Published in final edited form as:

Pers Individ Dif. 2013 January 1; 54(2): 153–157. doi:10.1016/j.paid.2012.08.009.

Evaluation of the Factor Structure of the Rosenberg Self-Esteem Scale in Older Adults

Sean P. Mullen, Neha P. Gothe, and Edward McAuley

University of Illinois at Urbana-Champaign

Abstract

The Rosenberg Self-Esteem Scale is the most utilized measure of global self-esteem. Although psychometric studies have generally supported the uni-dimensionality of this 10-item scale, more recently, a stable, response-bias has been associated with the wording of the items (Marsh, Scalas, & Nagengast, 2010). The purpose of this report was to replicate Marsh et al.'s findings in a sample of older adults and to test for invariance across time, gender and levels of education. Our results indicated that indeed a response-bias does exist in esteem responses. Researchers should investigate ways to meaningfully examine and practically overcome the methodological challenges associated with the RSE scale.

Keywords

global self-esteem; psychometrics; aging

1. Introduction

Self-esteem has been an integral construct in the field of psychology for decades. The field's most commonly used measure of global self-esteem is the 10-item Rosenberg's self-esteem (RSE) scale (Rosenberg, 1965). This scale has been used extensively with samples of all ages, from adolescents to older adults. Although the psychometric properties of this self-report measure have been rigorously tested, researchers have questioned whether the positively and negatively worded items are interchangeable, i.e., assess the same construct (Corwyn, 2000; DiStefano & Motl, 2009; Marsh, 1996; Motl & DiStefano, 2002; Wang, Siegal, Falck, & Carlson, 2001). Few researchers understand the origins of method effects, however there is a strong case for controlling consistent bias associated with the RSE scale, despite criticism of the general post-hoc approach to account for "common method variance" (Conway & Lance, 2010; Lance, Dawson, Birkelbach, & Hoffman, 2010; Richardson, Simmering, & Sturman, 2009). Overall, researchers agree that methodological decisions should be guided by substantive and theoretical arguments.

Recently, Marsh et al. (2010) systematically tested multiple models of global self-esteem based on the RSE scale to determine the extent to which method effects were ephemeral or stable. They used two approaches, correlated uniquenesses and latent method factors, to

© 2012 Elsevier Ltd. All rights reserved.

Address all correspondence to: Sean Mullen, University of Illinois at Urbana Champaign, Department of Kinesiology and Community Health, 906 S. Goodwin Ave., Urbana, IL 61801, Phone: (217) 244-4502, spmullen@illinois.edu, Fax: (217)333-3124.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

account for the hypothesized method effects. Each approach has strengths and weaknesses. Most notably, the correlated uniquenesses approach assumes different types of method bias are uncorrelated with each other, and the latent method approach relaxes this assumption. However, the latter solutions are prone to producing inadmissible solutions and do not converge as easily as the former approaches. Overall, Marsh et al found a consistent response-style bias associated with the item wording of the RSE. Furthermore, Marsh et al claimed these method effects call into question the vast literature based on the RSE and that “failure to control for them will bias the interpretations of RSE responses” (p. 378). However, one of the limitations of their study was that their findings were based solely on the responses of adolescent males.

To our knowledge, only one study (Whiteside-Mansell & Corwyn, 2003) has tested the possibility of differences in global self-esteem measurement across age. However, Whiteside-Mansell and Corwyn only tested the invariance of the RSE scale’s structure across a sample of 12 to 17 year olds ($M_{age} = 14.8$) and a sample of 18 to 80 year olds ($M_{age} = 33$). Thus, the “adult” sample contains young, middle, and older adults, and it cannot be determined whether the esteem construct has the same meaning for all age groups. It is reasonable to expect that older adults may interpret and respond to questionnaires differently, based on established age differences in “affective balance” during emotional self-report (Robinson & Clore, 2002) and memory for positive and negative events (Mather & Carstensen, 2005). Furthermore, no studies have explored potential individual differences in the interpretation of RSE, across gender or levels of education among older adults. The purpose of this study was to replicate Marsh et al.’s (2010) findings in a homogenous sample of older adults and to extend this work by exploring potential differences across subgroups. Given that there are considerable implications for older adults’ self-esteem, it is important to verify the structural integrity of the scale in older populations.

2. Method

2.1 Participants, procedure, and measures

Data were collected from sedentary older adults ($n = 603$) as part of a baseline questionnaire packet prior to participation in an exercise program; a smaller subsample ($n = 298$) completed the questionnaire packet 12 months later. The sample were community-dwelling older adults ($M_{age} = 69.94$, $SD = 5.66$; range = 60–95), mostly white (94.7%, vs. 3.7% Black/African-American, 1.3% Asian, .3% American Indian/Alaskan Native; .3% missing), female (72.5%, $n = 437$), married (59.4%), and earned an income of at least 40K (54.1%). Approximately half of the sample graduated from college or attained higher education (47.3%). Participants completed the original 10-item RSE (Rosenberg, 1965) along with demographic information. The responses on the scale were measured on a 5-point Likert scale: 1 (strongly agree), 2 (agree), 3 (neutral), 4 (disagree), and 5 (strongly disagree). Five of the items are positively-worded (items 1, 2, 4, 6, and 7) whereas the remaining five are negatively-worded (3, 5, 8, 9, and 10); negative items were reverse-coded prior to data analysis.

All modeling was conducted using raw data with version 6.1 of *Mplus* (Muthén & Muthén, 1998–2012) and we used full information for missing data with robust maximum likelihood estimation (MLR). We elected to use multiple criteria for evaluating model misspecification, including the chi-square statistic (χ^2), root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis Index (TLI). Accordingly, it is recommended that models should result in significant χ^2 values ($p > .05$), RMSEA of $< .06$, CFI values of $> .95$ (Hu & Bentler, 1999), and TLI values of $> .95$ (Marsh, Hau, & Grayson, 2005). Paralleling the procedures used by Marsh et al (2010), we tested eight models, including a 1-factor structure with no additional parameter constraints (Model 1), a model

involving a 2-factor latent structure, i.e., positive, negative (Model 2), a 1-factor model with correlated residuals among negative and among positive items (Model 3), a 1-factor model with only correlated negative residuals (Model 4), a 1-factor model with correlated positive residuals (Model 5), a 1-factor model with two method factors, i.e., representing systematic error in responses to positive and negatively-worded items (Model 6), a 1-factor model with only a negative method effect (Model 7), and a 1-factor with only a positive method effect (Model 8). Note that method factors were specified to covary a priori. Lastly, we also conducted invariance testing with the best-fitting models. The invariance routine we employed involved adding sequential restrictions to test equality of factor configurations (i.e., configural invariance), followed by loadings (i.e., metric invariance), intercepts (i.e., scalar invariance), residuals (i.e., strict invariance), and latent means and variances across measurement occasions. More restrictive models were deemed invariant from less restrictive models if the corrected Satorra-Bentler (S-B) χ^2 difference (Δ) test (Satorra & Bentler, 2001) was not significant ($p > .05$), $\Delta CFI < .01$ (Cheung & Rensvold, 2002) and $\Delta RMSEA < .015$ (Chen, 2007).

3. Results

3.1 Preliminary Analyses

The majority of items were found to be negatively skewed. In addition, the variability in item 2 (i.e., “I feel that I have a number of good qualities”) was restricted to just three responses (range = 1 to 3) at the second measurement occasion. We therefore employed MLR estimation.

3.2 Measurement Models

3.2.1 Overall sample—Model 1 showed a poor fit to the data (e.g., RMSEA = .185, CFI = .758) and was the poorest fitting model tested. Model 2 was an improvement, but still failed to fit the data according to χ^2 and recommended cutoff values for RMSEA and TLI. Model 3 failed to converge, as was the case in some of the samples reported by Marsh et al (2010). Similar to Model 2, Model 4 did not fit according to χ^2 , RMSEA, and TLI. However, Model 5 (χ^2 p value $> .05$; RMSEA = .011, CFI = .999, TLI = .999) and Model 6 provided an excellent fit (χ^2 p value $> .05$, RMSEA = .023, CFI = .997, TLI = .995). For Model 5, overall factor loadings ranged from .024 to .776) and the correlated uniquenesses among positive items ranged from .790 to .960, indicating a method effect. For Model 6, factors loadings for the overall model were low and not significant (range = .053 to .593), whereas significant loadings were found among positive items (range = .748 to .977) and negative items (range = .535 to .778), suggesting two method effects. Again, Model 7 failed to meet all recommended criteria (e.g., TLI = .919, RMSEA = .094), as did Model 8 (e.g., TLI = .917, RMSEA = .096). In sum, the best-fitting representations of RSE’s factor structure were Models 5, reflecting correlations among positively worded items, and Model 6, reflecting two underlying uncorrelated method factors. Fit indices for all measurement models are included in Table 1, whereas conceptual diagrams of the best-fitting models, i.e., Model 5 and 6 are displayed in Figure 1. See Table 2 for factor loadings and item uniquenesses associated with the best-fitting models.

In addition to examining cross-sectional measurement models, we attempted to replicate Marsh et al.’s findings showing temporal invariance which would indicate a stable response-style bias. Following conventional procedures, we first tested configural invariance (i.e., same items regressed on same constructs) for models 5 and 6. Model 5 provided an adequate fit to the model ($\chi^2 = 248.059$ (139), $p < .001$, RMSEA = .051, CFI = .940, TLI = .918), although the fit significantly worsened from the baseline model. Model 6 did not provide an admissible solution at this stage and invariance testing was terminated. For Model 5,

relatively speaking, the metric invariance model, i.e., loadings constrained to equality across time ($\chi^2 = 270.618$ (148), $p < .001$, RMSEA = .053, CFI = .933, TLI = .913) did not significantly differ from the configural invariance model based on S-B χ^2 test, Δ CFI or Δ RMSEA. The scalar invariance model, i.e., intercepts constrained to be equal across time ($\chi^2 = 282.557$ (157), $p < .001$, RMSEA = .052, CFI = .931, TLI = .916) also did not significantly change the fit. The same was also true for the strict invariance model, i.e., residual variances and correlated uniquenesses (method effect) constrained to be equal across time ($\chi^2 = 314.968$ (172), $p < .001$, RMSEA = .053, CFI = .921, TLI = .913), latent mean invariance ($\chi^2 = 316.767$ (173), $p < .001$, RMSEA = .053, CFI = .921, TLI = .913) and latent variance invariance ($\chi^2 = 318.998$ (174), $p < .001$, RMSEA = .053, CFI = .920, TLI = .913) based on all three criteria. Together, this evidence confirms Marsh et al.'s conclusions that these method effects are not ephemeral, but rather they reflect a stable response-bias over time.

3.2.2 Gender analyses—Gender invariance testing proceeded via the testing of identical model configurations in both groups for Models 1 through 8. At this first step, Model 1 showed a poor fit to the data ($\chi^2 = 818.031$ (70), $p < .001$; RMSEA = .188, CFI = .756, TLI = .686). With the exceptions of Model 3 (which did not converge) and Model 6 (produced a Heywood case, i.e., negative residual variance, for item 2), the remaining models improved the model-to-data fit, including Model 2 ($\chi^2 = 250.273$ (88), $p < .001$; RMSEA = .078, CFI = .947, TLI = .946), Model 4 ($\chi^2 = 215.227$ (50), $p < .001$; RMSEA = .105, CFI = .946, TLI = .903), Model 7 ($\chi^2 = 235.707$ (80), $p < .001$; RMSEA = .080, CFI = .949, TLI = .943), and Model 8 ($\chi^2 = 231.770$ (80), $p < .001$; RMSEA = .079, CFI = .950, TLI = .944). However, only Model 5 ($\chi^2 = 50.867$ (50), $p = .439$; RMSEA = .008, CFI = 1.000, TLI = .999) met recommended criteria.

Therefore, we proceeded with invariance testing for Model 5 only. Model 5 did not significantly change when constraints were added for metric invariance ($\chi^2 = 65.718$ (59), $p = .256$; RMSEA = .019, CFI = .998, TLI = .997), scalar invariance ($\chi^2 = 76.598$ (68), $p = .222$; RMSEA = .020, CFI = .997, TLI = .996), strict residual invariance ($\chi^2 = 88.087$ (78), $p = .204$; RMSEA = .021, CFI = .997, TLI = .996), latent mean invariance ($\chi^2 = 89.116$ (79), $p = .205$; RMSEA = .021, CFI = .997, TLI = .996), or latent variance invariance ($\chi^2 = 91.727$ (80), $p = .174$; RMSEA = .022, CFI = .996, TLI = .996). These results further support the findings above, suggesting that Model 5 best represents the response pattern of older adults. Furthermore, there were no gender differences in the levels of the latent means of GSE.

3.2.3 Educational analyses—The remaining model tests compare differences across educational levels (i.e., participants without a college degree vs. those with a college degree). Mirroring the procedures and findings above, Model 1, showed a poor fit to the data ($\chi^2 = 855.363$ (70), $p < .001$; RMSEA = .193, CFI = .753, TLI = .682) and Model 3 did not converge, whereas Model 6 provided a Heywood case. The remaining models improved the fit relative to Model 1, including Model 2 ($\chi^2 = 261.071$ (68), $p < .001$; RMSEA = .097, CFI = .939, TLI = .920), Model 4 ($\chi^2 = 230.661$ (50), $p < .001$; RMSEA = .109, CFI = .943, TLI = .898), Model 7 ($\chi^2 = 285.841$ (80), $p < .001$; RMSEA = .092, CFI = .935, TLI = .927), and Model 8 ($\chi^2 = 289.154$ (80), $p < .001$; RMSEA = .093, CFI = .934, TLI = .926). However, only Model 5 ($\chi^2 = 55.180$ (50), $p = .285$; RMSEA = .019, CFI = .998, TLI = .997) met recommended criteria.

Based on these initial findings, we again proceeded with invariance testing of Model 5 only. The model-to-data fit did not significantly change when restrictions were added for metric invariance ($\chi^2 = 65.267$ (59), $p = .268$; RMSEA = .019, CFI = .998, TLI = .997) or scalar invariance ($\chi^2 = 81.072$ (68), $p = .133$; RMSEA = .025, CFI = .996, TLI = .995). The

residual invariance model ($\chi^2 = 109.863$ (78), $p = .010$; RMSEA = .037, CFI = .990, TLI = .998) was significantly different from the scalar model based on S-B χ^2 test, but not Δ RMSEA or Δ CFI, and thus we feel this difference may be trivial. Latent mean invariance ($\chi^2 = 114.128$ (79), $p = .006$; RMSEA = .038, CFI = .989, TLI = .987) did not differ from the strict invariance model, and the addition of latent variance invariance did not change the fit ($\chi^2 = 115.551$ (80), $p = .006$; RMSEA = .038, CFI = .989, TLI = .987). Again, we found Model 5 to best represent the data, and we found no substantive differences across levels of education.

3.2.4 Exploratory Analyses involving Age and Functional Difficulty—Given the potential for differences in response patterns with increased age, we conducted further invariance testing. Specifically, we split the sample in half based on age (i.e., coded 0 if 70 years or younger, and 1 if older than 70) to test invariance across younger and older subgroups. However, the pattern of results remained nearly identical to the findings aforementioned. Additionally, we conducted a differential item functioning (DIF) analysis to assess the potential impact of having no versus one or more chronic functional impairments. Not a single item was influenced by the addition of this covariate to the model. Fit statistics and parameter estimates for these models are available from the first author.

4. Discussion

The purpose of this study was to replicate Marsh et al.'s (2010) findings that accounting for response-bias based on positive and negative items resulted in a better representation of RSE responses than the uni-dimensional model. Using the original RSE scale in a large sample of older adults, we found strong support for a response-style bias. Like Marsh and his colleagues, our best-fitting models with our overall sample included two unique approaches for assessing method factors: a 1-factor model with correlated positive residuals (Model 5), and a 1-factor model with two method factors, representing systematic error in responses to positive and negatively-worded items (Model 6). Furthermore, we found invariance of Model 5's factor structure across gender, levels of education, and age groups. Marsh et al stated that applied researchers should control for known method effects, "whether or not their meaning is understood" (p. 378). We agree and we offer some practical considerations.

As we expected, older adults exhibited a response bias. However, we felt that the most robust model reflected a positive response bias, which differs from findings reported by Marsh et al. (2010) and DiStefano and Motl (2009). However, their samples included all boys and young adults, respectively. It has been well-established that accessibility of emotions can greatly bias self-report (Robinson & Clore, 2002), and older adults tend to strategically focus on positive emotions more so than younger adults (Mather & Carstensen, 2005), which may cause them to interpret and respond to positive items in a similar fashion. One could speculate that, as a function of the aging process, negative self-appraisals may become more differentiated and highly contextualized relative to positive self-appraisals (i.e., serving an adaptive self-enhancement mechanism). If such a shift in interpretation is indeed part of a developmental process, one might expect similar response patterns across older adult subgroups, irrespective of demographics. We found some evidence of this with our invariance testing. It is also interesting to note that item 1 (i.e., I feel that I'm a person of worth, at least on an equal basis with others) and item 4 (i.e., I am able to do things as well as most people) prompt respondents to make social comparisons, and yet no differences were found between older adults with and without functional difficulties.

Some researchers have found different types of method effects, depending on the method of data collection (paper and pencil vs. computerized methods; Vispoel, Boo, & Bleiler, 2001). Practically speaking, it may be prudent to use just one item, an approach validated in prior

research (Robins, Hendin, & Trzesniewski, 2001). For example, singular items that do not require respondents to consider self-appraisals across a variety of situations such as, “Overall, I feel that I am a person of worth,” may accurately reflect global esteem in late-life. Certainly there are limitations to 1-item measures (e.g., reliability issues, reduced ability to capture depth and breadth of a construct), yet global self-esteem has been theorized to be reflective of everything from additive to interactive self-attributes and these are not being adequately measured with existing scales. Furthermore, global self-esteem should be the most contextually-independent assessment of esteem, and rating a general, neutrally-worded statement about overall worth may be sufficient to quickly assess esteem status within the context of clinical practice.

Research by Meier, Orth, Denissen, and Kuhnel (2011) suggests that self-esteem becomes more stable, less contingent, and higher with increasing age. It is important to note, however, that invariance across age groups should be conducted before we can make an unambiguous interpretation of mean level comparisons. Meier et al also reported that self-esteem instability and contingency was similar to the personality trait of neuroticism and inversely related to conscientiousness. These trends are consistent with DiStefano and Motl’s (2009) study, which showed that personality characteristics reflecting fear of negative evaluation, among a mostly young adult sample, were associated with a method effect for negatively worded items of the RSE scale. They also pointed out that there are neurological correlates associated with response styles (Motl & DiStefano, 2002). Given that the most robust method effect in our study was for positively worded items, and the past work supporting age differences in emotional regulation (Mather & Carstensen, 2005; Ryff, 1989), there may be true differences in these response-bias processes across the lifespan. For example, older adults may be less reserved, want to think of themselves in a positive light, or may spend more or less time carefully reading the questions than younger adults. In the future, self-esteem researchers may want to attempt to revise the scale so that the negatively worded items are reversed to reduce the social desirability bias. Item 2 (“I feel that I have a number of good qualities”), especially, deserves attention due to the limited response-range we observed at time 2. Personality researchers may also want to further delineate the brain regions associated with such response styles and the personality traits that encourage certain biases.

The major strength of this study is that it replicates Marsh et al.’s (2010) results within a large sample of older adults. However we acknowledge that the majority of our sample comprised of White, educated women earning a relatively high income. Thus, this sample was reasonably well-adjusted. Before one may draw conclusions about the generalizability of the measurement issues addressed herein, it is important to investigate the psychometric properties of the scale in more diverse samples that may be less equipped to adapt to life challenges. In spite of the limitations, we believe that these findings are critical to the use and interpretation of the RSE with samples of older adults who typically volunteer for physical activity research, as our sample characteristics mirror the characteristics reported by studies of older populations for the past two decades (Keysor, 2003; Liu & Latham, 2009).

Self-esteem measures are extensively used in the physical activity and quality of life research with older adult populations, and the response style may lead to incorrect interpretations by suppressing or inflating the RSE associations with other study variables. Future research is needed to address the generalizability of our findings. Finally we caution others to restrict comparisons of esteem levels to samples that have been tested for invariance and response-bias.

Acknowledgments

This project was partially supported by grants from the National Institute on Aging (R01 AG020118 and R37 AG025667).

References

- Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*. 2007; 14(3):464–504.
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9(2):233–255.
- Conway JM, Lance CE. What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*. 2010; 25(3):325–334.
- Corwyn RF. The factor structure of global self-esteem among adolescents and adults. *Journal of Research in Personality*. 2000; 34(4):357–379.
- DiStefano C, Motl RW. Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences*. 2009; 46(3):309–313.
- Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6(1):1–55.
- Keyser JJ. Does late-life physical activity or exercise prevent or minimize disablement?: A critical review of the scientific evidence. *American Journal of Preventive Medicine*. 2003; 25(3):129–136. [PubMed: 14552936]
- Lance CE, Dawson B, Birkelbach D, Hoffman BJ. Method effects, measurement error, and substantive conclusions. *Organizational Research Methods*. 2010; 13(3):435–455.
- Liu CJ, Latham NK. Progressive resistance strength training for improving physical function in older adults. *Cochrane Database Syst Rev*. 2009; (3) CD002759.
- Marsh HW. Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*. 1996; 70(4):810–819. [PubMed: 8636900]
- Marsh, HW.; Hau, KT.; Grayson, D. Goodness of fit in structural equation models. In: M.-O, A.; McArdle, J., editors. *Psychometrics A Festschrift to Roderick P McDonald*. Hillsdale, NJ: Erlbaum; 2005.
- Marsh HW, Scalas LF, Nagengast B. Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment*. 2010; 22(2):366–381. [PubMed: 20528064]
- Mather M, Carstensen LL. Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*. 2005; 9(10):496–502. [PubMed: 16154382]
- Meier LL, Orth U, Denissen JJA, Kuhnel A. Age differences in instability, contingency, and level of self-esteem across the life span. *Journal of Research in Personality*. 2011; 45(6):604–612.
- Motl RW, DiStefano C. Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling*. 2002; 9(4):562–578.
- Muthén, LK.; Muthén, BO. *Mplus (Version 6.0)*. Los Angeles, CA: 1998–2012.
- Richardson HA, Simmering MJ, Sturman MC. A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*. 2009; 12(4):762–800.
- Robins RW, Hendin HM, Trzesniewski KH. Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*. 2001; 27(2):151–161.
- Robinson MD, Clore GL. Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*. 2002; 128(6):934–960. [PubMed: 12405138]
- Rosenberg, M. *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press; 1965.

- Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*. 1989; 57(6):1069–1081.
- Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66(4):507–514.
- Vispoel WP, Boo J, Bleiler T. Computerized and paper-and-pencil versions of the Rosenberg Self-Esteem Scale: A comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement*. 2001; 61(3):461–474.
- Wang J, Siegal HA, Falck RS, Carlson RG. Factorial structure of Rosenberg's Self-Esteem Scale among crack-cocaine drug users. *Structural Equation Modeling*. 2001; 8(2):275–286.
- Whiteside-Mansell L, Corwyn RF. Mean and covariance structures analyses: An examination of the Rosenberg Self-Esteem Scale among adolescents and adults. *Educational and Psychological Measurement*. 2003; 63(1):163–173.

Highlights

RSE scale tested among older adults

Positive response bias

Scale revisions discussed

\$watermark-text

\$watermark-text

\$watermark-text

Model 5 **Model 6**

Figure 1.
Best-fitting structural models for Rosenberg Self-Esteem.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Fit indices for all measurement models based on the entire sample ($N = 603$)

Model	χ^2	df	p value	CFI	TLI	RMSEA (90% Interval)
1	756.283	35	<.001	.758	.689	.185 (.174-.196)
2	200.089	34	<.001	.944	.926	.090 (.078-.102)
4	179.140	25	<.001	.948	.907	.101 (.087-.115)
5	26.803	25	.368	.999	.999	.011 (.000-.035)
6	33.237	25	.125	.997	.995	.023 (.000-.043)
7	190.278	30	<.001	.946	.919	.094 (.082-.107)
8	195.239	30	<.001	.945	.917	.096 (.083-.109)

Note. Model 3 did not successfully converge.

Table 2

Parameter estimates for Models 5 and 6

Item	Model 5		Model 6	
	ROSE	Uniquenesses	ROSE	Uniquenesses
1	<i>ns</i>	.998	<i>ns</i>	.949
2	<i>ns</i>	.999	<i>ns</i>	.977
3	.666	.556	<i>ns</i>	.659
4	.083	.993	<i>ns</i>	.807
5	.539	.710	<i>ns</i>	.535
6	<i>ns</i>	.999	<i>ns</i>	.802
7	.082	.993	<i>ns</i>	.748
8	.665	.558	<i>ns</i>	.656
9	.776	.398	<i>ns</i>	.778
10	.762	.420	<i>ns</i>	.756

Note. ROSE = Rosenberg Self-Esteem latent factor loadings; POSITIVE = factor loadings for positively-worded items; NEGATIVE = factor loadings for negatively-worded items; *ns* = not significant.