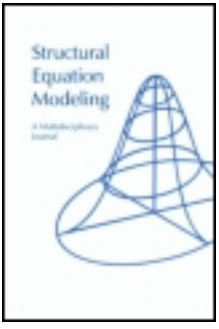
This article was downloaded by: [Temple University Libraries]
On: 26 October 2012, At: 00:31
Publisher: Psychology Press
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Structural Equation Modeling: A Multidisciplinary Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Rosenberg's self-esteem scale: Two factors or method effects

Jose M. Tomas [a] & Amparo Oliver [b]

[a] Department of Methodology, Psychobiology, and Social Psychology, Faculty of Psychology, University of Valencia, Avenida Blasco Ibáñez, 21, Valencia, 46010, Spain E-mail:

[b] Department of Methodology, Psychobiology, and Social Psychology, Faculty of Psychology, University of Valencia,

Version of record first published: 03 Nov 2009.

To cite this article: Jose M. Tomas & Amparo Oliver (1999): Rosenberg's self-esteem scale: Two factors or method effects, Structural Equation Modeling: A Multidisciplinary Journal, 6:1, 84-98

To link to this article: http://dx.doi.org/10.1080/10705519909540120

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution,

reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

STRUCTURAL EQUATION MODELING, 6(1), 84–98
Copyright © 1999, Lawrence Erlbaum Associates, Inc.

Rosenberg's Self-Esteem Scale: Two Factors or Method Effects

José M. Tomás and Amparo Oliver
Department of Methodology, Psychobiology, and Social Psychology
Faculty of Psychology
University of Valencia

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Self-esteem is one of the most studied constructs in psychology. It has been measured with a variety of methods and instruments. Although Rosenberg's (1965) self-report scale is one of the most widely used, empirical evidence on factor validity of this scale is somewhat contradictory, with either 1 or 2 factors. The results of this study suggest the existence of a global self-esteem factor underlying responses to the scale, although the inclusion of method effects is needed to achieve a good model fit.

Self-esteem is one of the most studied constructs in psychology. It has been measured with a variety of methods and instruments (Romero, Luengo, & Otero-López, 1994). Of those, Rosenberg's (1965) scale is the most widely used self-report instrument for assessing global self-esteem (Marsh, 1996). The scale is thought to measure the self-acceptance aspect of self-esteem (Crandall, 1973). Rosenberg's scale—a 4-point scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*)—was originally developed as a Guttman scale scored dichotomously (Fleming & Courtney, 1984; Rosenberg, 1965). However, most studies employ the scale as a Likert-type instrument. It consists of 10 items, 5 of them positively worded and 5 negatively worded. A positively worded item is, for example, "I feel good about myself"; a negatively worded item is, for example, "I certainly feel useless at times." Rosenberg developed this scale to measure a global self-esteem factor. However, empirical evidence is somewhat contradictory. Several studies defend the original one-factor structure, whereas others support a two-factor structure (positive and negative self-esteem). These contradictory results have been ex-

Requests for reprints should be sent to José Manuel Tomás, Department of Methodology, Psychobiology and Social Psychology, Avenida Blasco Ibáñez, 21, 46010, Valencia, Spain. E-mail: tomasjm@uv.es

plained in terms of response bias (response set, response effects, method effects) commonly present in rating scales containing positively and negatively worded items. When positively and negatively worded items are present in a self-report scale, factor analyses of responses frequently reveal different factors reflecting positive and negative items (Bachman & O'Malley, 1986; Bagozzi & Heatherton, 1994; Carmines & Zeller, 1979; Marsh, 1992). If this is the case with Rosenberg's scale, it must be decided whether the two factors are substantively meaningful rather than a method artifact.

Carmines and Zeller (1979) used exploratory factor analysis and criterion validity coefficients to study the dimensionality of the original 10-item scale. A principal-axis extraction indicated that there were two substantial empirical factors underlying the responses to the 10 items. They labeled Factor I as *positive self-esteem*, grouping items positively worded, and Factor II as *negative self-esteem*, grouping items negatively worded. When they factor analyzed the two sets of items separately, only one substantial factor emerged for each set of items. Regardless of these results, they considered an alternative interpretation: the possibility that the two-factor solution was a method artifact. To test this second interpretation, they argued that, if positive and negative self-esteem factors measured substantively different dimensions, they should relate differentially to some external variables (criteria). These two factors were correlated with 16 different criteria, but the difference in correlation was never statistically significant ($p > 0.25$). According to the results, Carmines and Zeller concluded that "the more appropriate interpretation is that the bifactorial structure of the items is a function of a single theoretical dimension of self-esteem that is contaminated by a method artifact, response set" (p. 69). This response set may have something to do with the participant's verbal ability (Kaufman, Rasinski, Lee, & West, 1991; Marsh, 1986).

Bachman and O'Malley (1977), using Cobb, Brooks, Kasl, and Connely's (1966) 6-item revision of the Rosenberg's scale, found empirical evidence for a one-factor solution to the responses of the scale, with factor loadings ranging from 0.38 to 0.69. Bachman and O'Malley (1986) found a good model fit of a confirmatory factor analysis (CFA) of the responses to the Rosenberg's scale only by fitting covariances among the residuals associated with the positively worded items and among residuals associated with the negatively worded items. This solution supports the presence of method effects in the scale.

Alvaro (1988) used principal component analysis with varimax rotation to study the structure of Rosenberg's scale in a Spanish sample. He used a Spanish translation of an eight-item version of the Rosenberg scale, developed by Warr and Jackson (1983). The sample consisted of employed and unemployed workers. Two components were retained, accounting for 56.8% of the variance together. The first component was defined by negatively worded items, whereas positively worded items defined the second component.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Kaufman et al. (1991) analyzed the responses to a revised version of the Rosenberg scale with three negatively worded items. On the basis of exploratory results, they proposed a different two-factor solution for responses to Rosenberg's items. The first factor was labeled *general evaluations of oneself* and the second factor was labeled *transient self-evaluations*. A CFA was then used to compare a one-factor and this two-factor solution. Their two-factor solution fit consistently better than the one-factor solution, but they concluded that neither was fully satisfactory.

Salgado and Iglesias (1995) used the Spanish translation of the original 10-item scale. They used CFA to compare the ability of four different measurement models to fit Rosenberg's scale data. Model 1 hypothesized that responses to the scale could be explained by a single self-esteem factor. Model 2 hypothesized a two-factor structure, a global factor of self-esteem explaining every item plus a method factor, response set, according to Carmines and Zeller's (1979) conclusions. Model 3 proposed a two-factor solution with positive and negative factors. Finally, Model 4 hypothesized that responses to the scale can be explained by two first-order factors (positive and negative self-esteem) and a second-order factor (global self-esteem). Several goodness-of-fit indexes were used to test these models. Model 3, positive and negative self-esteem, fit consistently better. However, most differences across indexes among Models 2, 3, and 4 were minimum. As an example, the comparative fit indexes (CFIs) were 0.93, 0.94, and 0.92, respectively; and the expected values of the cross-validation indexes were 0.88, 0.83, and 0.89, respectively. In addition, Model 2 only posited one method factor (it was not specified if this method factor was the one determining the negatively or positively worded items). There was no model with a global self-esteem factor and two method factors.

Marsh (1996) analyzed a seven-item scale with four positively worded items and three negatively worded items. These items were the self-esteem indicators of the Rosenberg scale, included in the data of the National Education Longitudinal Study of 1988. A set of CFA models was considered by Marsh (1996) to evaluate alternative interpretations of factor structure of Rosenberg's items. These models included: Model 1, single factor model; Model 2, two-factor model based on the transient and general evaluation items proposed by Kaufman et al. (1991); Model 3, positive and negative factors; Model 4, single self-esteem factor including correlated uniquenesses among negatively worded items; Model 5, single self-esteem factor including correlated uniquenesses among positively worded items; and Model 6, global self-esteem factor with correlated uniquenesses among negatively worded items and a pair of positively worded items. In Model 6, only one pair of correlated uniqueness between positively worded items was included to avoid identification problems (Marsh, 1996). Models 4, 5, and 6 were an application of the correlated uniqueness model to the analysis of multitrait–multimethod (MTMM) matrices in order to test for method effects or response bias (Marsh &

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Grayson, 1995; Wothke, 1996). The best-fitting model was Model 6. Marsh concluded that the results supported the existence of a global self-esteem factor underlying responses to Rosenberg's items, as well as the existence of method effects specially associated with negatively worded items.

The aforementioned review shows an ambiguous situation with competing models finding some empirical support. Unfortunately, some studies used Rosenberg's original 10-item scale, whereas others used some items from this scale (or revised scales).

This research builds on these studies, but it extends them in different ways. Some of these studies employed "exploratory factor analyses that are inherently weak in terms of distinguishing between competing factor structures" (Marsh, 1996, p. 811). CFA is an adequate statistical tool in this context. As a second contribution, this investigation analyzes the original 10-item scale, whereas most of the reviewed studies employed modifications of the scale and/or selected items. From a methodological point of view, the advantage of using the original scale is a higher ratio of indicators to factors. The ratio of observed variables to latent variables may be crucial in order to identify and estimate the CFA of MTMM data. From a substantive point of view, the construct validation of the scale requires, if possible, the complete scale rather than parts (or revisions) of it. An important contribution of this article is the fact that it focuses on studying a Spanish version of the scale. Method effects on Rosenberg's scale have to do with item wording, and therefore with linguistic issues. There is no need for results with the English version of the scale to generalize to other languages. In applied research with the Spanish version of the questionnaire, it is important to know if method effects are present in the scale and to what extent. There is only one article that employed CFA to analyze method effects on the Spanish version of the Rosenberg scale (Salgado & Iglesias, 1995). This study presented several shortcomings. First, the tested models did not consider any model positing both method effects; thus, conclusions about method effects were not comparable to other studies and so correlation among methods could not be tested. In other words, the proper and complete sequence of models was not used. Second, differences in fit among models were minimum and even contradictory. Third, the CFA employed to test for method effects was a modification of the correlated trait–correlated methods (CTCM) approach, which has several disadvantages. Other approaches were not employed (see the Method section for an explanation of CFA of MTMM data), and so the presence of potential multidimensional method effects was not tested. In sum, the purpose of this study is twofold: (a) to study the structure of Rosenberg's self-esteem scale in Spanish, and in particular to verify the presence of method effects; and (b) to compare MTMM approaches to the analysis of the questionnaire, assessing specifically the CFA approach with correlated traits and correlated methods and the correlated uniqueness model.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

METHOD

Data

The sample consisted of 640 high school students. Schools in the city of Valencia, Spain, were randomly sampled. In those schools that agreed to participate in the study, classrooms were randomly sampled. The mean age was 15.8. There was approximately the same number of male students (55.47%) as female students (44.53%). The survey was administered in classroom settings during a period without examinations. The data-gathering instrument consisted of a four-page booklet. The survey included sociodemographic variables, anxiety scales, Rosenberg's self-esteem scale, and other self-esteem questionnaires.

Rosenberg's scale was the translation into Spanish of the original 10-item scale (Rosenberg, 1965). It included five positively worded items (P) and five negatively worded items (N), with the sequence P–P–N–P–N–P–P–N–N–N. Data on the Rosenberg's scale were collected using a 4-point Likert scale response format. The common practice of reversing the negative items was done before the CFAs were run.

Statistical Analysis

CFA was employed in this study. The CFA was conducted using EQS 5.1 (Bentler & Wu, 1995). CFA is a well-known methodology (Bollen, 1989; Hayduk, 1987; Hoyle, 1995; Loehlin, 1987). Briefly, in CFA the researcher postulates a model (a particular linkage between observed variables and their underlying factors) and then tests this model statistically. The standard method of estimation is *maximum likelihood*, and it is the one used in this study. Maximum likelihood is based on the assumption that variables are multivariate normal distributed. However, there is growing evidence that it performs well under a variety of nonoptimal conditions, such as excessive kurtosis and so on (Hoyle & Panter, 1995). Mardia's coefficient of multivariate kurtosis for this data set was 24.24 (normalized estimate of 19.49), and so the data were not multivariate normal. An arbitrary distribution-free method was also applied to each model, but estimates and overall fit were very close to those of maximum likelihood and are not presented here.

A critical issue in any CFA is the assessment of model fit. Central to this assessment are the values of several goodness-of-fit indexes obtained from a specified model. The most common of these indexes is the chi-square test. If the model was specified correctly and the distributional assumptions for the data were satisfied, analysts could use a test statistic with an asymptotic chi-square distribution to test the null hypothesis that the specified model leads to a reproduction of the popula-

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

tion covariance matrix of the observed variables. A significant test statistic would cast doubt on the model specification (Bollen & Long, 1993). However, several problems arise with the use of the chi-square test, in particular that it is based on restrictive assumptions, depends on sample size, and that a model is an approximation of reality rather than an exact representation of the observed data (Bentler & Bonett, 1980; Cudeck & Browne, 1983; Jöreskog, 1969). To overcome these problems, a number of subjective fit indexes have been proposed and are commonly used. There is a broad consensus that no single measure of model overall fit should be relied on exclusively; therefore researchers are advised to use a variety of indexes from different families of measures (Marsh, Balla, & Hau, 1996; Marsh, Balla, & McDonald, 1988; Tanaka, 1993). From among the absolute fit indexes, the chi-square statistic, root mean-square residual (RMR), and the goodness-of-fit index (GFI) are reported. RMR is a summary statistic for the residuals, proposed by Jöreskog and Sörbom (1986); so the lower the value of the index, the better the model. Although the GFI is moderately related to sample size, it indexes the relative amount of the observed variances and covariances accounted for by a model, and it performs better than any other absolute index (Hoyle & Panter, 1995; Marsh et al., 1988). From among the incremental fit indexes, the Tucker-Lewis index (TLI) and CFI are reported. The TLI is a Type-2 incremental fit index that compares the lack of fit of the target model to the lack of fit of the independence model. The CFI is a Type-3 incremental fit index that indexes the relative reduction in lack of fit, as estimated by the noncentral chi-square of a target model versus a baseline model (see Hu & Bentler, 1995, for a classification of indexes). The TLI and CFI (the normed version of the RNI) differ primarily in that the TLI incorporates a correction for model complexity such that it rewards less complex models (Marsh, 1996). A value of 0.9 of these indexes has been proposed as a minimum for model acceptance (Bentler & Bonett, 1980). Following Hoyle and Panter's (1995) recommendations, no Type-1 index was considered.

There are several multivariate statistical models for analyzing method effects in MTMM data. CFA appears to be the most popular approach (Millsap, 1995; Schmitt & Stults, 1986; Wothke, 1996). Among the CFA models employed to analyze MTMM matrices, the CFA with correlated traits and correlated methods (CFA–CTCM, also known as general CFA model or block-diagonal model) is the most widely used. It hypothesizes that the total variation in observed variables can be written as a linear combination of trait, method, and error effects (Jöreskog, 1974). The CFA–CTCM model provides: (a) an explanation of MTMM matrices in terms of underlying factors, rather than observed variables; (b) evaluation of the convergent and discriminant validity at the matrix as well as at the parameter level; (c) the testing of hypotheses related to convergent and discriminant validity; (d) separate estimates of variance due to traits, methods, and uniqueness; and (e) estimated (disattenuated) correlations for both methods and trait factors (Byrne & Goffin, 1993). However, a major shortcoming of the

CFA–CTCM model is the frequent occurrence of ill-defined solutions, especially out-of-range values and convergence problems (Marsh, 1989; Marsh & Bailey, 1991; Marsh & Hocevar, 1988; Wothke, 1984). Another shortcoming is that the partitioning of variance into trait and method components may not, in general, yield trait-free and method-free interpretation (Bagozzi, 1993; Kumar & Dillon, 1992). Finally, it does not allow multidimensional method effects (Marsh & Grayson, 1995). Of those, it seems that the more general disadvantage is the high frequency of ill-defined solutions. Nevertheless, this does not have to be the case. Several authors have reported applications of the CFA–CTCM model that yield admissible estimates (Bagozzi, 1993; Bollen, 1989; Byrne & Goffin, 1993). However, Brannick and Spector (1990) suggested that researchers who use the CFA–CTCM model should assess the identification of their models, should be wary of computational problems and out-of-range parameters, and should avoid interpreting estimates when these problems occur. That was not the case of the CFA–CTCM models presented in this study, probably due to the high ratio of observed variables to factors (Brannick & Spector, 1990; Graham & Collins, 1991; Marsh, 1993; Marsh & Hocevar, 1988).

As a remedy for overfitting, ill-defined solutions, and the shortcomings of the CFA–CTCM model, a new approach to CFAs of MTMM data has been proposed (Marsh, 1988, 1989): the correlated-uniqueness model (CFA–CTCU). This model has three basic advantages (Bagozzi, 1993; Byrne & Goffin, 1993): (a) It seldom produces ill-defined solutions, (b) methods are not assumed to be unidimensional, and (c) the confounding of method variance with trait variance is avoided (when this is due to common trait variations across methods and traits are highly correlated). It also has some disadvantages (Bagozzi, 1993; Kenny & Kashy, 1992): (a) The interpretation of correlated uniqueness as method effects is not always clear and (b) it assumes that methods are uncorrelated. In this study, based on Marsh and Grayson's (1995) recommendation, both CTCM and CTCU models were proposed. This recommendation, applied to the present dataset, translates into the nine models presented in Figure 1.

Models

Nine a priori models were evaluated in this study (Figure 1). Model 1 presumed one global self-esteem factor. Model 2 posited two oblique factors, as in Carmines and Zeller (1979) or Salgado and Iglesias (1995). Model 3 also posited two factors, general evaluation and transient evaluation, as in Kaufman et al. (1991). Models 4 through 9 are models that posited a global self-esteem latent variable underlying responses to the items plus different method effects. Model 4 posited correlated uniquenesses among residual variances of the negatively worded items, whereas Model 5 posited correlated uniquenesses among residual variances of the positively worded items. Model 6 was a correlated uniqueness model in which method

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

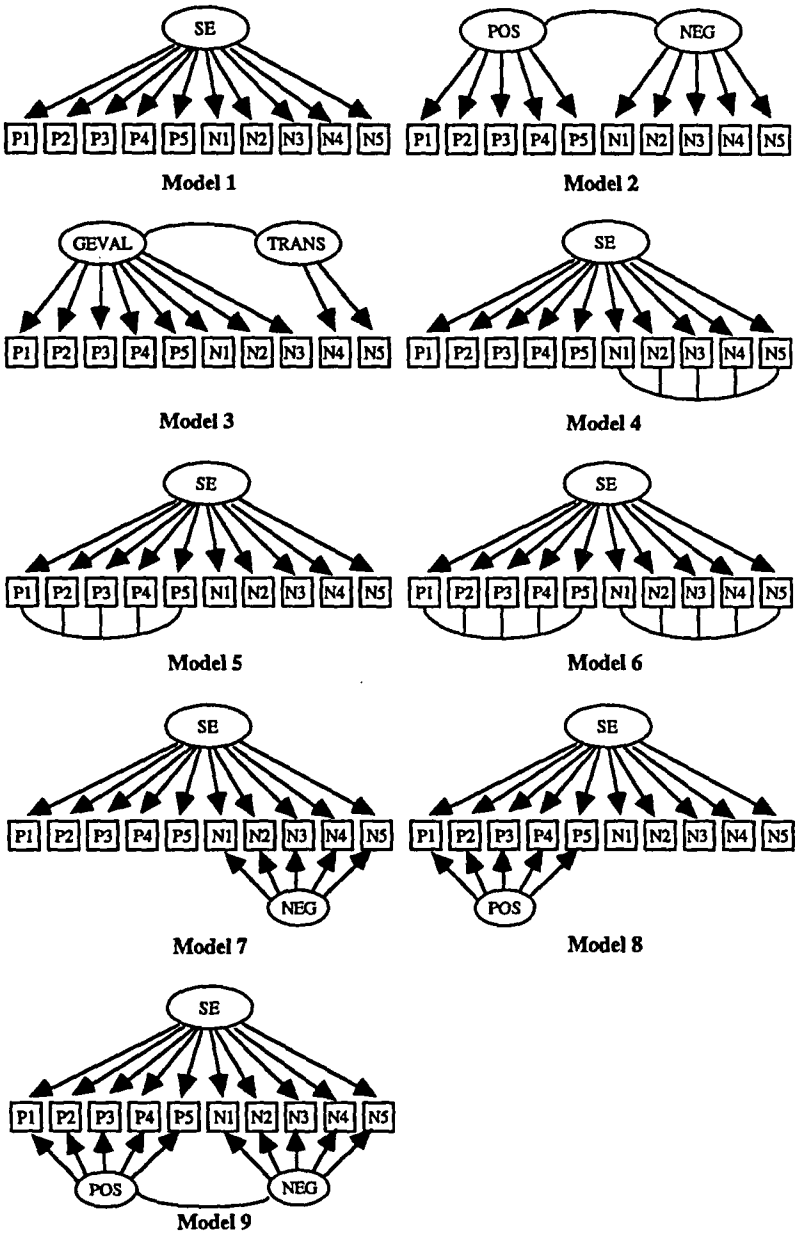


FIGURE 1 Tested models. P = positively worded items; N = negatively worded items; SE = global self-esteem; POS = positive self-esteem; NEG = negative self-esteem; GEVAL = general evaluation; TRANS = transient evaluation. For sake of clarity, uniqueness is not shown.

effects were inferred from correlated uniquenesses among positively and among negatively worded items. This model resulted in an improper solution, probably because this model is not identified in a global sense, with residual variance of Positive Item 4 (P4) as a Heywood case. Therefore, one of the correlated uniqueness of this item with Positive Item 2 (P2) was fixed to zero. This particular correlation among residuals was not statistically significant ($p > 0.1$), and was close to zero in Model 5. After that, a proper solution was obtained, and is the one reported in Table 1 as Model 6. Models 4, 5, and 6, or slight modifications of them, were tested by Marsh (1996). Model 7 presumed a trait factor of self-esteem underlying each item plus a method factor underlying negatively worded items. Model 8 posited a trait factor of self-esteem and a method factor underlying positively worded items. Finally, in Model 9 each observed variable was explained by a trait factor (self-esteem) and a method factor (depending on the item wording).

RESULTS

Overall fit results are shown in Table 1. The findings showed poor overall fit for both Model 1 (one global self-esteem factor) and Model 2 (positive and negative self-esteem). Moreover, the correlation between positive and negative self-esteem factors was 0.727 ($SE = 0.01$, correlation confidence interval limits: 0.704–0.746), indicating a pretty high amount of shared variance (a common underlying factor?).

All the models that posited method effects, Models 4 through 9, fit consistently better than Model 1. However, models positing method effects only among the positively worded items (Models 5 and 8) did not fit clearly better than Model 2. TLI, which incorporates a correction for model parsimony, evaluated these models as worse than Models 2 and 3 because of their complexity. In fact, Model 3 did

TABLE 1
Goodness-of-Fit Indexes for Models of Self-Esteem

	df	χ^2	χ^2/df	*RMR*	*GFI*	*TLI*	*CFI*
Model 1	35	352.80	10.08	.040	.880	.779	.828
Model 2	34	235.25	6.91	.036	.923	.856	.891
Model 3	34	136.19	4.01	.030	.955	.927	.945
Model 4	25	87.09	3.48	.019	.971	.939	.966
Model 5	25	188.22	7.53	.033	.940	.841	.912
Model 6	16	45.98	2.87	.015	.986	.954	.984
Model 7	30	94.37	3.14	.020	.968	.948	.965
Model 8	30	225.54	7.51	.034	.927	.841	.894
Model 9	24	55.52	2.31	.017	.983	.968	.983

Note: Chi-square tests statistically significant ($p < 0.001$); *df* = degrees of freedom; χ^2 = chi-square test; RMR = root mean-square residual; GFI = LISREL goodness-of-fit index; TLI = Tucker–Lewis index; CFI = comparative fit index.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

better than Models 5 and 8, supporting the idea that positive method effects are not enough to adequately represent the data.

It was when method effects among negatively worded items were considered that overall fit could be considered as adequate (consistent with findings by Marsh, 1996). Models 4 and 7, which posited method effects only among negatively worded items, fit the data better than models without method effects (Models 1, 2, and 3). However, models including method effects among both positively and negatively worded items (Models 6 and 9) were the best-fitting models; the overall fit indexes showed an excellent fit to data. It seems that positive as well as negative method effects are needed to achieve the best model fit. However, the fit of Models 4 and 7 (method effects only among negatively worded items) was also good, and much better than that of Models 5 and 8 (method effects only among positively worded items).

Both the CTCU model (Model 6) and CTCM model (Model 9) showed an excellent and nearly equal (in practical terms) fit to the data. A conjoint examination of both models' parameter estimates provided a much clearer idea of the nature and importance of method effects. In Model 6 (CTCU), several correlated uniquenesses were not statistically significant ($p > .05$). Specifically, 8 (of 9) correlated uniquenesses among positively worded items were not statistically significant, whereas only 2 (of 10) correlated uniquenesses among negatively worded items were not statistically significant ($p > .05$). Among them, all the correlated uniquenesses of P4 with the other positive items were found as statistically nonsignificant, and the nonsignificant correlated uniqueness among negatively worded items always implicated Negative Item 1 (N1). According to these results, it can be inferred that N1 and P4 did not present a significant method effect. The CTCM model (Model 9) is able to answer the question about the independence among methods and among traits. In this particular model, the correlation between method factors was not statistically significant ($\phi_{12} = -0.146$; $SE = 0.102$; $p > .05$). With respect to method loadings, all of them were statistically significant ($p \leq .05$) except for the loading of P4 ($\lambda = 0.107$; $SE = 0.054$; $p > .05$) and the loading of N1 ($\lambda = -0.083$; $SE = 0.031$; $p > .05$). These two items were also considered unaffected by method effects, taking into account the results from the CTCU model.

An important result to underline is the fit of Model 3 (proposed by Kaufman et al., 1991). This was the best-fitting model among the three models that only posited trait factors and no method effect (Models 1, 2, and 3). This result is consistent with findings by Marsh (1996), who found a CFI and a TLI of 0.979 and 0.966, respectively, indicative of acceptable model fit if considered in isolation. In our study, CFI, TLI, and GFI were over 0.9 for this model, and RMR was low, also indicative of acceptable fit (see Table 1). The transient self-evaluation factor in Model 3 has two indicators (Items N4 and N5), which are the same as Items NEG1 and NEG2 in Marsh's study. These two items have in common not only the psychological content proposed by Kaufman et al. (1991), but the negative item wording. Thus, the covariation

of these items may be due to substantive psychological aspects or to strong method effect. The factor loadings on the negative method factor in Model 9 should help to discern between both possibilities. As expected, Items N4 and N5 are those with the highest loadings in the negative method factor ($\lambda_{N4}=-0.664$ and $\lambda_{N5}=-0.682$), comparing with the other negatively worded items ($\lambda_{N1}=-0.083$, $\lambda_{N2}=-0.165$, and $\lambda_{N3}=-0.273$). They also have the pair of uniquenesses in Model 6 with the highest correlation. Under these circumstances it is normal that Model 3 obtained a better fit than Models 1 and 2, although it probably confounds pure method effects with psychological contents of substantive interest.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

CONCLUSIONS AND IMPLICATIONS

The results of this study suggest the existence of a single factor (global self-esteem) underlying responses to Rosenberg's scale. However, the inclusion of method effects is needed to achieve a good model fit. Method effects are associated with item wording, especially for negatively worded items. These results support findings by Marsh (1996). These results also highlight the need to explicitly study method effects in self-concept scales (and maybe other personality scales). Several of the reviewed studies did not test for method effects, and found empirical support for a two-factor structure of self-esteem. This result can be quite usual in the absence of method factors. In fact, Model 3 in our results is the best fitting model among the ones not including method factors. In other words, if method factors are not modeled in factorial validity studies in which positively and negatively worded items are present, the underlying structure of the responses to the scales can be obscured by method bias.

CFA of MTMM matrices has proved extremely useful when testing for response bias. Both CTCM and CTCU procedures were used in this study. Several points about the CTCM and CTCU models are worth noting. Although addressed to the same type of data (MTMM data) and the same type of research questions (method effects), they rely on different rationale and parameterizations. The CTCU model was mainly presented to overcome CTCM model shortcomings. However, both models are different, and they cannot be used interchangeably. The CTCM model addresses unidimensional method effects, whereas the CTCU model can handle both unidimensional and multidimensional method effects. When multidimensional method effects are present, the CTCU should be the method of choice. On the other hand, the CTCU is not adequate when method factors are correlated to some degree, whereas the CTCM model can handle method factor correlations. When method factors are oblique the, CTCM model should be the method of choice.

Unfortunately, there is no way to know if multidimensional method effects and/or correlations among method factors are present in the data before applying analytical tools to them. Therefore, an a priori election between the CTCM and

CTCU is not justified, unless other analytical tools are applied to the same sample or prior knowledge about the problem is available (i.e., it is a replication study). In other words, if the CTCM model was not estimated, how does one know that method factors were uncorrelated? Or, alternatively, if the CTCU model was not estimated, how does one know that the method effects were unidimensional? It is precisely in the comparison of the CTCM model and CTCU model where one can reinforce their conclusions.

In this study, both CFA approaches (Models 6 and 9) achieved similar results, and there was no significant correlation between the method factors. Comparing both models, it can be concluded that method factors were orthogonal and unidimensional (particular circumstances in which both models would theoretically perform well). In such circumstances there is probably no need to decide between them. In applied settings there is usually previous research that provides guidelines about statistical modeling. For instance, Marsh's (1996) results may help to discern the relative merits of the CTCM and CTCU models for our data. However, Marsh's study used an English version of the Rosenberg scale, whereas our study used a Spanish adaptation; Marsh's study employed a 7-item version of the scale, with 3 negatively worded items, whereas our study analyzed the responses to the original 10-item scale, with 5 negatively worded items; and finally, Marsh only applied the CTCU model, and so information about correlation between method factors was not available. In this context, we decided to apply both models. We suggest this strategy as a standard procedure, wherever the CTCM and CTCU models have no estimation or identification problems. The results show that, if there is no ill-defined solution, the CTCM model can still be useful when combined with the CTCU approach.

Finally, there are potential directions for further research. First, what needs to be considered again is whether negative items measure something different that is substantively important or a mere method effect. To further analyze this question, it would be useful to relate self-esteem as defined in each of the models here to other trait measures with positively and negatively worded items. Second, an effort is needed to study if this wording effect is also present in other social and personality scales, and if this effect is present across different ages and educational levels. Third, there is also a need to analyze the wording effect when Rosenberg's self-esteem scale is related to other variables and/or included into multivariate models. It is common practice to use an unweighted (or weighted) sum of the items as the variable representing self-esteem. This variable is then related to criteria or used in multiple regression, path analysis, and so on. If method effects are present in the scale, as the evidence suggests, then such a procedure confounds substantive factors (self-esteem) with wording effects (whenever the other variables had positively or specially negatively worded items, the relations with Rosenberg's scale may be biased). These biases may lead to misleading results and interpretations and are difficult to assess unless method factors were explicitly modeled.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

ACKNOWLEDGMENTS

This work was partly supported by Spanish Government PB 96–0791 (Programa Sectorial de Promocion del Conocimiento).

We thank Herbert W. Marsh and three anonymous reviewers for valuable and helpful suggestions in reviewing this article, and Eusebio Rial-González and Eva Oliver for helpful contributions in preparing the manuscript.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

REFERENCES

Alvaro, J. L. (1988). *Desempleo y bienestar psicológico [Unemployment and psychological well-being].* Unpublished doctoral thesis, Universidad Complutense de Madrid.

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology, 35,* 365–380.

Bachman, J. G., & O'Malley, P. M. (1986). Self-concepts, self-esteem and educational experiences: The frog pond revisited (again). *Journal of Personality and Social Psychology, 50,* 35–46.

Bagozzi, R. P. (1993). Assessing construct validity in personality research: Applications to measures of self-esteem. *Journal of Research in Personality, 27,* 49–87.

Bagozzi, R. P., & Heartherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1,* 35–67.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Bentler, P. M., & Wu, E. J. C. (1995). *EQS for Macintosh user's guide.* Encino, CA: Multivariate Software.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Brannick, M. T., & Spector, P. E. (1990). Estimation problems in the block-diagonal model of the multitrait–multimethod matrix. *Applied Psychological Mesurement, 14,* 325–334.

Byrne, B., & Goffin, R. (1993). Modeling multitrait–multimethod data from additive and multiplicative covariance structures: An audit of construct validity concordance. *Multivariate Behavioral Research, 28,* 67–96.

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

Cobb, S., Brooks, G. H., Kasl, S. V., & Connely, W. E. (1966). The health of the people changing jobs: A description of a longitudinal study. *American Journal of Public Health, 56,* 1476–1481.

Crandall, R. (1973). The measurement of self-esteem and related constructs. In J. P. Robinson & P. R. Shaver (Eds.), *Measures of social psychological attitudes* (2nd ed., pp. 45–167). Ann Arbor, MI: Institute for Social Research.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18,* 147–167.

Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem II: Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology, 46,* 404–421.

Graham, J. W., & Collins, N. L. (1991). Controlling correlational bias via confirmatory factor analysis of multitrait–multimethod data. *Multivariate Behavioral Research, 26,* 607–629.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore: John Hopkins University Press.

Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues and applications*. Thousand Oaks, CA: Sage.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 159–176). Thousand Oaks, CA: Sage.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1–56). San Francisco: Freeman.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood and least square methods.* Mooresville, IN: Scientific Software, Inc.

Kaufman, P., Rasinski, K. A., Lee, R., & West, J. (1991). *National Education Longitudinal Study of 1988. Quality of the responses of eighth-grade students in NELS88.* Washington, DC: U.S. Department of Education.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165–172.

Kumar, A., & Dillon, W. R. (1992). An integrative look at the use of additive and multiplicative covariance models in the analysis of MTMM data. *Journal of Marketing Research, 29,* 51–64.

Loehlin, J. C. (1987). *Latent variable models.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H. W. (1986).The bias of negatively worded items in rating scales for young children: A cognitive-developmental phenomenon. *Developmental Psychology, 22,* 37–49.

Marsh, H. W. (1988). Multitrait–multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement, and evaluation: An international handbook* (pp. 570–578). Oxford, England: Pergamon.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait–multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13,* 335–361.

Marsh, H. W. (1992). *Self-description questionnaire II: Manual.* MacArthur, Australia: Faculty of Education, University of Western Sidney.

Marsh, H. W. (1993). Multitrait–multimethod analyses: Inferring each trait/method combination with multiple indicators. *Journal of Applied Educational Measurement, 6,* 49–81.

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

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analysis of multitrait–multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15,* 47–70.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 102,* 391–410.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait–multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 177–198). Thousand Oaks, CA: Sage.

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait–multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73,* 107–117.

Millsap, R. E. (1995). The statistical analysis of method effects in multitrait–multimethod data: A review. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods and theory: A festchrift honoring Donald W. Fiske.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Romero, E., Luengo, M. A., & Otero-López, J. M. (1994). La medición de la autoestima: una revisión [Self-esteem measurement: A review]. *Psicologemas, 8,* 41–60.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012

Salgado, J. F., & Iglesias, M. (1995). Estructura factorial de la escala de autoestima de Rosenberg: Un análisis factorial confirmatorio [Factor structure of Rosenberg's self-esteem scale: A confirmatory factor analysis]. *Psicológica, 16,* 441–454.

Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement, 10,* 1–22.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Warr, P., & Jackson, P. R. (1983). Self-esteem and unemployement among young workers. *Le Travail Humain, 46,* 355–366.

Wothke, W. (1984). *The estimation of trait and method components in multitrait-multimethod measurement.* Unpublished doctoral dissertation, Department of Behavioral Science, University of Chicago.

Wothke, W. (1996). Models for multitrait–multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Downloaded by [Temple University Libraries] at 00:31 26 October 2012