

# UC Irvine

## UC Irvine Previously Published Works

### Title

Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: do they matter?

### Permalink

<https://escholarship.org/uc/item/4bc425gr>

### Journal

Personality and Individual Differences, 35(6)

### ISSN

0191-8869

### Authors

Greenberger, Ellen  
Chen, Chuansheng  
Dmitrieva, Julia  
[et al.](#)

### Publication Date

2003-10-01

### DOI

10.1016/s0191-8869(02)00331-8

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



## Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: do they matter?

Ellen Greenberger\*, Chuansheng Chen, Julia Dmitrieva, Susan P. Farruggia

*Department of Psychology and Social Behavior, School of Social Ecology, University of California,  
Irvine, CA 92697-7085, USA*

Received 11 March 2002; received in revised form 25 September 2002; accepted 29 October 2002

---

### Abstract

Some researchers contend that the Rosenberg Self-Esteem Scale taps two dimensions of self-image, whereas others argue that the two dimensions (positive and negative) are merely an artifact of item wording. To directly test these competing views, we had 741 ethnically diverse university undergraduates take one of three versions of the 10-item Rosenberg Scale: the original version comprised of five positively worded and five negatively worded items, or one of two alternative versions comprised of 10 positively worded or 10 negatively worded items. Analyses indicated that the original version fit a two-factor model, whereas the reworded versions generally fit a one-factor model. All three versions had high validity for different ethnic groups, but the revised-positive version had less overlap with a measure of depression, and both revised versions had less overlap with a measure of self-deception.

© 2003 Elsevier Ltd. All rights reserved.

*Keywords:* Self-esteem; Dimensionality; Item-wording; Construct validity; Confirmatory factor analysis; Rosenberg Scale

---

The widely used Rosenberg Self-Esteem Scale was conceptualized by its author as a single-factor scale with scores ranging along a continuum of low self-esteem to high self-esteem (Rosenberg, 1965). The “actual” or empirical factor structure of this scale has been the target of debate, however, for over 30 years (Owens, 1994). Several researchers who conducted factor analyses of the 10-item scale have suggested that the scale reflects a two-dimensional construct, comprised of positive and negative images of the self (Bachman & O’Malley, 1986; Goldsmith, 1986; Kaplan & Pokorny, 1969; Owens, 1993). In these and other studies, the five positively-worded items (e.g. “I feel that I have a number of good qualities”) loaded on one factor, referred

---

\* Corresponding author. Tel.: +1-949-824-6328; fax: +1-949-824-3002.

*E-mail addresses:* egreenbe@uci.edu (E. Greenberger), cschen@uci.edu (C. Chen).

to variously as “positive self-esteem,” “positive self-worth,” and “positive self-image,” whereas the five negatively worded items (e.g. “At times I feel that I am no good at all”) loaded on a separate factor, referred to variously as “negative self-esteem,” “self-derogation,” “self-deprecation,” and “negative self-image.”

In support of their contention that Rosenberg’s Scale taps two distinct dimensions of self-esteem, some researchers have argued that the positive and negative dimensions of this measure lead to different outcomes and are influenced by different experiences. For example, Owens (1994), based on a longitudinal study using an eight-item version of Rosenberg’s scale (with slight re-wording and a slightly different response-continuum), examined the reciprocal effects between three scores derived from the scale [total or Global Self-Esteem, “Self-Deprecation” (i.e. the subscale comprised of negatively worded items), and “Positive Self-Worth” (i.e. the subscale comprised of positively worded items)] and measures of grades, depression, and self-reported delinquency. Among the results of the study, self-deprecation scores were substantially more related to depression in both causal directions, i.e. self-deprecation led to increases in depression over time, and depression also led to increases in self-deprecation, than were scores on either the positive self-worth subscale or total (global) self-esteem scores. In a related study, using samples of British college students and adults, Sheasby, Barlow, Cullen, and Wright (2000) found evidence favoring a two-factor structure of the Rosenberg Scale, with negatively worded items all loading on one factor. However, one positively worded item emerged on the “negative items” factor for both samples, and a second positively worded item loaded on the factor for the adults in this study. In a recent, cross-cultural study (Farruggia, Chen, Greenberger, Dmitrieva, Tally, & Macek, unpublished manuscript), a single-factor solution for the Rosenberg Self-Esteem Scale did not fit the data for adolescent samples in any of the four countries (US, Czech Republic, China, and Korea), but a modified, nine-item, two-factor model fit the data adequately. Although the Owens study, in particular, appears to make a good case for the nuances of self-esteem that may be uncovered when both positive and negative dimensions of self-esteem are considered, the earlier mentioned studies did not test the possibility that the apparent bi-factorial structure of the Rosenberg scale may be an artifact of item-wording.

The possibility that the presumed two-factor structure of Rosenberg’s Self-Esteem Scale is an artifact of item-wording has been proposed by several researchers (e.g. Carmines & Zeller, 1979; Hensley & Roberts, 1976; Marsh, 1996; Tomas & Oliver, 1999). For example, the two factors that emerge may merely reflect issues of response set or response bias. Thus, positive and negative self-esteem factors might result from respondents’ tendency to agree with positively worded statements about the self and disagree with negatively worded statements about the self, or from more general tendencies to engage in “yea-saying” or “nay-saying” (Couch & Keniston, 1960). Marsh (1996) took a confirmatory analysis approach to examining the dimensionality of a seven-item version of the Rosenberg scale (four positively worded items and three negatively worded items) and compared the validity coefficients of Global (total) Self-Esteem, Positive Self-Esteem and Negative Self-Esteem) with 13 measures relevant to the validity of these three measures. Marsh’s analyses of Rosenberg’s scale reflected a single, substantively meaningful factor and the existence of a method effect, primarily due to the negative items, that complicates interpretation of the total or global Self-Esteem score. Marsh argued that agreeing with negative items such as “I feel I do not have much to be proud of,” but especially *disagreeing with* or negating negatively worded items, adds a degree of cognitive complexity to the task of responding to a questionnaire. In

support of this view, Marsh (1996) demonstrated that students with poorer verbal ability were especially susceptible to making responses to negative items that were inconsistent with their responses to positively worded items.

Two recent studies have built upon the work of Marsh (1996) by examining different models of the Self-Esteem Scale, focusing especially on method effects (i.e. Corwyn, 2000; Tomas & Oliver, 1999). Tomas and Oliver (1999) examined nine models, six of which took into account method effects, and Corwyn (2000) examined eight models, three of which incorporated method effects. Both studies concluded that the Rosenberg Self-Esteem Scale has a unidimensional factor structure, after controlling for method effects. The model that emerged as having the best fit for Tomas and Oliver (1999) included method effects for both positively worded and negatively worded items; however, Corwyn (2000) found the method effects to occur primarily for the negatively worded items.

In this study we pursue three objectives. First and most importantly, we undertake the only “direct” approach of which we are aware to examine whether item-wording accounts for the two-dimensionality of the Rosenberg Self-Esteem Scale. Specifically, based on Rosenberg’s original 10-item scale (original RSES), we created an all negatively worded version of the scale (Revised-negative version), in which previously positive items were re-written in the negative form, and an all-positively worded version (Revised-positive version), in which previously negatively worded items were written in the positive direction. The revised versions, therefore, are “uni-directional” in terms of their assessment of self-esteem. If a two-factor model should emerge only in relation to Rosenberg’s original scale (Original RSES), the artifactual nature of the positive and negative dimensions of self-esteem derived from his scale would be supported. That is, item-wording would “matter” in understanding why the original scale seems to reveal a bi-factorial structure. If, however, a two-factor model should fit equally well across *all* scale-versions, the contention that Rosenberg’s scale is two-dimensional, not uni-dimensional, would be supported. Such a result would imply that the two subsets of items, independent of the direction in which they are worded, tap somewhat different dimensions or domains of self-esteem.

Second, we investigate the construct validity of Self-Esteem scales based on the Original RSES and Revised (positive and negative) versions of the Rosenberg scale. For the construct validity tests, we chose measures that previously have been shown to have associations with total self-esteem scores (e.g. depressive symptomatology, as reported by Owens, 1994; and Rosenberg, 1965). If scores based on Original RSES and Revised versions were to show the *same* direction and magnitude of associations with validity-relevant variables, it would suggest that item-wording does not affect the validity of the Rosenberg scale. Such a finding, in combination with the finding that item-wording does not affect the factor-structure or dimensionality of the Rosenberg scale, would make a convincing argument for the integrity of the original scale. Of course, the combined results of our structural analysis of the three scale-versions and the findings from our validity analyses could lead to other conclusions.

Third, we explore whether there are gender and ethnic differences in the dimensionality and validity of the Original RSES and the two Revised versions of the scale. Previous research has revealed significant gender differences in mean self-esteem scores. Adolescent boys tend to have higher self-esteem than adolescent girls (Harter, 1990). However, this may not be true for all girls as ethnic differences in girls’ self-esteem have been found. Specifically, White and Latina girls have lower self-esteem and show a greater decline in self-esteem during adolescence than do

African-American adolescent girls (Gray-Little & Hafdahl, 2000). In the current study, we move beyond that issue and focus on whether the dimensionality and validity of the three versions of the Self-Esteem Scale are similar across gender and ethnic groups. This is an important issue inasmuch as the usefulness of a scale depends on its applicability across populations.

## **1. Methods**

### *1.1. Participants*

The sample consisted of 741 undergraduate students enrolled in social science-related courses at a public university in California. The mean age of participants was 20.1 years ( $SD = 2.88$ ). The gender and ethnic composition of the sample quite accurately reflected that of the Schools in which the participants were majoring. Thirty-eight percent of the sample was male, and 62% was female. The ethnic background of participants was: 31% East Asian (e.g. Korean, Chinese, Japanese), 19% Southeast Asian (e.g. Vietnamese, Filipino, Cambodian, Thai), 7% South Asian (Indian, Pakistani, Iranian), 20% European American, 11% Latino, 2% African American, and 11% Other (including having parents from two different ethnic categories, as defined earlier). Three-quarters of the students came from families with intact marriages. There was considerable variation in both fathers' and mothers' educational attainment: although 54% of fathers and 42% of mothers had a 4-year college or graduate degree, more than 10% had only a junior high school education. Twenty percent of the participants were born outside the US (i.e. first-generation immigrants); 46% were born in the US, but one or both of their parents were born outside the US (i.e. second-generation immigrants); the remaining 24% of respondents, as well as their parents, were US-born (i.e. third-generation or higher immigrants).

### *1.2. Procedures*

Participants completed an anonymous, self-report questionnaire administered in the classroom or in the laboratory to groups of participating students. Informed consent was obtained prior to the administration of the survey. Participants received either a small amount of extra course-credit, or none, and had the option to decline participation. Participants were randomly assigned to complete one of the three versions of the Rosenberg Self-Esteem Scale. All other measures were identical across the three conditions, as was the order in which the various measures appeared in the survey booklet.

### *1.3. Measures*

Three versions of the Self-Esteem Scale were administered (see Table 1). Original RSES ( $n = 257$ ) is the original 10-item Rosenberg Self-Esteem Scale, containing five positively worded items and five negatively worded items. Revised-negative version ( $n = 244$ ) is an adaptation of Rosenberg's scale in which the five positively worded items in the original scale have been rephrased in a negative direction, resulting in a 10-item scale with all negatively worded items. Revised-positive version ( $n = 240$ ) is an adaptation of Rosenberg's Scale in which the five negatively

Table 1  
The Rosenberg Self-Esteem Scale and two re-worded versions<sup>a</sup>

The original RSES	Revised-negative version	Revised-positive version
1. On the whole, I am satisfied with myself	1. On the whole, I am not satisfied with myself (–)	1. On the whole, I am satisfied with myself
2. At times I think I am no good at all (–)	2. At times I think I am no good at all (–)	2. At times I think I am pretty darn good
3. I feel that I have a number of good qualities	3. I don't feel that I have many good qualities (–)	3. I feel that I have a number of good qualities
4. I am able to do things as well as most other people	4. I am not able to do things as well as most other people (–)	4. I am able to do things as well as most other people
5. I feel I do not have much to be proud of (–)	5. I feel I do not have much to be proud of (–)	5. I feel I do have much to be proud of
6. I certainly feel useless at times (–)	6. I certainly feel useless at times (–)	6. I really feel useful at times
7. I feel that I'm a person of worth, or at least on an equal plane with others	7. I feel that I'm a person of little worth, not on an equal basis with others (–)	7. I feel that I'm a person of worth, or at least on an equal plane with others
8. I wish I could have more respect for myself (–)	8. I wish I could have more respect for myself (–)	8. I think I have enough respect for myself
9. All in all, I am inclined to feel that I am a failure (–)	9. All in all, I am inclined to feel that I am a failure (–)	9. All in all, I am inclined to feel that I am not a failure
10. I take a positive attitude toward myself	10. I take a negative attitude toward myself (–)	10. I take a positive attitude toward myself

<sup>a</sup> – Indicates items that were reverse-scored when calculating the total self-esteem score.

worded items in the original scale have been rephrased in a positive direction, resulting in a 10-item scale with all positively worded statements (Table 1). It should be noted that adapting previously positively worded items to a negative version, and vice versa, was usually a matter of inserting or deleting the word “not” or changing a negative word to a positive one (e.g. “useless” from Original RSES, to “useful” for the Revised-positive version). However, in some instances modest additional changes were necessary. In the case of item 8, for example, Rosenberg's original phrasing “I wish I could have more respect for myself” (a negative item) was revised to read, “I think I have enough respect for myself” for the Revised-positive version. A more literal “conversion,” such as “I do not wish I could have more respect for myself,” seemed linguistically awkward and difficult to comprehend. On all versions, respondents answered on a six-point Likert scale, from “1” = *Strongly disagree* to “6” = *Strongly agree*.

Total Self-Esteem scores were calculated as follows: For RSE, the mean score for the positively worded items and the reverse-scored, negatively worded items; for Revised-negative version, the mean for all items after each item had been reverse-scored; for Revised-positive version, the mean of all responses as reported by participants. Cronbach's alpha for total Self-Esteem was 0.88 for Original RSES, 0.91 for Revised-negative version, and 0.92 for Revised-positive version.

Parental warmth and acceptance (Greenberger, Chen, & Beam, 1998) was measured by an 11-item scale that includes items such as “My parents let me know through words or actions that they love me” and “I find it hard to please my parents” (reverse-scored). Respondents indicated their agreement on a six-point Likert scale, anchored by the terms “*strongly disagree*” and “*strongly agree*.” This scale has been shown to be significantly and inversely associated with depressive symptomatology in US and Chinese adolescents (Greenberger et al., 1998; Greenberger, Chen, Tally, & Dong, 2000) and with problem behavior in U.S., Taiwanese, and mainland Chinese youths (Chen, Greenberger, Lester, Dong, & Guo, 1998). Cronbach’s alpha for this scale in the current sample was 0.84.

Depressive symptomatology was assessed by means of the 20-item Center for Epidemiologic Studies Depression Scale (CES-D Scale; Radloff, 1977, 1991). Respondents indicated how frequently in the past month they had experienced each symptom, for example, “felt blue,” “could not get going.” The four-point Likert scale ranged from “*never*” to “*almost every day*.” Cronbach’s alpha for this sample was 0.89. Studies have consistently reported significant, inverse correlations between the CES-D and global self-esteem as measured by the Rosenberg scale. In one study based on a large university sample, the correlation between these measures was  $-0.54$  (Scheier, Carver, & Bridges, 1994). In a recent cross-cultural study of adolescents in eastern and western cultures, both the positive and negative subscales were significantly associated—in opposite directions—with depressive symptoms (Farruggia et al., unpublished manuscript).

Optimism was assessed by Scheier and Carver’s (1985) Life-Orientation Test (LOT). A sample item in this 10-item scale is “In uncertain times, I usually expect the best.” Respondents answered on a six-point Likert scale from *strongly disagree* to *strongly agree*. Scores on this scale correlated at about 0.43 with total self-esteem, as measured by the Rosenberg scale, in a sample of early adolescents (Carvajal, Clair, Nash, & Evans, 1998) and 0.54 in a sample of university students (Scheier et al., 1994). Cronbach’s alpha for the optimism scale in this sample was 0.86.

A five-item measure of life satisfaction (Lucas, Diener, & Suh, 1996) was administered. The seven-point Likert scale ranged from “1” = *strongly disagree* to “7” = *strongly agree*, with “4” = *neither agree nor disagree*. A sample item from this scale is “In most ways my life is close to my ideal.” Lucas et al. (1996) reported correlations between Life Satisfaction and Global Self-Esteem ranging from 0.32 and 0.65 in three studies of university students. Cronbach’s alpha for the life-satisfaction scale was .88 for the present sample.

Respondents also completed the 40-item Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991). The scale consisted of two subscales: self-deceptive enhancement (SDE) and impression management (IM). A sample item for SDE is “I never regret my decisions;” a sample item for IM is “I always obey laws, even if I’m unlikely to get caught.” Respondents indicated their answers on a seven-point Likert scale marked *not true* (scored “1”) to *very true* (score “7”). With items reversed as appropriate, SDE and IM scores were computed according to the author’s instructions: the number of items answered at the extreme end of the scales, i.e. responses of “6” or “7.” Coefficient alpha was 0.73 for the SDE and 0.79 for IM.

In order to evaluate method effects associated with use of both positively worded and negatively worded items, summary scores for the subset of negative items and positive items of the CES-D, optimism scale, and parental warmth scale were created. The negative summary scores (later referred to as “negative subscale”) were comprised of 16 negatively worded items for the CES-D (e.g. “I felt sad”), four negatively worded items for the optimism scale (e.g. “I rarely

count on good things happening to me”), and six negatively worded items for the parental warmth scale (e.g. “I find it hard to please my parents”). The positive summary scores (later referred to as “positive subscale”) were comprised of four positively worded items from the CES-D (e.g. “I was happy”), six positively worded items for the optimism scale (e.g. “I am always optimistic about the future”), and five positively worded items for the parental warmth scale (e.g. “My parents really understand me”). For ease of interpretation, positive items for the CES-D and negative items for the optimism and parental warmth scales were reverse coded.

Demographic information obtained from participants included their sex, age, racial or ethnic self-identification (11 separate ethnic designations, plus a write-in category, with data subsequently grouped into seven categories), and their generational status in the US. Respondents were also asked to report their grades, their biological parents’ marital status, and the highest level of education completed by their mother and father (stepmother/stepfather).

#### 1.4. *Plan of analysis*

Confirmatory factor analysis was conducted using AMOS 4.0 (Arbuckle & Wothke, 1999). Maximum likelihood (ML) method of estimation was used. This method places an assumption of multivariate normality of the data. Mardia’s normalized estimates of multivariate kurtosis for the samples in this study were 28.37, 27.58, and 36.82 for the Original RSES, Revised-negative, and Revised-positive samples, respectively. This indicated a departure from the multivariate normality assumption of the ML estimation method. Consequently, all analyses were confirmed with a Satorra–Bentler chi-square, using EQS 5.7 (Bentler, 1995). The Satorra–Bentler adjustment applies a correction for non-normal data and estimates rescaled chi-square statistic and standard errors (Satorra & Bentler, 1994). Results provided by this method did not produce conclusions different from the ones obtained with ML estimation and are, therefore, not presented in this paper.

Two types of models: (i.e. a two-factor model and a single-factor model) were evaluated for the three versions of the Self-Esteem Scale. Chi-square difference test was used to compare model fit for the two models. To explore the contribution of verbal abilities method effects as have been reported by Marsh (1996; and see earlier), CFA models for the Original RSES were compared across the three generations of immigrants and across two groups of students with high and low grades. The generational status and grade variables were used as a proxy for a verbal ability score of the participants. Next, mean comparisons were performed to evaluate whether scale revision for the Revised-negative and Revised-positive scale versions had resulted in different mean levels of reported self-esteem. Subsequently, construct validity of the revised scale versions was established by examining correlations between the self-esteem scores and validity-related scales. Finally, scale validity was evaluated separately by gender and ethnic groups.

## 2. Results

Using AMOS 4.0 (Arbuckle & Wothke, 1999), a two-factor model was compared to a single-factor model for the three versions of the Self-Esteem Scale. Table 2 shows that the original version of the scale, Original RSES, had a significantly better fit for the two-factor model than for



Table 2  
Summary of the fit statistics of the single-factor and two factor models

Scale version	Two-factor model (d.f. = 34)					Single-factor model (d.f. = 35)					$\Delta\chi^2$
	$\chi^2$	CFI	TLI	GFI	RMSEA	$\chi^2$	CFI	TLI	GFI	RMSEA	
The original RSES	91.34	0.951	0.935	0.928	0.081	241.50	0.823	0.773	0.793	0.152	150.16***
Revised-negative version	140.53	0.923	0.898	0.892	0.114	142.58	0.922	0.900	0.890	0.112	2.05
Revised-positive version	181.80	0.895	0.861	0.857	0.135	182.19	0.895	0.865	0.858	0.133	0.39

the one-factor model, with  $\Delta\chi^2(1) = 150.16$ ,  $P < 0.001$ . The correlation between the factors was 0.69. In contrast, for the two Revised versions, the two-factor model did not fit significantly better than the one-factor model ( $\Delta\chi^2(1) = 2.05$ ,  $P > 0.05$  and  $\Delta\chi^2(1) = 0.31$ ,  $P > 0.05$ , for Revised-negative and Revised-positive versions, respectively). The one-factor model for the two Revised versions had a better fit than that for Original RSES. However, the one-factor model fit was still not ideal for the two Revised versions, as indicated by significant chi-squares and out-of-range fit indices (see Table 2). These results suggest that the Revised-negative and Revised-positive versions might have an alternative factor structure that was not captured by our model specifications.

To further explore the factor structure of the Revised-negative and Revised-positive versions, we performed exploratory factor analysis using principal components factor extraction. The results of this analysis, however, did not shed light on the alternative factor structure of the scales. For both Revised versions, a single factor was extracted, which accounted for 58% (Revised-negative version) and 59% (Revised-positive version) of item variance. In other words, a single-factor solution was accepted by the exploratory factor analysis but was rejected by our prior confirmatory factor analysis.

The CFA two-factor and single-factor models for the Original RSES were compared across the three generations of immigrants (i.e. first-, second-, and third-generations of immigrants) and for students with high ("B" and above) and low grades. Results for these two indirect measures of verbal abilities did not show a consistent pattern. For all three generations, the two-factor model had a significantly better fit than a single-factor model. For the third generation of immigrants, the chi-square difference between the two-factor and single-factor models was much smaller ( $\Delta\chi^2(1) = 32.48$ ) than was the case for the first- and second- generations of immigrants ( $\Delta\chi^2(1) = 62.25$  and  $\Delta\chi^2(1) = 53.86$ , respectively). These results could indicate that third-generation students, who were presumably to higher verbal abilities, showed less of a method effect. However, results for grades led to the opposite conclusion: Students with higher grades showed a greater method effect,  $\Delta\chi^2(1) = 106.47$ , than students with lower grades,  $\Delta\chi^2(1) = 50.67$ .

### 2.1. Mean differences for the versions of the Self-Esteem Scale

Mean scores on the three versions of the Self-Esteem Scale did not differ significantly: for Original RSES,  $M = 4.65$ ,  $S.D. = 0.81$ ; for Revised-negative version,  $M = 4.61$ ,  $S.D. = 0.97$ ; and for Revised-positive version,  $M = 4.76$ ,  $S.D. = 0.73$ ,  $F(2,738) = 1.87$ , ns. That is, item-wording did not influence mean self-esteem scores.

Table 3

Construct validity of the three versions of the Rosenberg Self-Esteem Scale and their relations to socially desirable responding

	Self-esteem		
	The original RSES	Revised-negative version	Revised-positive version
Parental warmth	0.42***	0.40***	0.43***
Depressive symptoms	−0.64***	−0.60***	−0.47***
Optimism	0.61***	0.60***	0.60***
Life satisfaction	0.61***	0.55***	0.54***
Self-deception	0.61***	0.33***	0.48***
Impression management	0.25***	0.23***	0.20**

In calculating total self-esteem scores, negatively worded items were reverse-scored so that higher total scores reflect higher overall self-esteem.

\*\*  $P < 0.01$ .

\*\*\*  $P < 0.001$ .

## 2.2. Construct validity

Table 3 shows the correlations between total self-esteem scores and validity-related measures. As expected, total self-esteem scores derived from all three versions were correlated in the expected direction with parental warmth and acceptance (positively), depressive symptoms (negatively), optimism (positively), and life satisfaction (positively). In addition, self-esteem scores were positively related to self-deception (SDE), and to a lesser extent—perhaps due to the anonymous nature of the survey—impression management (IM). Pair-wise comparisons of the correlation coefficients were conducted, using Fisher's  $r$ -to- $z$  transformations. With two exceptions, the wording of items did not seem to affect the magnitude of correlations between the total self-esteem score and other measures. One exception was a significantly lower correlation between total self-esteem and depressive symptoms when all self-esteem items were worded positively (Revised-positive version  $<$  Original RSES,  $z = 2.76$ ,  $P < 0.01$ ). Second, Original RSES showed a significantly stronger association with SDE than the association between SDE and either the Revised-positive version,  $z = 2.05$ ,  $P < 0.05$ , or the Revised-negative version,  $z = 4.05$ ,  $P < 0.001$ .

To further examine whether the item-wording affects the validity of the self-esteem scale, we correlated the two subscales of self-esteem with the subscales of the validity-relevant variables (see Table 4). As expected, the correlations were all in the same direction and of similar magnitude for the subscales. The only evidence of item-wording effects was found for the correlations between the Original RSES and depressive symptoms: The correlations were higher when the wording was in the same direction (i.e.  $r = -0.56$  between positively worded items of self-esteem and positive items of the CES-D, and  $r = -0.57$  between negatively worded items of self-esteem and negative items of CES-D) than when the wording was in the opposite direction (i.e.  $r = -0.44$  between positively worded items of self-esteem and negative items of CES-D and  $r = -0.48$  between negatively worded items of self-esteem and positive items of CES-D). These differences, however, were not statistically significant. Overall, there was little evidence that the two self-esteem subscales had differential associations with the validity-relevant measures.

Table 4

Correlations of the Negative and Positive Self-Image Scales with Positively- and Negatively-Worded Items of the CES-D, Optimism, and Parental Warmth Scales

	The original RSES		Revised-negative version		Revised-positive version	
	Positive	Negative	Positive	Negative	Positive	Negative
<i>Depressive symptoms</i>						
Positive items	−0.56***	−0.48***	−0.53***	−0.49***	−0.48***	−0.47***
Negative items	−0.44***	−0.57***	−0.51***	−0.53***	−0.43***	−0.36***
<i>Optimism</i>						
Positive items	0.55***	0.46***	0.54***	0.52***	0.56***	0.50***
Negative items	0.42***	0.45***	0.52***	0.52***	0.46***	0.44***
<i>Parental warmth</i>						
Positive items	0.37***	0.32***	0.39***	0.32***	0.38***	0.38***
Negative items	0.33***	0.36***	0.37***	0.35***	0.39***	0.37***

\*\*\*  $P < 0.001$ .

### 2.3. Dimensionality and validity of self-esteem scores, by ethnicity and gender

Multigroup comparisons were used to compare the factor structure of the Self-Esteem Scale for the Asian American and European American groups. The sample size for the rest of the ethnic groups was not large enough to allow a meaningful model fit. The Asian American group was comprised of the East Asians and Southeast Asians. Factor loadings for the two-factor model for Original RSES and a single-factor model for the two Revised versions were compared across the two ethnic groups. For all three versions, there were few ethnic differences in the magnitude of factor loadings [for Original RSES,  $\Delta\chi^2(8) = 12.34$ , n.s.; for Revised-negative version,  $\Delta\chi^2(9) = 18.60$ ,  $P < 0.05$ ; and for Revised-positive version,  $\Delta\chi^2(9) = 10.01$ , n.s.].

Table 5 shows the correlations of self-esteem scores based on the three versions of the Self-Esteem Scale with other variables separately for each of three ethnic groups: European Americans, Latinos, and a group comprised of east and southeast Asian Americans [In addition to Asian and European Americans, we included Latinos, whose sample size ( $n = 25$ – $29$ , depending on the version) was too small for factor analysis but adequate for correlational analysis.] In general, similarities across groups were more notable than differences. Only three out of 54 possible ethnic differences in correlations (3 versions  $\times$  3 groups  $\times$  6 correlates) were significant. Using Fisher's  $r$ -to- $z$  transformations to examine differences between pairs of correlations, we found that the correlation between self-esteem and life satisfaction was lower for Asian Americans than the other two groups on the Revised-positive version, with  $z = 2.65$ ,  $P < 0.01$  for the Asian-European American comparison, and  $z = 2.00$ ,  $P < 0.05$  for the Asian-Latino comparison. On the Original RSES, the correlation of self-esteem and impression management was significantly lower for European Americans than Latinos,  $z = 2.01$ ,  $P < 0.05$ .

Similar analyses were performed to evaluate gender differences in the factor structure and construct validity of the three versions of the Self-Esteem Scale. For all three versions, there were

few significant gender differences in the magnitude of factor loadings,  $\Delta\chi^2(8) = 7.97$ , n.s.,  $\Delta\chi^2(9) = 8.06$ , n.s., and  $\Delta\chi^2(9) = 19.72$ ,  $P < 0.05$ , for Original RSES, Revised-negative, and Revised-positive versions respectively.

Table 6 shows the correlations of the three versions of the Self-Esteem Scale with other variables (i.e. parental warmth, depressive symptoms, optimism, life satisfaction, self-deception, and impression management) by gender. There were no significant gender differences in the degree of association between self-esteem (for all three versions) and other variables.

Table 5

Convergent validity of the three versions of the Rosenberg Self-Esteem Scale and their relations to socially desirable responding by ethnicity<sup>a</sup>

Variable	The original RSES			Revised-negative version			Revised-positive version		
	EA	L	AA	EA	L	AA	EA	L	AA
Parental warmth	0.21	0.53**	0.36***	0.41**	0.29	0.44***	0.48***	0.43*	0.35***
Depressive symptoms	-0.69***	-0.67***	-0.58***	-0.69***	-0.73***	-0.58***	-0.53***	-0.67***	-0.39***
Optimism	0.69***	0.65***	0.57***	0.79***	0.66***	0.59***	0.69***	0.67***	0.52***
Life satisfaction	0.54***	0.62***	0.57***	0.63***	0.52**	0.55***	0.72***	0.71***	0.42***
Self-deception	0.56***	0.61***	0.60***	0.52***	0.22	0.32**	0.33*	0.49**	0.50***
Impression management	0.06	0.52***	0.23*	0.33*	-0.07	0.23*	0.19	0.08	0.13

EA, European Americans; L, Latinos; AA, Asian Americans.

<sup>a</sup> Sample sizes for each version of the self-esteem scale ranged from 45 to 51 for EA; 23–29 for L; and 115–131 for AA.

\*  $P < 0.05$ .

\*\*  $P < 0.01$ .

\*\*\*  $P < 0.001$ .

Table 6

Construct validity of the three versions of the Rosenberg Self-Esteem Scale and their relations to socially desirable responding by gender<sup>a</sup>

Variable	The original RSES		Revised-negative version		Revised-positive version	
	Males	Females	Males	Females	Males	Females
Parental warmth	0.54***	0.37***	0.47***	0.38***	0.51***	0.38***
Depressive symptoms	-0.70***	-0.62***	-0.70***	-0.54***	-0.51***	-0.44***
Optimism	0.59***	0.61***	0.64***	0.59***	0.54***	0.64***
Life satisfaction	0.64***	0.60***	0.58***	0.56***	0.50***	0.57***
Self-deception	0.60***	0.60***	0.16	0.43***	0.39***	0.53***
Impression management	0.38***	0.19*	0.43***	0.26**	0.22*	0.21*

<sup>a</sup> Sample sizes for each version of the self-esteem scale ranged from 84 to 99 for males and from 138 to 163 for females.

\*  $P < 0.05$ .

\*\*  $P < 0.01$ .

\*\*\*  $P < 0.001$ .

### 3. Discussion

The results of this study indicate that the two-factor structure of the Rosenberg Self-Esteem Scale found by some researchers is an artifact of the two types of item-wording (positive and negative) used in that scale. Once the wording of the scale is altered, so that all items are written in a consistent direction, it is a single-factor scale. One might argue that re-wording of the scale items (i.e. changing positively worded items into negatively worded items and vice versa) essentially changes the construct being measured by these items. In other words, the Revised-negative version may measure only “negative self-image,” whereas the Revised-positive version may measure only “positive self-image.” However, given the extreme similarity of the two Revised versions with respect to their relations with other variables (i.e. construct validity), one is hard-pressed to argue that the positively worded and the negatively worded versions are tapping into different dimensions of self-esteem.

Not only was the validity of the two Revised versions similar, but their validity was also generally similar to that of Original RSES. This finding is not particularly surprising because the two factors of the original Rosenberg Scale are highly correlated ( $r=0.69$  in this sample). Nonetheless, two exceptions were uncovered that may have important implications. These exceptions concern the magnitude of associations between self-esteem scores and measures of depressive symptoms and socially desirable responding. Researchers are sometimes uneasy about treating self-esteem and depression as separate constructs, because of overlap in the attributes that characterize low self-esteem and high depression and overlap among the items that are commonly used to assess these constructs. A measure of self-esteem that minimizes or reduces the degree of overlap with a measure of depression would be desirable for researchers and practitioners alike. Data from this study suggest that a Revised-positive version of the Rosenberg Scale results in the lowest correlation with the CES-D. The “superiority” of Revised-positive version in this regard may be due to its having less shared method variance with the CES-D, a scale in which 16 of 20 items are worded *negatively*. Given the similarity in several items for the two scales, however, some overlap is inevitable. For example, one pair of similar CES-D and Rosenberg Self-Esteem items, respectively, is “I felt that I was just as good as other people” and “I feel that I am a person of worth, at least on an equal plane (basis) with others.” Another similar pair of items is “I thought my life had been a failure” and “All in all, I am inclined to feel that I am a failure.”

Researchers may also wish to use a measure of self-esteem that minimizes socially desirable responding. The original Rosenberg Scale yielded a higher correlation with a measure of desirable responding than the revised versions of the scale, especially the Revised-negative version.

To summarize, all three versions of the Rosenberg Self-Esteem Scale have high construct validity and did not differ substantially in their dimensionality and validity as a function of gender or ethnicity. Results of this study suggest that there is no absolute superiority of any one version of the scale over the others. The original scale, which was counter-balanced with positively and negatively worded items, was not different from the uni-directional versions in terms of the scale means and validity. In other words, the presumed benefit of counter-balanced wording was not evident. On the other hand, the similar validity of the three versions also suggests that the uni-directional versions which generated “single-factor” models showed no improvement in the scale’s validity.

This study had several limitations. It focused exclusively on a sample of college students. Thus, findings may not generalize to other populations. Results of our study failed to show a consistent

pattern of method effects for respondents with different levels of verbal ability. Our use of immigration generation status and grades as a measure of verbal abilities may well be the reason for this lack of meaningful findings. Future research using more clear-cut measures of verbal skills may be able to shed more light on this issue. Future studies should also examine the convergent validity of the three versions of the Rosenberg Self-Esteem Scale, for example, their association with other measures of self-esteem.

Does the “real” structure of the Rosenberg Scale matter? We suggest that it does, insofar as researchers who focus on separate measures of negative and positive self-image based on the Rosenberg scale may be drawing attention to distinctions that are spurious. Additionally, previous advice to program planners and practitioners that they focus on *positive* self-image if they are attempting to maximize certain outcomes, and on *negative* self-image in order to optimize other outcomes (Owens, 1994), may be misplaced. Future efforts, instead, should focus on refining the Rosenberg Self-Esteem Scale by further investigating and hopefully eliminating some of the measurement errors to create a uni-dimensional scale of self-esteem as Rosenberg originally intended.

## Acknowledgements

We thank a dedicated group of undergraduates for assistance with data collection and data entry.

## References

- Arbuckle, J. L., & Wothke, W. (1999). *AMOS users' guide (Version 4.0)*. [Computer software manual]. Chicago: Small Waters Corporation.
- Bachman, J., & O'Malley, P. (1986). Self-concepts, self-esteem, and educational experiences: the frog pond revisited (again). *Journal of Personality and Social Psychology*, *50*, 35–46.
- Bentler, P. M. (1995). *EQS program manual*. Encino, CA: Multivariate Software.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Carvajal, S., Clair, S. D., Nash, S. G., & Evans, R. I. (1998). Relating optimism, hope and self-esteem to social influences in deterring substance use in adolescents. *Journal of Social and Clinical Psychology*, *17*, 443–465.
- Chen, C., Greenberger, E., Lester, J., Dong, Q., & Guo, M. (1998). A cross-cultural study of family and peer correlates of adolescent misconduct. *Developmental Psychology*, *34*, 770–781.
- Corwyn, R. F. (2000). The factor structure of global self-esteem among adolescents and adults. *Journal of Research in Personality*, *34*, 357–379.
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, *60*, 151–174.
- Farruggia, S. P., Chen, C., Greenberger, E., Dmitrieva, J., Tally, S. R., & Macek, P. Self-esteem in cross-cultural perspective: Measurement equivalence and the mediator role of self-esteem between parental warmth and adolescent depressed mood (unpublished manuscript).
- Goldsmith, R. E. (1986). Dimensionality of the Rosenberg Self-Esteem Scale. *Journal of Social Behavior and Personality*, *1*, 253–264.
- Gray-Little, B., & Hafdahl, A. (2000). Factors influencing racial comparisons of self-esteem: a quantitative review. *Psychological Bulletin*, *126*, 26–54.
- Greenberger, E., Chen, C., & Beam, M. (1998). The role of “very important” nonparental adults in adolescent development. *Journal of Youth and Adolescence*, *27*, 321–343.

- Greenberger, E., Chen, C., Tally, S. R., & Dong, Q. (2000). Family, peer, and individual correlates of depressive symptomatology among US and Chinese adolescents. *Journal of Consulting and Clinical Psychology, 68*, 209–219.
- Harter, S. (1990). Cause, correlates, and the functional role of global self-worth: a life-span perspective. In J. Kolligan, & R. J. Sternberg (Eds.), *Competence considered* (pp. 67–97). New Haven, CT: Yale University Press.
- Hensley, W. E., & Roberts, M. K. (1976). Dimensions of Rosenberg's Self-Esteem Scale. *Psychological Reports, 38*, 583–584.
- Kaplan, H. B., & Pokorny, A. D. (1969). Self-derogation and psychosocial adjustment. *Journal of Nervous and Mental Disease, 149*, 421–434.
- Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71*, 616–628.
- Marsh, H. (1996). Positive and negative global self-esteem: substantively meaningful distinction or artifacts. *Journal of Personality and Social Psychology, 70*, 810–819.
- Owens, T. J. (1993). Accentuate the positive-and the negative: rethinking the use of self-esteem, self-deprecation, and self-confidence. *Social Psychology Quarterly, 56*, 288–299.
- 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.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Radloff, L. S. (1977). The CES-D Scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
- Radloff, L. S. (1991). The use of the Center for Epidemiologic Studies Depression scale among adolescents and young adults. *Journal of Youth and Adolescence, 20*, 149–166.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Satorra, A., & Bentler, P. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. Clogg (Eds.), *Latent variables analysis* (pp. 399–441). Thousand Oaks, CA: Sage.
- Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health Psychology, 4*, 219–247.
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*, 1063–1078.
- Sheasby, J. E., Barlow, J. H., Cullen, L. A., & Wright, C. C. (2000). Psychometric properties of the Rosenberg Self-Esteem Scale among people with arthritis. *Psychological Reports, 86*, 1139–1146.
- Tomas, J. M., & Oliver, A. (1999). Rosenberg's Self-Esteem Scale: Two factors or method effects. *Structural Equation Modeling, 6*, 84–98.