

Factor Structure of the Rosenberg Self-Esteem Scale

By: [Andrew J. Supple](#), Jinni Su, Scott W. Plunkett, Gary W. Peterson, Kevin R. Bush

Supple, A.J., Su, J., Plunkett, S.W., Peterson, G.W., & Bush, K.R. (2013) Factor Structure of the Rosenberg Self-Esteem Scale. *Journal of Cross-Cultural Psychology*, 44(5), 748-764.

Made available courtesy of Sage Publications & the International Association for Cross-Cultural Psychology: <http://jcc.sagepub.com/content/44/5/748>

*****Reprinted with permission. No further reproduction is authorized without written permission from Sage Publications. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. *****

Abstract:

This study evaluated the factor structure of the Rosenberg Self-Esteem Scale (RSES) with a diverse sample of 1,248 European American, Latino, Armenian, and Iranian adolescents. Adolescents completed the 10-item RSES during school as part of a larger study on parental influences and academic outcomes. Findings suggested that method effects in the RSES are more strongly associated with negatively worded items across three diverse groups but also more pronounced among ethnic minority adolescents. Findings also suggested that accounting for method effects is necessary to avoid biased conclusions regarding cultural differences in self-esteem and how predictors are related to the RSES. Moreover, the two RSES factors (positive self-esteem and negative self-esteem) were differentially predicted by parenting behaviors and academic motivation. Substantive and methodological implications of these findings for cross-cultural research on adolescent self-esteem are discussed.

Keywords: bidimensional | Rosenberg | self-esteem | two-factor | construct validity | cross-cultural validity | Adolescents

Article:

The purpose of this study is to evaluate the factor structure and cross-cultural validity of the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1989). Numerous studies employ the RSES to examine the etiology of self-esteem and how self-esteem relates to a variety of mental health, academic, or social outcomes (Farruggia, Chen, Greenberger, Demitrieva, & Macek, 2004). Despite evidence from factor analysis studies suggesting that the RSES is bidimensional (i.e., having two factors), researchers continue to treat the RSES as a unidimensional scale. A current debate centers on whether or not bidimensional findings demonstrate that the RSES assesses two substantively distinct elements related to self-esteem or result from method effects that occur due to the inclusion of positively and negatively worded items in the same scale. Furthermore, while the RSES is commonly used in studies of adolescents from diverse cultural and ethnic groups, few studies have evaluated the cross-cultural construct validity of the RSES.

For cross-cultural researchers, the issue of possible method effects-particularly those that vary across cultural groups-has the potential to create bias when assessing self-esteem. That is, some cross-cultural comparisons of mean levels of self-esteem may be biased if the RSES produces different scores across groups due to method effects and not due to culturally meaningful differences in self-esteem (van de Vijer & Tanzer, 2004). Moreover, to the extent that two RSES-related subdimensions may have substantive meaning across cultures, the use of the full RSES unidimensional scale may lead to concluding that differences exist in predictors of self-esteem across cultures that are due to measurement artifacts and not cultural differences.

Factor Structure and Dimensionality of the RSES

A number of factor analysis studies have suggested that, rather than simply assessing a unidimensional positive self-evaluation construct, the RSES may actually have two factors. One factor is proposed to assess positive self-esteem (comprised of the five positively worded RSES items, e.g., “*On the whole, I am satisfied with myself*”) with another assessing negative self-esteem (comprised of the five negatively worded items, e.g., “*I certainly feel useless at times*”). Researchers also find differential patterns of association between positive self-esteem and theoretically related constructs when compared to similar associations with negative self-esteem (Ang, Neubronner, Oh, & Leong, 2006; Farruggia et al., 2004; Owens, 1993, 1994). For example, Owens (1993, 1994) found that positive self-esteem was significantly associated with grades and delinquent behaviors but unrelated to depressive symptoms. Negative self-esteem, on the other hand, was positively related to depressive symptoms and delinquent behaviors but not grades.

While Owens (1993, 1994) argued that such findings revealed two RSES developmental constructs, other researchers argue that bidimensionality in the RSES simply points to ignorable method effects (Carmin & Zeller, 1979; Corwyn, 2000; Marsh, 1996; Tomás & Oliver, 1999), which often result when positively and negatively worded items appear in the same scale. For certain respondents, the inclusion of negatively worded items seems to induce a response set (see Quilty, Oakman, & Risko, 2006, for a review) who differentially respond to such items and in a manner that is unrelated to item content. As such, studies finding two RSES subdimensions may not be revealing a substantively meaningful RSES subdimension (related to negative self-esteem) but rather method effects that should either be ignored or can be controlled using advanced quantitative analytical techniques.

The issue of bidimensionality in the RSES has significant implications for researchers who study adolescent self-esteem generally and for cross-cultural research on self-esteem specifically. A major substantive implication relates to the suggestion that the RSES assesses two related yet distinct aspects of development related to self-esteem (Ang et al., 2006; Owens, 1993, 1994). Individuals may differentially score high or low on assessments of their positive self-evaluations as well as in assessments of not liking themselves or feeling inferior to others. Moreover, the associations that either positive or negative self-esteem have with other elements of development might vary and thus present different opportunities for research. For example, while self-deprecating behaviors among adolescents might be more strongly related to the development of depressive symptoms, typical conceptions of self-esteem (as a positive self-evaluation) might not be (Owens, 1993, 1994), and such associations could vary across cultures.

The methodological implications of bidimensionality in the RSES center on how researchers choose to address and understand how method effects may be operating in their own data to avoid introducing bias when assessing self-esteem. Confirmatory factor analysis (CFA) can account for method effects in SEM analyses (see below), whereas little guidance is available for researchers using ANOVA or OLS regression when summing or averaging across RSES items. Researchers may inadvertently introduce error into their measures when treating the RSES as unidimensional and not accounting for method effects as certain respondents may be scored as lower in terms of self-esteem when what differentiates them from other respondents is related to some other factor (i.e., a response style). Perhaps most troubling, for valid comparisons of self-esteem across cultures, is the possibility that these method effects might vary across cultural groups and the failure to account for these effects could bias conclusions, such as falsely concluding that one cultural group has higher self-esteem than another, when what truly differentiates the two groups could be a response set or differences in two related yet distinct dimensions of self-esteem being assessed by the RSES.

Assessing Method Effects in the RSES

The most common strategy for assessing method effects in the RSES involves comparing alternative CFA models. The seminal study was conducted by Marsh (1996), who used the 1988 National Educational Longitudinal Study to compare the fit of a bidimensional RSES factor model to several alternative unidimensional models using four positively worded and three negatively worded RSES-based items. The general approach, in sum, compares (a) a unidimensional RSES model, (b) a bidimensional RSES model, and (c) unidimensional models that either contain a method latent factor or models that involve correlating residual uniquenesses (error terms) among either positively or negatively worded items (and in some cases, both). To the extent that specifying either a method effect factor or correlated uniquenesses among items (e.g., the negatively worded items) results in a well-fitting unidimensional model, there is evidence that unidimensional models will fit if method effects are accounted for (and thus method effects are related to item wording and not indicative of two RSES factors). Marsh concluded that while a bidimensional model fit better than the baseline unidimensional model with no correlated uniquenesses, the best fit was obtained for a unidimensional model with correlated uniqueness among all three negative items and between two of the four positive items (see Figure 1). In a more recent study, Marsh, Scalas, and Nagengast (2010) replicated this approach and added additional models assessing fit when method effects were specified as latent factors (rather than correlated uniquenesses) using longitudinal data to find that method effects are highly stable among respondents over time.

While Marsh's studies did not employ the actual RSES, other studies have adopted a comparable approach to find method effects when using the RSES. For example, Tomas and Oliver (1999) found that correlated uniquenesses among negatively worded items yielded an acceptable fit to their data (using a Spanish version of the RSES with high school students). Corwyn (2000) also found method effects most prominently associated with negatively worded RSES items with U.S. adults and adolescents. On the other hand, Aluja, Rolland, Garcia, and Rossier (2007) found that the best fit among French college students (predominantly female) was a one-factor RSES model with correlated errors among three of the positive items and for two of the negative items.

Finally, Ang et al. (2006), in addition to reporting superior fit of a bidimensional RSES model compared to a unidimensional model (seventh-grade students from Singapore) also found differential associations between positive versus negative self-esteem (positive self-esteem was associated with academic self-efficacy, while negative self-esteem was related to deviant behaviors).

Despite ample evidence regarding method effects in the RSES, little is known regarding what consequences exist for research conclusions (e.g., how associations are attenuated or biased by combining together conceptually distinct measures into the same construct) when studies employ a unidimensional RSES scale versus two RSES subdimensions. As a result, the extent to which two RSES subdimensions might have theoretically meaningful or substantively different associations with a number of other constructs is not well understood. A recent study by Quilty et al. (2006) reported associations between method effects associated with negatively worded items (by specifying a latent method effects factor) and avoidance motivation, conscientiousness, and emotional stability. Such findings suggest that individual response tendencies are related to personality dimensions and an alternative RSES factor has substantive meaning because of its association with personality.

[Figure 1 Omitted]

Overall, CFA models have been helpful in identifying method effects in the RSES and hold promise in identifying covariates associated with method effects to better understand respondent characteristics associated with biased self-esteem scores. Relative to cross-cultural research, these approaches can be used to consider cross-cultural variation or similarities in RSES method effects and, perhaps more importantly, whether or not cultural differences in RSES scores are biased due to method effects.

Cross-Cultural Validity of the RSES

Despite its wide use with adolescents from diverse cultural groups, few studies have considered to what extent method effects in the RSES might vary across cultural groups. Within-culture studies (reviewed above) suggest that bidimensionality tends to be found across cultures as do method effects predominantly associated with negatively worded RSES items. Other studies have focused on measurement invariance analyses (are factor loadings and item intercepts comparable across cultural groups) with limited research comparing factor structures with diverse samples in the same study. One cross-cultural study comparing 11th-grade students from the United States, Czech Republic, China, and Korea concluded that a bidimensional RSES model was a better fit than a unidimensional model across all of the samples (Farruggia et al., 2004). To achieve a well-fitting bidimensional model, however, these authors deleted a noninvariant item (due to low factor loadings in the non-U.S. samples; “*I wish I had more respect for myself*”) and added correlated uniquenesses. Using this final model, these authors found that two RSES subdimensions (i.e., positive or negative self-esteem) varied in their associations with depression, particularly cross-culturally. In the Czech and Chinese samples, negative self-esteem was positively related to depressed mood, but positive self-esteem was unrelated. In the U.S. and Korean samples, on the other hand, both positive self-esteem (a

negative association) and negative self-esteem (a positive association) were significantly related to depression.

Other studies focusing on cross-cultural measurement equivalence of RSES items with samples of adults (Schmitt & Allik, 2005, across 53 countries; Greenberger, Chen, Dmitrieva, & Farruggia, 2003, comparing Asian American and European American college students) reported similar item loadings across cultural groups. Moreover, Schmitt and Allik (2005) found a consistent factor structure, factor loadings (except for the *wish I had more respect* item), and associations with personality indicators across 53 countries. This study also found a similar pattern of association between two RSES subdimensions and indicators of personality, but negative item biases (responding differently to the negatively worded items) and lower reliability in the RSES were more common among respondents from collectivistic cultures and developing nations. The only study to consider ethnic variation in self-esteem items across diverse samples of adolescents within the United States (Russell, Crockett, Shen, & Lee, 2008) used the Add Health data set to demonstrate measurement invariance across European American, Filipino American, and Chinese American adolescents on self-esteem items (not the RSES items).

Summary and Goals of This Study

Despite its wide use in studies of adolescent self-esteem across diverse cultural and national settings, there is little research regarding the cross-cultural validity of the RSES. More specifically, few studies have addressed the extent to which possible method effects in the RSES are evident in cross-cultural studies, whether or not differential patterns of association exist across two RSES subdimensions, or what implications there are for cross-cultural research when employing alternative specifications of the RSES. To address these issues, the current study examined (a) the factor structure and bidimensionality of the RSES; (b) possible method effects associated with item wording as an alternative explanation to bidimensionality; (c) patterns of association among two RSES subdimensions with age, gender, two parenting behaviors (i.e., parental support and psychological control), and an academic outcome (i.e., academic motivation); and (d) how alternative specifications of the RSES may alter conclusions regarding predictors and self-esteem. We also consider *a posteriori* models to explore whether relatively modest modifications to either the unidimensional or bidimensional model will result in good model fit and consider all these questions using three ethnically diverse groups of adolescents.

Method

Participants

The sample for this study consisted of 1,248 high school students (Grades 9 through 12) residing in a large metropolitan area in Southern California. Adolescents were selected from a larger sample for inclusion if they self-identified as Mexican, Guatemalan, Salvadoran, Armenian, Iranian, or European American (hyphenated identities such as Mexican American also were included). Other immigrant ethnic or racial participants were excluded due to small sample size. Mexican, Guatemalan, and Salvadoran adolescents were combined into a group labeled *Latinos*, whereas Armenians (or Armenian Americans) and Iranian (or Persians) were also combined into a larger group. These combinations were justified given cultural similarities among these groups

and due to preliminary analyses suggesting similarity in terms of configural measurement invariance (details available from the first author).

Of the 790 Latino adolescents in the study, 73% were Mexican origin, 18% were Salvadoran, and 9% were Guatemalan. The majority (66%) of respondents were born in the United States, but few (9% and 7%, respectively) had mothers or fathers who were U.S. born. The average age was 15, and 59% were female. The Armenian/Iranian sample included 334 adolescents, 59% were female, and with an average age of 15. Less than half of these participants (46%) were born in the United States (28% were born in Armenia and 17% in Iran, with the remainder born elsewhere in the Middle East), and only a few (3%) indicated that either parent was born in the United States. The subsample of 124 European Americans was 67% female, with an average age of 15.5. All of the European American adolescents were born in the United States, with the vast majority having parents who also were born in the United States (95% and 90% of mothers and fathers, respectively).

Procedures

School administrators from four schools agreed to allow student participation. Teachers distributed and collected parental consent forms in three of the schools. Approximately 1 to 2 weeks later, research assistants visited schools, explained the project, and administered assent forms and questionnaires during the class time of a required course. Researchers in the fourth school explained the project and gave consent forms to students, which they were asked to take home for their parent's voluntary consent signature. Students who returned signed parental consent forms and then signed assent forms themselves were given questionnaires to complete on their own time as part of a school-based service credit.

Measures

The following measures were included in a larger survey on parental influences on academic outcomes and psychological well-being across diverse groups of adolescents. As such, the available constructs to link to self-esteem were centered on academic outcomes and adolescent reports of parenting behaviors (see below). In preliminary analyses (available from the first author), the measurement equivalence of these measures was established across the three groups of adolescents, suggesting that any differential associations between the selected predictors and the RSES would not be biased across groups based on differences in item parameters (Sass, 2011; Steenkamp & Baumgartner, 1998).

Self-esteem. The RSES includes 10 Likert-type scale, self-report items designed to assess positive evaluations of the self. Response options range from *strongly disagree* (1) to *strongly agree* (4). Five items indicate greater positive self-esteem (e.g., “*I feel that I have a number of good qualities*”), and five items indicate greater negative self-esteem (e.g., “*I certainly feel useless at times*”). Cronbach's alphas for the full scale in these data were .79, .82, and .86 for the Latino, Armenian/Iranian, and European American samples, respectively. Consistent with conventional use of the RSES, negatively worded items were reverse coded so that higher scores on this RSES subdimension would actually indicate *lower* negative self-esteem.

Academic motivation. Four items assessed adolescents' effort in school, the importance of grades and education, completing homework on time, and liking school (Plunkett & Bamaca-Gomez, 2003). Response options ranged from *strongly disagree* (1) to *strongly agree* (4), with Cronbach's alphas for this scale ranging (across groups) from .76 to .84. Items were averaged to create a summary variable, with higher scores indicating greater academic motivation.

Maternal behaviors. The Parental Behavior Measure (Bush, Peterson, Cobas, & Supple, 2002) was used to assess youth reports of maternal support and psychological control. Mothers' support was assessed with four items (e.g., warmth, encouragement, praise), whereas psychological control was assessed by six items representing intrusive and manipulative behavior. Sample items included "*This parent seems to approve of me and the things I do*" (support) and "*This parent tells me if I love her, I would do what she wants me to do*" (psychological control). Cronbach's alphas for these variables in the current samples ranged from .78 to .90. Items were averaged to create summary variables, with higher scores indicative of greater perceived maternal support and psychological control.

Results

All analyses reported below were conducted using Mplus version 6 and the weighted least squares estimator. This estimator is appropriate for CFA analyses with ordinal items, like those used in the RSES (Muthen, 1984). The first set of analyses addressed possible method effects in the RSES and whether method effects were similar or different across samples. Following the procedure outlined by Marsh (1996) and Marsh et al. (2010), we evaluated the relative fit of a series of CFA models that varied in how RSES items were specified. The central question examined in these analyses is whether or not unidimensional models can fit when method effects are taken into account and whether or not a bidimensional model provides a better fit than unidimensional models. According to this approach, the first two possible solutions consist of a unidimensional model (Model 1, see Figure 1) with all 10 RSES items loading on a single self-esteem factor and a bidimensional model (Model 2) specifying the five negatively worded RSES items as loading onto one factor and the five positively worded RSES items loading onto a second factor. A significant improvement in the fit of Model 2 over Model 1 would suggest that the RSES is better represented by a bidimensional factor structure.

Subsequent models assessed the fit of a unidimensional model when correlated uniquenesses are included in the model for the negatively and positively worded items (Models 4 and 5, respectively). The extent to which either of these unidimensional models provides the best fit to the data would provide evidence that bidimensionality is due largely to method effects. We also considered models proposed by Marsh and others that included correlated uniquenesses among both positively and negatively worded items in the same model (Model 3) and models where, rather than specifying method effects via correlated uniquenesses, method effects are specified as a latent factor representing shared covariation among RSES items beyond what is explained by the target RSES factor (see Models 6, 7, and 8 in Figure 1). A comparison of model fit allows for an empirical assessment of which alternative specification best fits the data and to what extent unidimensional RSES models fit once method effects are taken into account.

Unfortunately, some of the models proposed by Marsh (Model 3 and Model 6) are often difficult to evaluate given convergence problems in their estimation. Such was the case in the present analyses, and as a result, Models 3 and 6 could not be specifically evaluated. In lieu of specifying Model 3 and consistent with Marsh (1996), we used modification indices provided in the Mplus output to add in correlated uniquenesses that would lead to a statistically significant improvement in model fit until no more such modifications could be made. This model (Model 9) is consistent with the intent of Marsh's Model 3 by specifying correlated uniquenesses among both positively and negatively worded RSES items. In addition, we also included an *a posteriori* bidimensional model that included suggested correlated uniquenesses (Model 10, see Table 1). This latter model offers a more parsimonious approach to improving model fit (e.g., adding two correlated errors rather than 5 to 10) and does not include specifying nonsignificant correlations among item uniquenesses.

Acceptable model fit is generally established via nonsignificant χ^2 values; however, in large samples relatively minor misspecifications lead to overrejection of relatively well-fitting models (Kline, 2011). Additional fit indices included the Comparative Fit Index (CFI; acceptable > .90, good fit > .95), the Root Mean Square Error of Approximation (RMSEA; acceptable < .08, good fit < .05), and the Tucker-Lewis Index (TLI; acceptable > .90, good fit > .95).

Results Comparing Alternative CFA Models to Assess Dimensionality and Method Effects

Consistent with previous research, these results (see Table 1) suggested that in each of the three ethnic groups, Model 1 was a poor fit to the data, with Model 2 demonstrating a vastly superior fit. Moreover, Model 2 demonstrated a marginally acceptable fit across the three samples based on CFI and TLI values that were above .90. Model 3 was not possible to evaluate as analyses for this model failed to converge when specifying maximal correlated uniquenesses across all items. Model 4, which specified correlated uniquenesses among the negatively worded items, demonstrated a good fit in all the samples based on the CFI and TLI fit indexes and a marginal to good fit based on the RMSEA. Moreover, Model 4 demonstrated a better fit than did Model 5, which specified correlated uniquenesses among the positively worded items. The *a posteriori* model (Model 9) was specified using modification indices resulting from Model 1 and by adding correlations one by one based on the size of suggested modifications from the Mplus output. This process continued in each group until there were no more significant modifications to make. Across groups the largest suggested modifications involved correlating uniquenesses among negatively worded items. In the Armenian/Iranian and European American groups, *all* suggested modifications were centered on negatively worded items (for the Armenian/Iranian group, Model 9 is identical to Model 5), whereas in the Latino sample in addition to correlated uniquenesses among all negatively worded items (suggested first), three additional correlated uniquenesses among positively worded items were also suggested. Taken together, these results suggest, consistent with previous studies, that modifications to a unidimensional RSES model were indicative primarily of method effects associated with negatively worded items, and a model with all uniquenesses correlated was not necessary to achieve maximal fit.

The next set of models specified a latent method factor rather than accounting for method effects via incorporating correlated uniquenesses. These models involve specifying all 10 RSES items as loading onto a target self-esteem factor and then the negatively (in the case of Model 7) or

positively (in the case of Model 8) worded items as loading onto a latent method factor. It is also sometimes possible to consider two method latent factors, but in these data, this model could not be evaluated in all groups as it resulted in a negative error variance in the Latino sample. The advantage of these models over the correlated uniquenesses models includes the ability (in subsequent analyses) to link method effects to possible covariates as a means to understand how these effects might vary across respondents (Marsh et al., 2010; Quilty et al., 2006). The Models 7 and 8 (specifying a negative items method factor and a positive items method factor, respectively) demonstrated adequate fit in each sample, although the fit of these models was not as good compared to comparable models specifying correlated uniquenesses. Moreover, specifying method factors in RSES models generally led to undesirable changes in factor loadings on the RSES factor. For example, specifying a negative items method factor in Model 7 led to decrement in factor loadings of negative worded items on the RSES factor (some factor loadings become lower than .30). Similarly, specifying a positive items method factor in Model 8 led to lower factor loadings of positive worded items on the RSES factor.

[Table 1 Omitted]

The last model considered, Model 10, specified a bidimensional RSES model and achieved a good fit across all of the samples after adding additional correlated uniquenesses to the model (only among negatively worded items). Overall, the findings from this first set of analyses suggested that a bidimensional factor structure was superior, with the best fitting model being those that account for method effects among the negatively worded items. In the two ethnic minority samples, the best fitting models either accounted for method effects primarily in the negatively worded items or specified the RSES as bidimensional with two correlated uniquenesses. In sum, across the three samples, evidence suggested that method effects exist across these diverse samples of adolescents and that method effects were predominantly associated with negatively worded items.

The next set of analyses evaluated what consequences there are for predicting self-esteem based on alternative specifications of the RSES. For example, are there differences in associations between a predictor and self-esteem when method effects are controlled (i.e., Model 7) compared to when method effects are not controlled (i.e., Model 1). We also considered how age, gender (boys coded 1, girls 0), parenting, and academic motivation may be related to methods effects (using Model 7) and also how these factors might differentially predict two RSES factors (Model 10). These analyses were conducted using the entire sample, to consider how a grouping variable (European Americans as the reference group) would be associated with both the RSES and method factors and also via multigroup analyses to evaluate coefficients within each group and to test for statistical differences across groups in coefficients. Predictors of method effects may indicate a response style among certain respondents that show that method effects vary across respondents (Quilty et al., 2006), while an examination of predictors of two RSES subdimensions addresses the question of whether or not the RSES has two substantively meaningful factors that are differentially predicted by typical factors previously found to associate with self-esteem (Owens, 1993, 1994).

Given that measurement nonequivalence can lead to biased associations when comparing regression coefficients across cultural groups, we first evaluated whether the RSES items

demonstrated measurement equivalence (also called invariance) across the three groups in this study (Sass, 2011; Steenkamp & Baumgartner, 1998). Multigroup CFAs were conducted to evaluate equivalence/invariance of item thresholds and factor loadings across the three groups. Invariance testing generally involves comparing the change in χ^2 across two models with one model specifying item parameters (factor loadings and intercept/thresholds) as freely estimated across groups and a subsequent model where these same parameters are constrained to equality across groups. If the addition of parameter equality constraints results in a decrement in model fit (the change in χ^2 per the change in df across the models), then researchers conclude that invariance does not hold and there is not measurement equivalence. Invariance testing in Mplus with ordinal indicators involves the comparison of a default model where both factor loadings and item thresholds are assumed equal across groups (which would indicate what researchers term *strong measurement invariance*; Steenkamp & Baumgartner, 1998) to a model where both loadings and thresholds are set to be freely estimated (threshold and loadings are recommended considered jointly). The DIFFTEST command in Mplus provides the chi-square difference test across freely estimated and constrained models (Muthen & Muthen, 1998-2010).

Measurement invariance analyses were considered for Models 1, 7, and 10 across the Latino, European American, and Armenian/Iranian samples. Results revealed significant differences in χ^2 between the constrained and free models, suggesting significant variations in factor loadings and item thresholds across ethnic groups (strong invariance did not hold for these models across ethnic groups). Modification indices suggested that the source of nonequivalence was primarily centered on the negatively worded item, “*I wish I could have more respect for myself*” (respect). Freeing the constraints on this item’s parameters resulted in a nonsignificant difference test for Models 7 and 10, suggesting partial strong invariance (Steenkamp & Baumgartner, 1998). For Model 1, however, there were two additional items that contributed to nonequivalence (“*I am inclined to feel that I am a failure*” and “*I take a positive attitude toward myself*”). These analyses suggested that the RSES alternative specifications demonstrate partial strong invariance (mostly invariant factor loadings and item thresholds across groups), which is acceptable for group comparisons in structural parameters like regression coefficients (Steenkamp & Baumgartner, 1998). Subsequent analyses including group comparisons were based on models where all factor loadings and item thresholds were constrained to equality across groups except for the relevant parameters previously found to be noninvariant.

Predictors of Alternative Specifications of the RSES, Method Effects, and Positive Versus Negative Self-Esteem

Using the overall sample and predicting self-esteem using Model 1 (see Table 2), there was a significant group coefficient suggesting that, relative to the European Americans, the Armenian/Iranian group had higher self-esteem with no difference between Latinos and the European American group. Findings also suggested that being a boy, maternal support, and academic motivation were positively associated with self-esteem, while psychological control was negatively related. Subsequent multigroup analyses suggested that two associations varied across cultural groups, however. Modification indices resulting from a model that constrained coefficients to equality across the three groups suggested significant variation that suggested age and academic motivation were positively related to self-esteem in the Latino sample only.

Predictors linked to Model 7 examined how gender, age, maternal support, psychological control, and academic motivation would predict a 10-item RSES factor with method effects taken into account and also how these same predictors may be linked to a negative items method effects latent factor. The coefficient previously suggesting (using Model 1) a cultural difference in self-esteem was not significant when including the negative items method factor, suggesting that, after accounting for method effects, the group difference in self-esteem is no longer significant. Findings also suggested, again, that boys and those reporting greater maternal support and academic motivation tended to score higher on the self-esteem factor, while psychological control was negatively related. The coefficient linking academic motivation with self-esteem, however, increased in size after controlling for method effects associated with the negatively worded items, and the association between psychological control and self-esteem was weaker. Significant coefficients also suggested that age, maternal support, and being ethnic minority (either in the Latino or Armenian/Iranian group) were positively associated with the negative items method factor, while academic motivation and psychological control were negatively related. Multigroup analyses with coefficients in Model 7 constrained to equality across groups resulted in an acceptable model fit ($\chi^2 = 676.85$, $df = 284$, CFI = .94, TLI = .94, RMSEA = .06), however modification indices suggested that paths linking age to the RSES factor and academic motivation to both the RSES and method effect factors varied across groups. Once those modifications were added, the model demonstrated a slightly improved fit ($\chi^2 = 643.62$, $df = 278$, CFI = .95, TLI = .95, RMSEA = .06), suggesting that academic motivation was more strongly related to self-esteem and the method effect factor for Latino respondents. To consider the possibility that the noninvariant item (respect) would change these conclusions, these analyses were replicated with that item's parameters constrained to equal across the groups and with that specific item omitted from the analyses. Results were the same, suggesting that noninvariance in the respect item was not responsible for the group differences in coefficients.

[Table 2 Omitted]

The last set of analyses included predictors of two RSES subdimensions using Model 10 to consider substantively meaningful differences in patterns of association between predictors of positive self-esteem versus negative self-esteem (Owens, 1993, 1994). An additional question considered whether there would be variation across groups in coefficients linked to both positive and negative self-esteem. Specifying specific paths to equality both within groups (e.g., comparing links to positive versus negative self-esteem) and across groups allowed for both a test of differential patterns of association between the two RSES subdimensions and also whether or not such differences are observed across groups. Using the entire sample, there was no group difference in predicting positive self-esteem, whereas there was a positive coefficient suggesting higher negative self-esteem (although these are reversed) among the Armenians compared to the European Americans. In addition, gender, maternal support, and academic motivation were positively associated with both RSES factors, while psychological control was negatively related. Based on these results, associations across the two RSES factors were compared on a sample-by-sample basis due to variation in coefficients across groups.

Cultural differences in paths linking predictors to the two RSES factors were evident based on a multigroup analysis using Model 10. A model that constrained the coefficients to equality (while also freeing the parameters associated with the respect item) demonstrated a good fit to these

data ($\chi^2 = 494.28$, $df = 282$, CFI = .97, TLI = .97, RMSEA = .04). Examination of the modification indices, however, suggested that paths varied across groups linking academic motivation and maternal support to positive self-esteem and age to both self-esteem factors. After freeing paths between groups linking academic motivation and positive self-esteem (stronger in the Latino group) and maternal support to positive self-esteem (stronger in the Armenian/Iranian group), the model fit improved ($\chi^2 = 455.51$, $df = 276$, CFI = .97, TLI = .97, RMSEA = .04). In addition, the extent to which older adolescents may score higher on both RSES subdimensions appeared most relevant to the Latino subsample. These results were similar even when forcing the respect item to equality across groups or by omitting that item.

Moving to comparisons of coefficients from predictors to the two RSES subdimensions, these analyses were considered on a group-by-group basis since some coefficients varied across cultural groups. In the Latino group, a model constraining paths to equality across RSES subdimensions (i.e., the path linking academic motivation to positive self-esteem was constrained to be equal to the path linking academic motivation to negative self-esteem) demonstrated a marginal fit to these data ($\chi^2 = 352.35$, $df = 77$, CFI = .93, TLI = .92, RMSEA = .07) and suggested freeing two model constraints. After freeing the equality constraints linking academic motivation to the two RSES factors and linking psychological control to the two RSES factors, the model demonstrated a good fit ($\chi^2 = 293.58$, $df = 75$, CFI = .97, TLI = .96, RMSEA = .04). The results, consequently, suggested that for the sample of Latino adolescents, academic motivation was more strongly associated with positive self-esteem than with negative self-esteem (see Table 2), while psychological control by mothers was more strongly associated with negative self-esteem compared to positive self-esteem.

The same analyses conducted on the European American sample suggested an acceptable fit and no modification indices suggesting that predictors largely were comparable to both the positive and negative self-esteem factors ($\chi^2 = 115.98$, $df = 77$, CFI = .96, TLI = .95, RMSEA = .06). Moreover, freeing the constraints suggested using the Latino group above did not result in an improved model fit. For the Armenian/Iranian sample, the overall model with equality constraints also demonstrated an acceptable fit to these data ($\chi^2 = 168.48$, $df = 77$, CFI = .96, TLI = .94, RMSEA = .06), although modification indices suggested that freeing the paths linking psychological control to both the positive and negative self-esteem factors and linking maternal support to the two RSES factors would improve model fit ($\chi^2 = 1,221.17$, $df = 75$, CFI = .98, TLI = .97, RMSEA = .04). Similar to findings in the Latino sample, this suggested that psychological control was more strongly related to negative self-esteem than to positive self-esteem. In addition, there was a stronger association with maternal support and positive self-esteem when compared to the coefficient linking support and negative self-esteem.

Related to findings linking predictors to two RSES subdimensions, these findings suggested that, across groups, age and gender differences were related in a similar manner to both positive and negative self-esteem, although age was more positively related to both factors in the Latino group. In two of the three groups, psychological control was more strongly related to negative self-esteem than to positive self-esteem, suggesting that this differential association may be particularly pronounced among ethnic minority groups. Findings in the Latino group also suggested that academic motivation is more strongly related to positive self-esteem, whereas in

the Armenian/Iranian group, maternal support was more strongly related to positive self-esteem than negative self-esteem.

Discussion

Consistent with previous research, these results suggested that method effects should be accounted for when using the RSES and that such a conclusion is consistent across cultural groups. Findings also suggested, however, that the method effects may be more pronounced among ethnic minority respondents and that a failure to account for method effects may lead to biased conclusions regarding how theoretically related covariates are associated with self-esteem. In addition, results also were consistent with arguments that when specified separately, positive and negative self-esteem factors are differentially related to predictors/covariates, however those differences may vary across cultural groups. Perhaps most importantly, these findings suggested that a possible group difference in self-esteem was confounded with method effects so that, once method effects were taken into account, there was no longer a significant group difference in self-esteem when comparing Armenian/Iranians and European Americans.

The first implication for cross-cultural researchers who study adolescents is that method effects in the RSES seem most pronounced in the negatively worded items and may be related to cultural group membership. Moreover, accounting for these method effects may reduce cultural differences in self-esteem and also alter the magnitude of associations between predictors and the RSES. In particular, if researchers conceptualize their target construct as positive self-evaluations, failure to account for method effects among the negatively worded items may bias overall “self-esteem” when using the RSES. One possible adverse consequence of ignoring these method effects is to conclude greater self-esteem in group versus another, when the group difference may result from different response tendencies and not actual differences in self-esteem. In addition, separating out possibly confounding effects of negatively worded items (either by specifying a latent factor accounting for method effects or a separate RSES factor) may lead to an RSES factor with greater validity. For example, associations with factors that are theoretically linked to self-esteem (academic motivation) get stronger, as was the case in the Latino sample. This was also evident when linking predictors to Model 10 where academic motivation was more strongly related to positive self-esteem than with negative self-esteem for Latinos, whereas maternal support was more strongly linked to positive self-esteem for Armenian/Iranians, and psychological control was more strongly related to negative self-esteem for Latinos and Armenian/Iranians. As such, a failure to account for method effects may bias conclusions regarding how factors such as academic success are differentially linked to self-esteem (when self-esteem is conceptualized as positive self-evaluations) across groups.

While it is difficult to definitively conclude that bidimensionality is related to method effects or substantive differences in two RSES subdimensions based on these analyses (Tomas & Oliver, 1999), the differential pattern of association was consistent with previous studies suggesting that two RSES factors vary in their associations with other constructs (Ang et al., 2006; Farruggia et al., 2004; Owens, 1994; Quilty et al., 2006). Based on such findings, researchers may argue that negatively worded RSES items are assessing a different set of psychosocial processes (possibly related to depressive symptoms or certain personality traits) related to negative self-evaluations that are substantively distinct from the typical conceptualization of self-esteem. For example,

these data suggest that adolescents who experience greater psychological control by mothers may be more adversely affected by their resulting self-deprecating thoughts rather than by declines in their positive self-evaluations (given the negative association and reverse coded items). Similarly, positive self-esteem may be more related to motivation to succeed in school as compared to the extent to which feelings of negative self-esteem can be harmful to academic motivation. To the extent that negative RSES items assess a different aspect of self-esteem, a consideration of how developmental antecedents and consequences of two distinct elements of self-esteem may vary would be needed (see Owens, 1993, 1994). Such a possibility raises research possibilities for cross-cultural researchers who could focus on identifying cultural groups that may encourage adolescents to develop both positive and negative self-evaluations (e.g., China) or identify cultures in which self-criticism and negative self-esteem may be more common (e.g., collectivistic cultures; Twenge & Crocker, 2002) versus other cultures (e.g., the United States) that may emphasize enhancing positive views of the self (Boucher, Peng, Shi, & Wang, 2009).

Implications related to measurement of self-esteem using the RSES are equally important. The inclusion of both positively and negatively worded items in the RSES may induce response effects in particular individuals, and as a result, researchers may attenuate associations between what is conceptualized as an assessment of positive self-evaluations and other theoretically related constructs when summing across RSES items. Previous research has suggested that biased responding related to negatively worded RSES items may be more common for individuals with specific personality traits (Quilty et al., 2006) and also may be related to cognitive ability or age (Marsh, 1996). Findings from this study also suggest that cultural group membership may be a factor related to method effects in the RSES. In addition, these findings suggested that a predisposition to answer negatively worded items differently was related to age, greater maternal support, and lower academic motivation and psychological control by mothers. These latter associations have questionable applicability to real-world research settings and, combined with the results including two RSES subdimensions, may suggest that the RSES is better specified as having one dimension (with only the positive worded items) if one is interested in treating negative self-esteem as a construct worthy of study. In addition, the findings related to predictors of a method effects may be proxies for other variables. For example, adolescents who are low in academic motivation and high in psychological control may have different personality profiles that explain these links (Quilty et al., 2006). While some authors have speculated that method effects are related to cognitive or language ability, recent studies suggest that method effects are highly stable across time (Marsh et al., 2010). The current study did consider generational status (born in the United States or not) and language use in the home (predominantly English or not) as linked to the latent method factor from Model 7 but found (although with imperfect measures of language or cognitive ability) no associations.

Limitations to this study include a limited focus on only three cultural groups. Although moving beyond previous studies by highlighting within-U.S. differences, research is still needed with more diverse groups to better understand cultural variation in bidimensionality and method effects in the RSES. Another limitation involved the differential sample sizes and a relatively small sample of European American adolescents that may have led to some biased conclusions due to lower statistical power (e.g., associations between separate RSES subdimensions and academic orientation were not different in this group). In particular, modification indices in

single-group analyses will reveal greater frequency of possible changes to the model with larger samples. In addition, the analysis strategy selected, while consistent with previous research, makes it difficult to draw conclusions regarding what are substantive differences in RSES subdimensions and what are method effects (Tomas & Oliver, 1999).

These findings are consistent with previous studies and provide further evidence that researchers should be aware of possible biases and error introduced to their measures when summing across RSES items to create a summary variable (Ang et al., 2006; Greenberger et al., 2003; Quilty et al., 2006). Researchers using the RSES in cross-cultural research on adolescents (particularly in comparative research) should develop a strategy to account for method effects associated with the negatively worded RSES items. One strategy for future research would be to follow suggestions by Owens (1993, 1994) and Boucher et al. (2009) to further develop conceptual models explaining how negative self-esteem is a different developmental issue for adolescents than is positive self-esteem. Such an approach would open up rich research opportunities regarding cultural differences in elements of self-esteem that emphasize positive or negative self-evaluations. Moreover, such a line of research might also lead to the development of more culturally appropriate assessments of self-esteem for cross-cultural research.

Additional recommendations focused on eliminating or accounting for method effects would be more appropriate for researchers who have already collected data such as using only the positively worded items to assess self-esteem (if the main study construct is positive self-esteem; Corwyn, 2000) or to use SEM-based approaches to either model method effect factors or to include correlated uniquenesses to achieve an adequately fitting unidimensional model. For researchers who are planning a data collection, altering the wording of the negative RSES items so that they describe a positive self-evaluation has also been proposed (Greenberger et al., 2003). At a minimum, cross-cultural researchers need to be aware that administering and scoring the full version of the RSES has the potential to produce biased research conclusions across cultural groups related to both mean differences in self-esteem and also correlational analyses linking external factors to self-esteem.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Aluja, A., Rolland, J., Garcia, L. F., & Rossier, J. (2007). Brief report: Dimensionality of the Rosenberg self-esteem scale and its relationships with the three- and the five-factor personality models. *Journal of Personality Assessment*, 88, 246-249.

Ang, R. P., Neubronner, M., Oh, S., & Leong, V. (2006). Dimensionality of Rosenberg's self-esteem scale among normal-technical stream students in Singapore. *Current Psychology, 25*, 120-131.

Boucher, H. C., Peng, K., Shi, J., & Wang, L. (2009). Culture and implicit self-esteem: Chinese are "good" and "bad" at the same time. *Journal of Cross-Cultural Psychology, 40*, 24-45.

Bush, K. R., Peterson, G. W., Cobas, J. A., & Supple, A. J. (2002). Adolescents' perceptions of parental behaviors as predictors of adolescent self-esteem in mainland China. *Sociological Inquiry, 72*, 503-526.

Carmine, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage Publications.

Corwyn, R. F. (2000). The factor structure of global self-esteem among adolescents and adults. *Journal of Research in Personality, 34*, 357-379.

Farruggia, S. P., Chen, C., Greenberger, E., Dmitrieva, J., & Macek, P. (2004). Adolescent self-esteem in cross-cultural perspective: Testing measurement equivalence and a mediation model. *Journal of Cross-Cultural Psychology, 35*, 719-733.

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg self-esteem scale: Do they matter? *Personality & Individual Differences, 35*, 1241-1254.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Marsh, H. W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology, 70*, 810-819.

Marsh, H. W., 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.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthen, L. K., & Muthen, B.O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: Muthen & Muthen Owens, T. J. (1993). Accentuating the positive and the negative: Rethinking the use of self-esteem, selfdepreciation, and self-confidence. *Social Psychology Quarterly, 56*, 598-609.

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

- Plunkett, S. W., & Bamaca-Gomez, M. Y. (2003). The relationship between parenting, acculturation, and adolescent academics in Mexican-origin immigrant families in Los Angeles. *Hispanic Journal of Behavioral Sciences, 25*, 222-239.
- Quilty, L. C., Oakman, J. M., & Risko, T. (2006). Correlates of the Rosenberg self esteem scale methods effects. *Structural Equation Modeling, 13*, 99-117.
- Rosenberg, M. (1989). *Society and the adolescent self-image*. Middleton, CT: Wesleyan University Press.
- Russell, S. T., Crockett, L. J., Shen, Y., & Lee, S. (2008). Cross-ethnic invariance of self-esteem and depression measures for Chinese, Filipino, and European American adolescents. *Journal of Youth and Adolescence, 37*, 50-61.
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347-363.
- Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89*, 623-642.
- Steenkamp, J.- B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research, 25*, 78-90.
- Tomas, J. M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling, 6*, 84-98.
- Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: Meta-analyses comparing Whites, Blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000). *Psychological Bulletin, 128*, 371-408.
- Vijver, F. J. R. van de, & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54*(2), 119-135.