive symptoms, while negative self-esteem was positively related to depressive symptoms and delinquent behaviors, but not with school grades. Supple et al. (2012), in a study of self-esteem among teenagers, also found different patterns of correlation, with positive and negative self-esteem being differently predicted by parental behavior and academic motivation.

This controversy over the factorial structure is reflected in the debate about whether the bifactorial structure indeed encompasses two theoretically distinct elements related to self-esteem or results from a type of method effect that occurs when including items that are positively and negatively worded in the same scale (DiStefano & Motl, 2009; Urbán, Szigeti, Kökönyei & Demetrovics, 2014). This method effect refers to the variance that arises from, among other sources, the positive or negative phrasing of the items, instead of the variance related to the construct of interest, resulting in a systematic variation that is undesirable and impairs the construct's measurement (Lindwall et al., 2012). This influence has been called the wording effect, and is a type of measurement method effect (Wu, 2008).

The use of positive and negative items in psychological measures is a common practice, seeking to avoid acquiescence error, i.e., the tendency of respondents to agree or disagree with all the items in general, regardless of their content (Podsakoff, MacKenzie & Podsakoff, 2012). However, the inclusion of positive and negative items can introduce systematic measurement errors that affect the analysis and interpretation of the data, besides the effects of the construct being assessed. This can result, for example, in the emergence of factors that separately group positive and negative items, even when the content of these items is congruent (DiStefano & Motl, 2006).

One of the explanations for the phenomenon of negative and positive items forming separate factors involves the difficulty or the way respondents cognitively process items written inversely, introducing variation in responses that is not associated with the construct under analysis (Wu, 2008). Therefore, the negative items function as cognitive "speed bumps" that require a more elaborate cognitive process instead of an automatic response at the moment of answering the item, thus resulting in variation due to the measurement effect (Podsakoff et al., 2012). Although this effect is primarily related to negatively worded items, some studies have indicated that positively worded items also can bias the responses, because that effect is

mainly related to the participants' interpretation of the item's content (Wu, 2008).

Some strategies have been used to evaluate the wording effect of the items on the factorial structure of scales, such as using confirmatory factor analysis (CFA), in which it is possible to control for the unique variance associated with negative items, positive items, or both together (DiStefano & Motl, 2006; Wu, 2008). Unlike exploratory factor analysis, CFA permits specifying possible method effects and comparing alternative factor models that specify these effects or not. Using CFA, two methodological strategies have been developed, derived from multitrait–multimethod models, to separate the variance derived from the construct from that derived from the method of wording the items (Lindwall, 2012). One of them is the correlated traits-correlated uniqueness (CTCU) model, in which the method is modeled by establishing covariances between the errors of the affected items. In contrast, the correlated traits-correlated methods (CTCM) model treats the response effect of the items as a latent variable that is incorporated in the analysis as a distinct factor, orthogonal to the factor of the construct to be evaluated.

According to Lindwall et al. (2012), both strategies allow comparing whether the models in which the method effect is specified better fit the data than models that do not include this effect. If the model presents better adjustment, the existence of the method effect can be inferred. Moreover, the method effect in CTCM models can be quantified and correlated with other variables, such as gender or socioeconomic class, which cannot be done by using CTCU models. On the other hand, CTCM models can cause problems in performing the analysis, such as under-identification and negative variance estimates, among others (Marsh, Scalas & Nagengast, 2010). For these reasons, we believe that both approaches should be employed, even though Lance, Noble and Scullen (2002) stated that when convergent and acceptable factor solutions are obtained, the use of a CTCM model is preferable.

Various studies aimed at assessing the method effect on the factor structure of the RSES, utilizing both CTCU and CTCM models, have indicated that the unifactorial structure controlling for the method effect of the wording of the negative items presents the best fit to the date than a bifactorial structure or unifactorial structure without that control (DiStefano & Mottl, 2009; Lindwall et al., 2012; Marsh et al., 2010; Michaelides et al., 2016; DiStefano & Motl, 2009;

Tomás, Oliver, Galiana, Sancho & Lila, 2013; Urbán et al., 2014; Wu, Zuo, Wen & Yan, 2017). On the other hand, some studies have also indicated that the models that control for both the method effects associated with negatively and positively worded items fit the data well (Lindwall et al., 2012; Wu, 2008).

According to some researchers, the method effect related to the wording of items can vary in function of the population or be more preponderant in certain groups than in others, such as between men and women (DiStefano & Motl, 2009; Tomás et al., 2013). Differences between men and women have been observed in relation to global self-esteem, with men generally presenting higher levels (Bleidorn et al., 2016; Tomás et al., 2013). According to DiStefano and Motl (2006), the differences in average RSES scores between men and women might be associated with the different responses to negatively worded items. In other words, people of different gender can provide stronger or weaker responses to negatively worded items, hence causing differences in the average self-esteem scores. Studies examining the method effect associated with both negative and positive items (DiStefano & Motl, 2009; Lindwall et al., 2012) have not observed differences in the adjustment of the models to the data when testing the invariance in relation to gender, although they have observed differences in the average raw scores of men and women. On the other hand, Michaelides et al. (2016), when testing the gender invariance in a model controlling for the method effect of positive and negative items, observed configurational and metric invariance but not scalar invariance, making it impossible to compare the latent means.

In short, there is no consensus regarding the factor structure of the RSES, with studies adopting an exploratory approach indicating a two-factor structure, while others that have controlled for the wording effect of items have indicated a single-factor structure is most suitable. The studies of the RSES conducted in Brazil have not considered the method effect associated with the negative and positive items when testing the factor structure. Therefore, this study sought to gather evidence about the factor structure of the RSES in Brazil by testing the fit of bifactorial and unifactorial structures, and of unifactorial solutions while controlling for the method effect associated with the items through the CTCU and CTCM strategies. We also evaluated whether the factor structure of the RSES is invariant between men and women.

Method

Participants

A total of 689 people took part in the study, consisting of university students and people from the general population, all residing in the city of Fortaleza, Ceará. The participants' age ranged from 18 to 70 years, with average of 25.5 years ($SD = 8.06$), the majority being women (77.1%). The sample was chosen by convenience, so it was non-probabilistic.

Instruments

The participants received a booklet containing the Rosenberg Self-Esteem Scale (RSES) and sociodemographic questions. The RSES (Rosenberg, 1965), translated into Portuguese and adapted for Brazil by Hutz and Zanon (2011), is composed of ten items with statements related to feelings of self-esteem and self-acceptance to assess global self-esteem, containing five items related to a positive personal view (e.g., *I feel that I have various good qualities; On the whole, I am satisfied with myself*) and five items with a negative personal view (e.g., *I feel I do not have much to be proud of; On the whole, I am inclined to believe that I am a failure*). The participants were asked to state how much they agree or disagree, with each statement, on a seven-point scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Procedure

The participants were told that taking part in the survey was voluntary and that the information collected would be treated as confidential, only used for academic purposes. The data were gathered in various places, such as classrooms and public venues. All the items were answered individually. The research procedures were in line with the ethical rules on research with human beings, as specified in Resolution 510/2016 from the National Health Council. According to article 1 of the resolution, this study did not require registration in the CEP/CONEP system (Research Ethics Committee/National Research Ethics Commission) because of data aggregation without the possibility of individual identification.

Data analysis

We initially calculated descriptive statistics for the RSES items and tallied the overall score, followed by computing the asymmetry and kurtosis values and constructing the Pearson pairwise correlation matrix between the items and between them and the overall score. We used Chronbach's alpha to evaluate the internal consistency of the responses, followed by application of AMOS 18 to evaluate the factorial validity of the RSES. To compare the alternative factor models, we performed multiple confirmatory factor analysis, considering the covariance matrix and employing the maximum likelihood (ML) method. Only 0.5% of the observations had missing data, which were filled with the mean of the responses obtained in the corresponding item.

To ascertain the fit of the proposed model and compare it with alternative models, we used the following indicators: chi-square (χ^2), normalized chi-square (χ^2/df), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), Akaike information criterion (AIC), and the expected cross-validation index (ECVI). These indices have often been used in previous studies, and although each one has merits and limitations, when used together they provide strong indication of the fit of models to the data (Byrne, 2010). For the normalized chi-square metric (χ^2/df), values smaller than 5 indicate adequate adjustment of the model, although values smaller than 3 are desirable. Values of the CFI and TLI greater than 0.90 indicate acceptable fit, while values higher than 0.95 indicate good fit. For the RMSEA, values up to 0.08 denote acceptable fit, while values up to 0.06 indicate good fit. Finally, in relation to the indices for comparison between the models (AIC and ECVI), lower values indicate the model has better fit.

We tested the fit of eight factorial solutions for the RSES, presented in Figure 1. Model 1 is a single-factor solution while Model 2 represents a structure with two correlated factors, when dividing the items between positive self-esteem and negative self-esteem. Models 3 and 4 represent a single-factor structure of the RSES with the measurement errors correlated between positive items (Model 3) and between negative items (Model 4). Model 5 denotes a unifactorial structure with measurement errors correlated simultaneously between positive items and between negative items. Models 6 and 7 represent a unifactorial structure in which the method effect is controlled via a latent factor associated with the positive items (Model 6) and negative items (Model 7). Finally, Model 8 denotes a unifactorial structure with correlated latent factors to control for the method effect in the positive and negative items simultaneously.

Finally, to assess whether the factor structure of the RSES was equivalent between men and women,
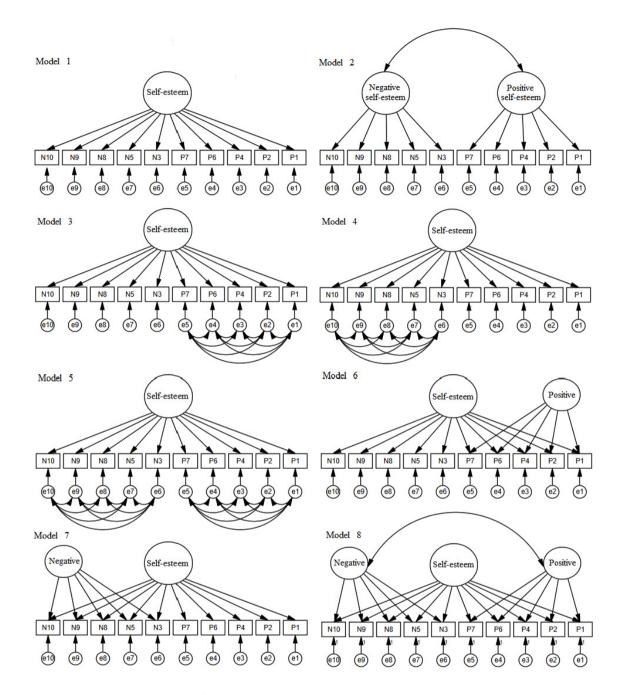
Figure 1. Schematic representations of the models tested.

we applied multigroup confirmatory factor analysis (MCFA). For this we tested four hierarchical invariance models (Byrne, 2010): configurational, metric, scalar and residual. The configurational invariance assesses whether the factor structure of the instrument has good fit to the data in both groups. The metric invariance evaluates whether the regression weights of the items are statistically equivalent between the groups analyzed. The scalar invariance measures whether the observed scores

are related with the latent scores, i.e., if the individuals that obtained the same score for the latent variable also obtained the same score for the observed variables, irrespective of the group. Finally, the residual invariance of the items evaluates to what extent the measurement errors are equal for different groups.

When the models tested are nested, the invariance between the models can be ascertained by the difference of the chi-square value and degrees of

freedom $\Delta\chi^2(\Delta df)$ of the model evaluated in relation to the previous (less restricted) one. If the chi-square value resulting from this difference is not significant (considering the reference values in relation to the degrees of freedom resulting from the difference between the two models), this indicates the existence of invariance in relation to the model tested (Byrne, 2010). However, the use of only $\Delta\chi^2(\Delta df)$ has been criticized, with the recommendation to use other metrics, such as the difference between the CFI of the models (ΔCFI). In this respect, CFI differences below 0.1 provide evidence of invariance between the models tested (Byrne, 2010).

Results

Table 1 presents the fit indices for the eight models tested. Initially, Model 1 (simple unifactorial) did not adequately fit the data, while Model 2 (bifactorial, based on positive and negative items) presented a marginal fit to the data. The correlation between the two factors was moderate ($r = -0.25$; $p < 0.001$), indicating an orthogonal structure. The models including the method effect (3 and 4: controlling for the method using correlations between the errors; and 6, 7 and 8: controlling for the method effect through latent factors) presented generally better adjustment to the data than Models 1 and 2. However, the models that controlled for the method effect of the negative items (Models 4 and 7) had better fit than those that controlled for the method effect associated with the positive items (Models 3 and 6). Furthermore, Models 3 and 6 presented slightly better fit than Model 2, indicating that controlling only for the method effect associated with the positive items did not significantly improve the fit to the data.

Finally, in relation to the models that controlled simultaneously (Models 5 and 8) for the effect of the positive and negative items, Model 5 was an unidentified model, as also observed in other studies (Michaelidis et al., 2016; Vasconcelos-Raposo, Fernandes, Teixeira & Bertelli, 2012; Supple et al., 2012; Wu et al., 2017). Due to the absence of non-significant correlations between the errors observed in Models 3 and 4, it was not possible to respecify Model 5, as proposed by Tomás and Oliver (1999), who recommended setting the non-significant correlations of unidentified models at zero. Therefore, Model 5 was excluded from further analysis. In turn, Model 8 presented adjustment indices considered good in the literature, although these indices were similar to the Model 4 indices, which had the best fit

among the eight models tested. The value of Cronbach's alpha for global self-esteem was 0.82.

Since Model 4 best fit the data, we decided to evaluate whether this model was invariant between men and women. The fit indices for the hierarchical invariance models are presented in Table 1. From the most restrictive model (configurational) to the least restrictive one (residual), the fit indices were consistent and within the levels recommended in the literature, indicating the suitability to conduct a factorial invariance test for gender. The configurational model presented good adjustment indices, indicating that the one-factor structure, with control for the effect of the negative items through the correlations between the errors, was adequate for men and women. The metric invariance also was supported by the data, since the differences of the chi-square values were not significant and the CFI continued having the same value.

In the scalar variance model, the chi-square value increased significantly in comparison with the previous one, $\Delta\chi^2 = 21.4$ and $\Delta df = 10$, $p < 0.05$. On the other hand, the change in the CFI was smaller than 0.01. Based on these inconsistent results, even considering that χ^2 is a sensitive and often biased indicator, we decided to conduct individual analyses of the intercepts of each item to identify which was degrading the model's fit. We observed that by allowing the intercept of item 3 to vary freely among the groups, the chi-square value diminished considerably, ceasing to be significantly different in relation to the metric variance model ($\Delta\chi^2 = 16;30$ and $\Delta df = 9$, $p = 0.061$). Furthermore, the ΔCFI value less than 0.01 indicated that RSES had partial scalar invariance according to our data.

Due to this partial scalar invariance between men and women, we compared the latent means, setting the average for women to zero. The latent mean estimated for men was not significantly higher than that for women (average difference = 0.04; standard error = 0.07; $Z = 0.48$; $p = 0.63$). For illustrative purposes, the raw scores observed for men ($M = 5.27$; $SD = 1.07$) and women ($M = 5.08$; $SD = 1.08$) presented a marginally significant difference, $t(685) = 1.90$, $p = 0.058$, Cohen's $d = 0.18$. Finally, to test for residual invariance (errors of the items were equal for the different groups), we restricted the correlations between the negative items of Model 4. There was a significant increase in the chi-square value ($\Delta\chi^2 = 47.8$ and $\Delta df = 10$, $p < 0.01$) and an increase of the CFI to almost 0.01, indicating that the invariance of the residuals between men and women was not supported by the data.

Table 1.
Fit indices of the models and for the measurement invariance test

	χ^2	gl	χ^2/df	CFI	TLI	RMSEA [IC 90%]	ECVI	AIC
Model 1	1156.5	35	33.0	0.629	0.523	0.216 0.205 - 0.227	1.739	1196.5
Model 2	261.9	34	7.70	0.925	0.900	0.099 0.088 - 0.110	0.442	303.94
Model 3	177.8	25	7.112	0.949	0.909	0.094 0.081 - 0.108	0.346	237.81
Model 4	115.96	25	4.638	0.970	0.946	0.073 0.060 - 0.086	0.256	167.43
Model 5				Unidentified				
Model 6	263.65	30	8.788	0.923	0.884	0.106 0.095 - 0.118	0.456	313.65
Model 7	230.3	30	7.677	0.934	0.901	0.099 0.087 - 0.111	0.407	280.3
Model 8	115.56	24	4.82	0.971	0.943	0.074 0.061 - 0.088	0.258	177.56

	χ^2	df	$\Delta\chi^2(\Delta df)$	CFI	ΔCFI	TLI	RMSEA [IC 90%]
Configurational	166.40	50	-	0.962	-	0.932	0.058 0.049 - 0.068
Metric	175.50	59	9.1(9)	0.962	0.0	0.943	0.054 0.045 – 0.063
Scalar	196.90	69	21.4(10)[*]	0.959	0.003	0.946	0.052 0.044 - 0.061
Partial scalar	191.80	68	16.3(9)	0.960	0.002	0.947	0.052 0.043 – 0.060
Residual	239.60	89	47.8(21)[**]	0.951	0.009	0.951	0.058 0.049 - 0.068

Discussion

We investigated the factor structure of the Brazilian version of the RSES by testing the fit of two-factor and one-factor structures, as well as the fit of single-factor structures in which we controlled for the method effect associated with negative and positive items, through the correlated traits-correlated uniqueness (CTCU) and correlated traits-correlated methods (CTCM) strategies. We also evaluated whether the factor structure of the RSES presents invariance between men and women. The results indicated that the one-factor structure with correlated measurement errors for the negative items (Model 4; CTCU) presented better fit than the other models. Nevertheless, Model 8 (CTCM),

unifactorial with two latent factors to control for the method effect of the positive and negative items, also presented good fit to the data. The analyses also indicated that the model that best fit the data (Model 4) was equivalent between men and women, with only invariance of the residuals not being observed.

With respect to the factor structure, our results are in line with previous studies that have indicated better fit of single-factor solutions for the RSES that include control for the method effect associated mainly with negative items, but also with positive ones (Lindwall et al., 2012; Marsh et al., 2010; Michaelides et al., 2016). In this sense, the two-factor structure of the RSES, commonly observed in Brazilian studies, can be understood as an artifact derived from the wording of the items

in positive and negative form, not as two distinct self-esteem factors. We therefore recommend that the RSES be considered a general measure of self-esteem, in a more parsimonious representation of this construct, as originally proposed by Rosenberg (1965).

The method effect associated with the wording of the items in the structure of the RSES is not a specific problem of this scale. It has been observed in other scales that use items phrased in positive and negative terms with similar number, such as the General Health Questionnaire and the Need for Cognition Scale. Some studies have indicated that these scales are better interpreted through a single-factor structure with control for the effect of the negative items, instead of the two-factor structure commonly seen in studies that adopt exploratory factor techniques (Gouveia, Lima, Gouveia, Freires & Barbosa, 2012; Molina, Rodrigo, Losilla & Vives, 2014; Zhang, Noor & Savalei, 2016).

Regarding the studies conducted in Brazil about the factor structure of the RSES, most have pointed to positive and negative self-esteem as two distinct factors (Avancini et al., 2007; Meurer et al., 2012; Sbicigo et al., 2010). Only one study, conducted by Hutz and Zanon (2011), indicated a single-factor structure of the RSES. However, those authors, in analyzing the data via exploratory factor analysis, used criteria without strong robustness (scree plot) to justify retaining only one factor, disregarding more reliable techniques like parallel analysis (Tabachnick & Fidell, 2012), and contrary evidence from their own data (e.g., the values higher than 1 obtained by two factors). In this respect, our study provides adequate empirical evidence to treat the Brazilian version of the RSES as a single-factor measure when controlling for the wording effect of the negative and positive items.

Furthermore, the results described here support the use of correlated traits-correlated uniqueness (CTCU) and correlated traits-correlated methods (CTCM) models to study the wording effect of the items (Marsh et al., 2010). By using these two types of methods to examine the wording effect, we aimed to establish whether the models that include correlated measurement errors or latent factors related to the method effect better fit the data in comparison with models that do not include these methods (Lindwall et al., 2012). The results indicated that Model 4, in which the errors of the negative items were correlated (CTCU), presented slightly better fit indices than those of Model 8, which had two latent factors predicting the effect of negative and positive items (CTCM). Other

studies involving the RSES have observed that both the CTCM model, similar to Model 8 (Michaelides et al., 2016; Salerno, Ingoglia & Coco, 2017; Wu et al., 2017), and the CTCU model, similar to Model 4 (DiStefano & Motl, 2006), have the best fit, although the indices of the two are close.

Some studies, however, have indicated that CTCU models in general present better fit than CTCM models (Lindwall et al., 2006; Molina et al., 2014), as was observed in this study. A possible explanation is that in allowing the residuals of the items to be correlated, the CTCU models might take into consideration not only the variance related to the wording effect of the items, but also unknown factors (Wu et al., 2017). In counterpart, the CTCM models specify the wording effect of the items explicitly, which allows examining the empirical relations of this effect with relevant measures of attitude or personality (Lindwall et al., 2012). However, a worse fit can be observed when other unknown factors are present, but not specified in the model (Molina et al., 2014). According to Wu et al. (2017), CTCU models are more stable than CTCM models and less susceptible to generating non-convergent solutions in the CFA. However, our Model 5, which is a CTCU model, did not converge to a viable solution because it was unidentified. Similar results have been observed in other studies (Michaelidis et al., 2016; Vasconcelos-Raposo et al., 2012; Supple et al., 2012; Wu et al, 2017), so additional research is necessary to identify the causes behind the problems of non-convergence found in the models.

It should also be noted that both the wording of the negative items (Models 4 and 8) and positive ones (Model 8) is related to the method effect, although the negative items had a more significant impact, given the better fit of Model 4 compared to Model 2. In fact, both the negative and positive items can introduce variation in the measurement that is not associated with the construct being studied (Wu, 2008). The possible sources of the method effect associated with the negative items include careless responses and acquiescence bias (Wu, 2008). The negative items require the respondent engage in a more elaborate cognitive process at the moment of responding to the item, resulting in additional variance in the measurement process (Podsakoff et al., 2012; Weijters, Baumgartner & Schillewaert, 2013). This effect is also related to the positive items, because the presence of the method effect is related to the way the participants interpret the content of the items, which in this case can also be influenced by the

modesty bias, commonly observed in countries with collectivist orientation (Wu, 2008).

The invariance analyses indicated that the model that best fit the data (Model 4) was equivalent between men and women at the configurational, metric and partial scalar levels, when the method effects associated with negative items were controlled. This is the first evidence of invariance of the RSES in a Brazilian sample. The equivalence between men and women indicates that the measurement of self-esteem with the Brazilian version of the RSES has the same underlying theoretical structure and psychological significance between men and women, corroborating other findings in the international literature (DiStefano & Motl, 2009; Vasconcelos-Raposo et al., 2012; Michaelidis et al., 2016; Tomás et al., 2013). Nevertheless, when evaluating the scalar invariance, we observed that item 3 (*On the whole, I am inclined to believe that I am a failure*) was the only one that operated differently between men and women. Similar results were obtained by Vasconcelos-Raposo et al. (2012) in a sample of Portuguese men and women. According to the authors, when considering the content of this item, it is possible to suggest that the differences in its scoring between men and women can be the consequence of distinct social and cultural practices underlying gender roles.

Due to the partial scalar invariance between men and women, we compared the latent means and observed that the differences between men and women were not significant. However, the raw scores indicated a marginally significant difference. These results differ from others reported in the literature, in which comparisons of the means of the latent factors have indicated that men have significantly higher levels of self-esteem than women (DiStefano & Motl, 2009; Lindwall et al., 2012; Michaelidis et al., 2016; Tomás et al., 2013). Nevertheless, some studies conducted in Brazil have not indicated significant differences between men and women (Sbicigo et al., 2010; Hutz, Zanon & Vazquez, 2014). This disparity of findings indicates that the differences in self-esteem observed in other studies should not be interpreted as something inherent to gender, but instead as due to contextual factors.

Although the results presented here provide satisfactory evidence for the method effect of the negative and positive items in the factor structure of the RSES and about the gender invariance of this measure, some limitations should be mentioned. Previous studies have indicated that wording effect is associated with the characteristics of the respondents, not with a statistical

property of the scale (Wu, 2008). In this sense, the present study has a limitation since we did not evaluate possible dispositional and situational determinants related to the wording effect of the items. For example, certain personality traits are apparently associated with the method effect, such as conscientiousness and neuroticism, which predict, respectively, high scores on negative items and low scores on positive ones (Michaelides et al., 2016). Differences in relation to the wording effect of items were also observed among respondents from collectivist and individualist societies (Wu, 2008).

Another limitation is related to our sample, mostly composed of university students. As proposed in previous studies (DiStefano & Motl, 2006; Wu et al., 2016), the method effects can vary among populations and be more important for certain groups than for others. For example, Lindwall et al. (2016) indicated that negatively worded items can be problematic for groups with relatively low verbal skills, but not for those with higher skills. The absence of evaluation of these dispositional factors and the high variability in the sample hinder comprehension of possible individual differences related to the method effect. In this respect, future studies could evaluate the relationship of dispositional, contextual and sociodemographic factors in the responses to negative and positive items in more diverse Brazilian samples.

Furthermore, the methods used here (CTCU and CTCM), although controlling for the wording effect of the items, do not exclude this effect in the evaluation of the fit of the factor structure. In other words, there is a tendency for methods based on CFA to present a medium to poor fit to the factor structure proposed in which the effect of the items wording is controlled (Maydeu-Olivares & Steenkamp, 2018). Future studies could adopt other modeling strategies, such as random intercept, which permits the exclusion of the variance related to the wording effect of the items, besides enabling controlling for other method effects (such as acquiescence bias), enabling maintaining the proposed factor structure and improving the fit indices.

Final Considerations

In summary, the results obtained indicate the adequacy of a single-factor model for the Rosenberg Self-Esteem Scale when controlling for the effect of the wording of the negative items, suggesting that the RSES is a measure of global self-esteem. The results

also indicate that this scale is gender invariant, i.e., the RSES presents the same underlying theoretical structure and the same psychological meaning for women and men.

References

Avancini, J., Assis, S., Santos, N. & Oliveira, R. V. (2007). Adaptação Transcultural de Escala de Auto-Estima para Adolescentes. *Psicologia: Reflexão e Crítica, 20*, 397-405. doi: 10.1590/S0102-79722007000300007

Bleidorn, W., Arslan, R. C., Denissen, J. J., Rentfrow, P. J., Gebauer, J. E., Potter, J. & Gosling, S. D. (2016). Age and gender differences in self-esteem-A cross-cultural window. *Journal of Personality and Social Psychology, 111*, 396-410. doi: 10.1037/pspp0000078

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic concepts, applications and programming.* New York: Routledge.

Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319. doi: 10.1037/1040-3590.7.3.309

DiStefano, C. & Motl, R. (2006). Further Investigating Method Effects Associated With Negatively Worded Items on Self-Report Surveys. *Structural Equation Modeling: A Multidisciplinary Journal, 13*, 440-464. doi: 10.1207/s15328007sem1303_6

DiStefano, C. & Motl, R. (2009). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences, 46*, 309-313. doi: 10.1016/j.paid.2008.10.020

Erol, R. Y. & Orth, U. (2016). Self-Esteem and the Quality of Romantic Relationships. *European Psychologist, 21*, 274-283. doi: 10.1027/1016-9040/a000259

Falk, C. F., Heine, S. J., Takemura, K., Zhang, C. & Hsu, C. W. (2015). Are Implicit Self-Esteem Measures Valid for Assessing Individual and Cultural Differences? *Journal of Personality, 83*, 56-68. doi: 10.1111/jopy.12082

Garson, G. D. (2012). *Structural Equation Modeling.* Asheboro: Statistical Publishing Associates.

Gebauer, J., Riketta, M., Broemer, P. & Maio, G. (2008). "How much do you like your name?" An implicit measure of global self-esteem. *Journal of Experimental Social Psychology, 44*, 1346-1354. doi: 10.1016/j.jesp.2008.03.016

Gouveia, V. V., Lima, T. J. S., Gouveia, R:, Freires, L. & Barbosa, L. (2012). Questionário de Saúde Geral (QSG-12): o efeito de itens negativos em sua estrutura fatorial. *Cadernos de Saúde Pública, 28*, 375-384. doi: 10.1590/S0102-311X2012000200016

Hutz, C. S. & Zanon, C. (2011). Revisão da adaptação, validação e normalização da Escala de Autoestima de Rosenberg. *Avaliação Psicológica, 10*, 41-49. Recuperado de: http://pepsic.bvsalud.org/pdf/avp/v10n1/v10n1a05.pdf

Hutz, C. S., Zanon, C. & Vazquez, A. C. (2014). Escala de Autoestima de Rosenberg. Em C. Hutz, *Avaliação em Psicologia Positiva* (pp. 85-94). Porto Alegre: Artmed.

Kleiman, E. M. & Riskind, J. H. (2013). Utilized social support and self-esteem mediate the relationship between perceived social support and suicide ideation. A test of a multiple mediator model. *Crisis, 34*, 42-49. doi: 10.1027/0227-5910/a000159.

Kong, F., Zhao, J. & You, X. (2013). Self-Esteem as Mediator and Moderator of the Relationship Between Social Support and Subjective Well--Being Among Chinese University Students. *Social Indicators Research, 112*, 151-161. doi: 10.1007/s11205-012-0044-6

Lance, C. E., Noble, C. L. & Scullen, S. E. (2002). A critique of the correlated trait– correlated method and correlated uniqueness models of multitrait–multimethod data. *Psychological Methods, 7*, 228–244. doi: 10.1037/1082-989X.7.2.228

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J. & Thøgersen-Ntoumani, C. (2012). Method Effects: The Problem With Negatively Versus Positively Keyed Items. *Journal of Personality Assessment, 94*, 196-204. doi: 10.1080/00223891.2011.64593

Marsh, H., Scalas, L. F. & Nagengast, B. (2010). Longitudinal Tests of Competing Factor Structures for the Rosenberg Self-Esteem Scale: Traits, Ephemeral Artifacts, and Stable Response Styles. *Psychological Assessment, 22*, 366-381. doi: 10.1037/a0019225

Maydeu-Olivares, A. & Steenkamp, J. E. M. (2018, 31 August). An integrated procedure to control for common method

variance in survey data using random intercept factor analysis models. Recuperado de https://www.academia.edu/36641946/An_integrated_procedure_to_control_for_common_method_variance_in_survey_data_using_random_intercept_factor_analysis_models?auto=download

Meurer, S. T, Luft, C. B., Benedetti, T. R. & Mazo, G. Z. (2012). Validade de construto e consistência interna da escala de autoestima de Rosenberg para uma população de idosos brasileiros praticantes de atividades físicas. *Motricidade*, *8*, 5-15. doi: 10.6063/motricidade.8(4).1548.

Michaelides, M. P., Zenger, M., Koutsogiorgi, C., Brähler, E., Stöbel-Richter, Y. & Berth, H. (2016). Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Personality and Individual Differences*, *97*, 13-18. doi: 10.1016/j.paid.2016.03.011

Moksnes, U. K. & Espnes, G. A. (2013). Self-esteem and life satisfaction in adolescents-gender and age as potential moderators. *Quality of Life Research*, *22*, 2921-2928. doi: 10.1007/s11136-013-0427-4

Molina, J. G., Rodrigo, M. F., Losilla, J. M. & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological Assessment*, *26*, 1031-1037. doi: 10.1037/a0036472.

Owens, T. J. (1994). Two dimensions of self-esteem: Reciprocal effects of positive self-worth and self-deprecation on adolescent problems. *American Sociological Review, 59*, 391–407. doi: 10.2307/2095940

Podsakoff, P. M., MacKenzie, S. B. & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569. doi: 10.1146/annurev-psych-120710-100452

Rieger, S., Gollner, R., Trautwin, U. & Roberts, B. W. (2016). Low self-esteem prospectively predicts depression in the transition to young adulthood: A replication of Orth, Robins, and Roberts (2008). *Journal of Personality and Social Psychology*, *110*, 16-22. doi: 10.1037/pspp0000037

Robins, R. W., Hendin, H. M. & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*, 151-161. doi: 10.1177/0146167201272002

Rosenberg, M. (1965). *Society and the Adolescent Self-image*. Princeton: Princeton University Press.

Salerno, L., Ingoglia, S. & Coco, G. L. (2017). Competing factor structures of the Rosenberg Self-Esteem Scale (RSES) and its measurement invariance across clinical and non-clinical samples. *Personality and Individual Differences*, *113*, 13-19. doi: 10.1016/j.paid.2017.02.063

Sbicigo, J. B., Bandeira, D. R. & Dell'Aglio, D. D. (2010). Escala de Self-esteem de Rosenberg (EAR): validade fatorial e consistência interna. *Psico-USF*, *15*, 395-403. doi: 10.1590/S1413-82712010000300012

Supple, A. J., Su, J., Plunkett, S. W., Peterson, G. W. & Bush, K. R. (2012). Factor structure of the Rosenberg Self-Esteem Scale. *Journal of Cross-Cultural Psychology*, *44*, 748-764. doi: 10.1016/j.paid.2012.08.009

Tabachnick, B. & Fidell, L. (2012). *Using Multivariate Statistics*. Harlow: Pearson Education Limited

Tomás, J. M. & Oliver, A. (1999). Rosenberg's Self-Esteem scale: Two factors or method effects. *Structural Equation Modeling*, *6*, 84–98. doi: 10.1080/10705519909540120

Tomás, J. M., Oliver, A., Galiana, L., Sancho, P. & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem Scales. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 299–313. doi: 10.1080/10705511.2013.769394

Urbán, R., Szigeti, R., Kökönyei, G. & Demetrovics, Z. (2014). Global self-esteem and method effects: competing factor structures, longitudinal invariance, and response styles in adolescents. *Behavior Research Methods*, *46*, 488–498. doi: 10.3758/s13428-013-0391-5.

Vasconcelos-Raposo, J., Fernandes, H. M., Teixeira, C. M. & Bertelli, R. (2012). Factorial Validity and Invariance of the Rosenberg Self-Esteem Scale Among Portuguese Youngsters. *Social Indicators Research*, *105*, 483-498. doi: 10.1007/s11205-011-9782-0

Weijters, B., Baumgartner, H. & Schillewaert, N. (2013). Reversed Item Bias: An Integrative Model. *Psychological Methods*, *18*, 320-334. doi: 10.1037/a0032121

Wu, C. H. (2008). An Examination of the Wording Effect in the Rosenberg Self-Esteem Scale among Culturally Chinese People. *The Journal of Social Psychology*, *148*, 535-552. doi: 10.3200/SOCP.148.5.535-552

Wu, Y., Zuo, B., Wen, F. & Yan, L. (2017). Rosenberg Self-Esteem Scale: Method Effects, Factorial Structure and Scale Invariance Across Migrant Child and Urban Child Populations in China. *Journal of Personality Assessment*, *99*, 83-93. doi: 10.1080/00223891.2016.1217420

Sobre os autores:

Tiago Jessé Souza de Lima. PhD in Social Psychology from Universidade Federal da Paraíba (UFPB), Brazil, with an internship at ICS-UL, Portugal. Professor at Universidade de Fortaleza, teaching in undergraduate program in Psychology and in the professional master in business administration. Researches interests focus on the areas of Social Psychology (intergroup relations, social exclusion and psychosocial consequences; human values), Psychological Assessment (construction and adaptation of psychological and psychosocial variables evaluation measures), Organizational and Work Psychology (workers' health and well-being).
ORCID: https://orcid.org/0000-0001-8840-4285
E-mail: tiago.souzalima@hotmail.com

Luana Elayne Cunha de Souza. PhD in Social Psychology from Universidade Federal da Paraíba (UFPB), Brazil, with an internship at ICS-UL, Portugal. Professor at Universidade de Fortaleza (UNIFOR), teaching in undergraduate and graduate program in Psychology. Coordinates the Laboratório de Estudos dos Processo de Exclusão Social (LEPES). Researches are concentrated in the areas of Social Psychology (social exclusion; legitimation of social inequalities; human values) and Psychological Assessment.
ORCID: https://orcid.org/0000-0001-9425-9598
E-mail: luana_elayne@hotmail.com

Contato com os autores:

Universidade de Fortaleza, Departamento de Psicologia
Av. Washington Soares, 1321, Edson Queiroz
Fortaleza-CE, Brasil
CEP: 60811-905