Stability of Self-Esteem: Demonstrating the Effects of Various Assumptions in Regression-Style Analyses

Lee M. Wolfle
Virginia Polytechnic Institute and State University

Every statistic requires some assumptions. This paper examines some of the assumptions in regression-style analyses of the stability of self-esteem, and inspects the consequences of some of the assumptions one makes with regard to measurement error and the distribution of variables.

Every statistic requires some assumptions. And to the extent the assumptions are not met, the statistics calculated will vary consequently. Robustness studies can estimate to some extent those consequences by systematically varying the assumptions against known but arbitrary parameters. In most situations, however, one is faced with sample data with unknown parameters, with assumptions met with varying degrees of accuracy, and with unknown inferential consequences.

Background

One of the most frequently studied constructs in psychology is self-esteem. It has most often been measured with some form of Rosenberg's (1965) instrument, in which respondents rate themselves on a Likert-type scale from strongly disagree to strongly agree in response to items such as "I feel I am a person of worth, on an equal plane with others."

Just this question was asked in the National Longitudinal Study (NLS) of the High School Class of 1972 (Riccobono, Henderson, Burkheimer, Place & Levinsohn, 1981), which was designed to provide data on the development of educational, vocational, and personal aspects of the lives of adolescents as they made the transition from high school to the adult world. Included among the many items of data collected for over 20,000 respondents were four self-esteem items, including the one quoted above. The analysis reported here was restricted by listwise deletion of missing data to 3,511 U.S. white males who answered these items completely in 1972 during their senior year of high school and seven years later in a 1979 follow-up study, along with information provided about their postsecondary educational attainment.

The apparently simple question to be addressed here is the stability of self-esteem for U.S. white males in the seven years following high school. Do those who exhibit high (or low) self-esteem in 1972 continue to do so seven years later?

One approach toward addressing such a question would be to simply regress the 1979 responses on their 1972 counterparts and thus estimate the average rate of change of self-esteem in metric or standardized form (more complete information is conveyed, of course, by reporting both). The closer the estimated coefficient comes to unity, the greater the agreement between 1972 and 1979 responses.

However, such a regression approach requires certain assumptions that may or may not be met in varying degrees. Indeed, any analytic approach toward estimating the stability of self-esteem will involve certain assumptions, and it is the purpose of this short paper to briefly touch upon these as they affect the motivating question about the stability of self-esteem. The purpose is not to find a definitive answer, but rather to inspect the consequences of the assumptions one adopts in seeking an answer.

Regression Examples

Take the regression of 1979 responses to "I feel I am a person of worth..." on the 1972 responses to the same question. For this sample, the resulting regression equation was estimated to be:

$$S2B = a + \underset{(.016)}{.150}\, S1B + e \qquad (\hat{\beta} = .160;\ R^2 = .026) \qquad (1)$$

in which S2B stands for self-esteem measured at time 2 on item B (of four self-esteem items included in the NLS), and S1B stands for the same item at time 1; "a" is an intercept or constant; .150 is the estimated metric regression coefficient and its standard error is shown below it in parentheses; "e" is the error of prediction or residual; $\hat{\beta} = .160$ is the standardized regression coefficient, and $R^2 = .026$ is the coefficient of determination.

The same equation can be shown diagrammatically in Figure 1-A, in which the arrow from S1B to S2B indicates that S1B is thought to be a cause of, or to cause changes in, S2B; and the short disconnected arrow represents all sources of variation in S2B not explained by, and not correlated with, S1B.

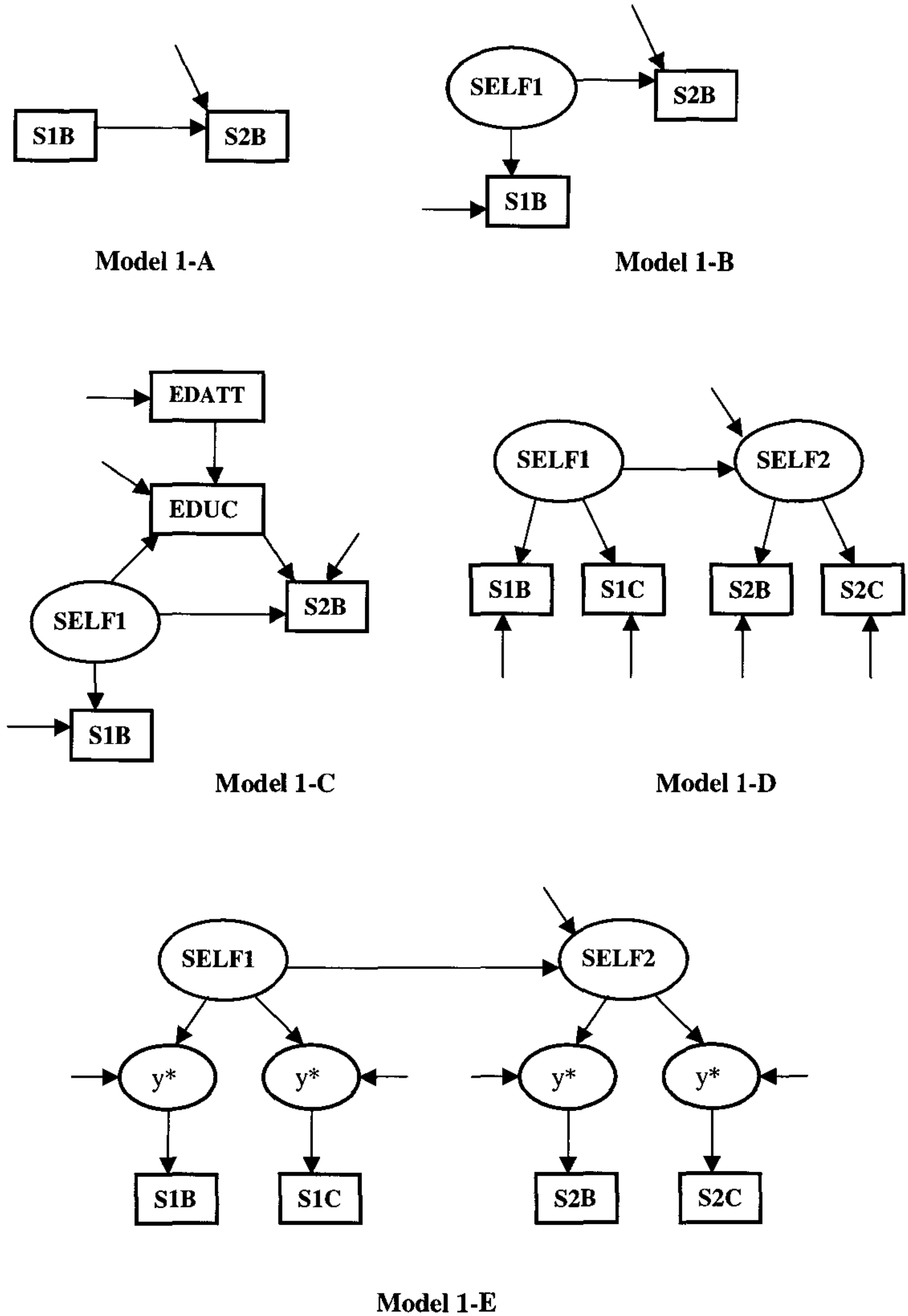

Figure 1. Models of the Stability of Self-Esteem

These results indicate that a one-unit change in self-esteem in 1972 may be expected to produce an average change of .150 units of self-esteem reported in 1979, and that less than 3 percent of the variation in self-esteem in 1979 is explained by self-esteem in 1972. The remaining 97 percent of the variation in self-esteem in 1979 is due to all other unmeasured (in this equation) sources of variation, including random changes, measurement error in self-esteem, deviations from linearity, and all other unspecified causes of self-esteem in 1979 uncorrelated with self-esteem in 1972.

The statement that a one-unit change in self-esteem in 1972 may be expected to produce an average change of .150 units of self-esteem reported in 1979, however, is based on an implicit assumption about the accuracy with which self-esteem was measured. As pointed out, for example, some time ago (Costner, 1969; Werts, Rock, Linn, & Jöreskog, 1976; Wolfle, 1979) and more recently (Rigdon, 1994), the regression approach assumes, in this instance, that self-esteem was measured perfectly! More specifically, it assumes that S1B was measured with reliability equal to 1.0. But as Schumacker and Lomax (1996, *p.* 38) have pointed out, the effect of unreliable variables on statistics can sometimes have dramatic effects.

Adding a measurement component to the analysis could be shown diagrammatically in Figure 1-B, in which Self1 is considered to be a so-called latent, unmeasured variable, thought to be a cause of S1B, the manifestly measured variable. In the previous equation, the coefficient thought to relate Self1 to S1B would be 1.0, and the residual of S1B (or, more accurately, the residual variance) would be zero, indicating no error of measurement. In this case, then, Figures 1-A and 1-B would be identical.

What if we relax the assumption of perfect measurement of the independent variable? What if we assume the reliability of the measurement of self-esteem in 1972 was less than unity? We can do this. Let us assume the reliability was 0.70. (This is not completely arbitrary, and is approximated from a confirmatory factor analysis with these data of the four self-esteem items included in the NLS.)

Assuming that the reliability of S1B was r_{xx} = .70 would imply that the error variance for S1B would be (1 - r_{xx})(variance of S1B) = (1 - .70)(.315) = .0945 (see Jöreskog & Sörbom, 1993, *p.* 37). The model thus implied in Figure 1-B was estimated with LISREL 8.30 (Jöreskog & Sörbom, 1993) using maximum likelihood estimates derived from the variances and covariances for the data described above.

The resulting (structural portion in LISREL terminology) equation was estimated to be:

$$\text{S2B} = a + \underset{(.022)}{.215}\ \text{Self1} + e \qquad (\hat{\beta} = .192;\ R^2 = .037) \qquad (2)$$

In this case, assuming S2B was regressed on S1B corrected for measurement error (i.e., Self1), one would expect that a one-unit change in self-esteem (Self1) in 1972 would be expected to produce a change of .215 units of self-esteem in 1979 (S2B) with an R-square of .037.

That errors in the independent variable reduce the coefficient in a bivariate regression is well known (e.g., Walker & Lev, 1953, *p.* 305). As seen here, the uncorrected regression coefficient of .150 underestimates the corrected (by measurement error) estimate of .215 by 30%. The converse is not true — errors in the dependent variable have no effect on ordinary least squares regression estimates, since such errors are absorbed as ordinary disturbances of prediction (Goldberger, 1964, *p.* 284).

The extension of this example to the case of two or more explanatory variables introduces unknowns into the system of equations that involve varying degrees of measurement error and multicollinearity among the predictors (Namboodiri, Carter & Blalock, 1975, *pp.* 541*ff*). In the present case, for example, one might expect that additional years of formal postsecondary education from 1972 to 1979 (see, e.g., Pascarella & Terenzini, 1991, *pp.* 162*ff*) might partially mediate (e.g., Baron & Kenny, 1986) the effect of self-esteem in 1972 on self-esteem in 1979, or equivalently said, that a part of the causal relationship of self-esteem in 1972 and 1979 occurs indirectly (e.g., Duncan, 1975) through the intervening accumulation of additional years of education.

Yet the effects of measurement error on such estimates is not necessarily predictable *a priori*. In this instance, consider the model shown in Figure 1-C, in which S2B is seen to be caused by Self1 and Educ, a measure of additional years of formal postsecondary education. Initially, the variable Educ is thought to be a latent variable measured without error, as is Self1. That is, initially Educ and EDATT are thought to be equivalent in which the error of EDATT is zero; similarly, Self1 and S1B are thought to be equivalent. The resulting set of two equations was estimated to be:

$$\text{Educ} = a + \underset{(.040)}{.204}\ \text{Self1} + e \qquad (R^2 = .007)$$

$$\text{S2B} = a + \underset{(.016)}{.141}\ \text{Self1} + \underset{(.006)}{.044}\ \text{Educ} + e \qquad (R^2 = .038) \qquad (3)$$

The difference in the estimated stability of .150 of self-esteem in Equation (1) and that of .141 in Equation (3) represents the indirect effect of 1972 self-esteem through educational attainment, namely (.204)(.044) = .009. In standard form, the same set of equations would be:

$$\begin{aligned} \text{Educ} &= .204\ \text{Self1} \\ \text{S2B} &= .151\ \text{Self1} + .113\ \text{Educ} \end{aligned} \tag{4}$$

If we no longer assume perfect measurement, the results are not as tractable. If we assume, as before, that self-esteem in 1972 was measured with reliability of .70, and that educational attainment was measured with reliability of .85, then the model shown in Figure 1-C can be re-estimated by specifying that the errors of S1B and EDATT, respectively, are not zero, but rather (1 - .70)(.315) = .0945 for S1B and (1 - .85)(1.818) = .2727 for EDATT. The resulting set of two equations was estimated to be:

$$\begin{aligned} &\text{Educ} = a + .295\ \text{Self1} + e \\ &(R^2 = .012) \qquad (.058) \\ &\text{S2B} = a + .203\ \text{Self1} + .049\ \text{Educ} + e \\ &(R^2 = .051) \quad (.023) \qquad (.008) \end{aligned} \tag{5}$$

and in standard form:

$$\begin{aligned} \text{Educ} &= .111\ \text{Self1} \\ \text{S2B} &= .180\ \text{Self1} + .117\ \text{Educ} \end{aligned} \tag{6}$$

These results are not easily related to those previously reported, except by their relative magnitudes, due to variations in reliabilities and multicollinearity (except to note its near absence in this instance), and have led many researchers to assume, for example, that "all instrumental variables are measured without error" (Wonnacott & Wonnacott, 1970, p. 371). Another approach (not pursued here) would be to construct overidentified models that could allow the estimation of, and correction for, random and systematic measurement error in variables (e.g., Wolfle, 1982).

Returning to the consideration of the stability of self-esteem without an intervening variable, with more information than just the single covariance of responses to the self-esteem items in 1972 and 1979 (and an assumed estimate of reliability of self-esteem in 1972 imposed on the model) one could estimate both the reliabilities of multiple self-esteem items and the stability of the latent estimates of self-esteem in 1972 and 1979. This can be accomplished by taking advantage of responses to multiple self-esteem items at the two time periods. Another such stem item was "I am able to do things as well as most other people." If we incorporate that item into the analysis, we could represent it appropriately as shown in Figure 1-D.

Figure 1-D specifies that a latent self-esteem variable, Self1, is the cause of two manifest items in the 1972 survey, S1B and S1C; similarly, Self2 is seen to be the cause of two identically worded items in the 1979 survey, S2B and S2C. In order to establish a metric for the latent variables, the slopes relating Self1 to S1B and Self2 to S2B were set to unity. The other two measurement parameter slopes were free to be estimated, as was the parameter relating Self1 to Self2. There were also two variances of latent variables to be estimated, as well as four error variances for the four manifest variables. With ten variances and covariances among the four manifest variables, this model is actually overidentified with one degree of freedom. Variations of this model have appeared before, some frequently with standardized variables, as early as some of Wright's (1934) work, and early work in the literature of path analysis in sociology (Siegel & Hodge, 1968). In standardized form, with minor restrictions, this is, of course, Spearman's (1907) correction for attenuation. It is also the model introduced by Costner (1969, Figure 4) that came to be known as the walking dog model, because of the visual appearance of the diagram.

This model (shown in Figure 1-D), unlike that of the model implied by Figure 1-A, makes no a priori assumption of error-free measurement. The estimate of the stability of self-esteem is thus adjusted for measurement error. The resulting structural equation was estimated to be:

$$\begin{aligned} &\text{Self2} = a + .291\ \text{Self1} + e \\ &(\hat{\beta} = .255;\ R^2 = .065) \quad (.029) \end{aligned} \tag{7}$$

This result was obtained with LISREL 8.30 (Jöreskog & Sörbom, 1993) using maximum-likelihood estimates from the covariance matrix, resulting in a model that exhibited a likelihood-ratio chi-square of 10.54 with 1 degree of freedom and a root mean square error of approximation (RMSEA; Steigler & Lind, 1980) of .052, which is numerically less than the cutoff value close to .06 recommended by Hu and Bentler (1999).

In substantive terms, then, having relaxed the constricting assumption of error-free measurement of self-esteem, one would expect that a one-unit change in self-esteem in 1972 (Self1) would be expected to produce a change of .291 units of self-esteem in 1979 (Self2), or .255 standard deviations, with an R-square of .065.

This latest estimate, however, is itself not free of assumptions of some kind. In particular, by estimating the associations by the method of maximum likelihood, one assumes that the manifest variables are distributed multivariate normally, and as we shall see, this is an unrealistic assumption to make with regard to the measurement of self-esteem. For example, the NLS respondents were asked to respond to "I feel I am a person of worth, on an equal plane with others," on a 4-point scale, to which they could respond (1) agree strongly, (2) agree, (3) disagree, or (4) disagree strongly. (For this analysis, the items were reverse coded so that higher scores indicated higher self-esteem.) But these variables are highly skewed and kurtotic; for example, the estimate of skewness was -.377 (z = -4.66) for S2B and the estimate of kurtosis was -1.237 (z = -14.96), as calculated by PRELIS 2.30. Indeed, for these subjects no one agreed strongly with this statement! While non-normal bivariate distributions can occur with normal marginals (Kowalski, 1973), it may be said with near certainty that the non-normal univariate distributions seen here insure non-normal multivariate distributions.

Possible Solutions

A new (actually, a renewed) feature in PRELIS 2.30 (Jöreskog, Sörbom, du Toit & du Toit, 1999, pp. 162ff) provides one possible solution to this violation of the assumption of non-normality, namely to normalize the variables before analysis. The idea would be to substitute normal scores as a continuous variable rather than ordinal scores, but it is doubtful that this tact would work in general and simply does not work in the present case since the ordinal and the so-called normalized variable are correlationally equivalent.

A more useful approach would be to treat self-esteem scores as if they were ordinal and censored measures of latent, continuous normal distributions. Diagrammatically, this may be shown in Figure 1-E, in which the y^* variables represent unmeasured estimates of continuous normal variables thought to be caused by latent self-esteem factors, and thought in turn to underlie the ordinally, manifestly measured self-esteem scores (Muthén, 1984; Jöreskog, 1990, 1994).

Estimates of the moment matrix of the associations among the four y^* variables may be obtained with the use of polychoric correlations among the four variables, a procedure that is available in PRELIS 2.30 (Jöreskog & Sörbom, 1996). We are assuming in this case that there exist normally distributed, continuous y^* variables that underlie the ordinal self-esteem variables; furthermore, the associations among these four y^* variables can be estimated, but their metric is unknown, hence their associations are measured in correlational terms, specifically with polychoric correlations, where an underlying bivariate normal distribution is assumed for each pair. The assumption of bivariate normality among the y^* variables may be tested with a chi-square goodness-of-fit test implemented in PRELIS; for these variables even this assumptions is questionable since all such bivariate tests should be rejected at the .01 level.

In order to estimate the structure implied by Figure 1-E for these variables, a general fit function called "asymptotically distribution free" by Browne (1982, 1984) or "weighted least squares" by Jöreskog and Sörbom (1996) was employed. This procedure requires a weight matrix for the polychoric correlations obtained from PRELIS 2.30. As a practical matter, this weight matrix increases rapidly in size as the number of variables increase. In this instance, we have
$k = 4$ variables with $p = k(k + 1)/2 = 10$ unique moments, and the weight matrix is of the order p x p = 100 with $p(p + 1)/2 = 55$ unique elements. As a further practical matter, to estimate moments of the fourth order with reasonable precision requires very large samples (Jöreskog & Sörbom, 1996, *p*. 28).

With these caveats in mind, estimates for the model implied by Figure 1-E were obtained with PRELIS 2.30 and LISREL 8.30. The resulting structural equation in standardized form was estimated to be:

$$\text{Self2} = a + .292\ \text{Self1} + e$$
$$(R^2 = .085) \qquad (.031) \tag{8}$$

The full model exhibited a likelihood-ratio chi-square of 8.58 with 1 degree of freedom and an RMSEA of .046. These results can be compared to the standardized estimate shown in Equation (7), and we see that the estimated standardized stability of self-esteem is now estimated to be .292 rather than .255. Which is to say, by specifying a model with self-esteem at two points in time, with two fallible indicators each, all distributed bivariate normally, but measured with censored ordinal variables, then the standardized estimate of stability is .292 with an R-square of .085.

Conclusions and Implications

That last sentence may be difficult to read with all of its clauses, but it represents most of the travails that got us to this point. We began by estimating the stability of self-esteem by assuming it was measured perfectly. That assumption was relaxed a bit by imposing a degree of measurement error on self-esteem as measured in 1972. With more than one predictor, the estimated stability of self-esteem changes, and changes yet again depending on assumptions of measurement error imposed on the model. If multiple indicators of self-esteem are brought to the analysis, one no longer must assume lack of measurement error, or impose arbitrarily estimated levels of measurement error, but can estimate a model incorporating estimates of both the measurement properties and stability of self-esteem. But those estimates were purchased at the cost of distributional assumptions among the manifest variables. If the manifest variables cannot be assumed to be normally distributed, perhaps their underlying distributions can be, but the data-collection costs to obtain such estimates can be high.

Thus, the estimate of the stability of self-esteem for a sample of high school males depends on the assumptions one is willing to make about the variables involved. In a simulation study, one would start with known parameters and examine on the average how estimates deviated as a consequence of the effects of varying assumptions. In the present case, however, the parameters are unknown and we simply do not know if the correct (standardized) measure of stability is .160, .192, .151, .180, .255, .292, or some other value. It depends on the assumptions one makes, and is willing to defend. I don't think this is a reason to quit in frustration as I perceive some critics of regression-style structural modeling would have us do. Rather, I think it merely requires that one acknowledges the assumptions involved in any statistical application, and addresses them appropriately in the design of the research.

References

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*. Cambridge: Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.

Costner, H.L. (1969). Theory, deduction, and rules of correspondence. *American Journal of Sociology*, *75*, 245-263.

Duncan, O.D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Goldberger, A.S. (1964). *Econometric theory*. New York: John Wiley & Sons, Inc.

Hu, L-T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jöreskog, K.G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, *24*, 387-404.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381-389.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International, Inc.

Jöreskog, K.G., & Sörbom, D. (1996). *PRELIS 2: User's reference guide* (3rd ed.). Chicago: Scientific Software International, Inc.

Jöreskog, K.G., Sörbom, D., duToit, S., & duToit, M. (1999). *LISREL 8: New statistical features*. Chicago: Scientific Software International, Inc.

Kowalski, C.J. (1973). Non-normal bivariate distributions with normal marginals. *American Statistician*, *27*, 103-106.

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

Namboodiri, N.K., Carter, L.F., & Blalock, H.M., Jr. (1975). *Applied multivariate analysis and experimental designs*. New York: McGraw-Hill Book Company.

Pascarella, E.T., & Terenzini, P.T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass Publishers.

Riccobono, J., Henderson, L.B., Burkheimer, G.J., Place, C., & Levinsohn, J.R. (1981). *National longitudinal study: Base year (1972) through fourth follow-up (1979) data file user's manual*. Washington, D.C.: National Center for Education Statistics.

Rigdon, E.E. (1994). Demonstrating the effects of unmodeled random measurement error. *Structural Equation Modeling*, *1*, 375-380.

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

Schumacker, R.E., & Lomax, R.G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Siegel, P.M., & Hodge, R.M. (1968). A causal approach to the study of measurement error. In H.M. Blalock, Jr. & A.B. Blalock (Eds.), *Methodology in social research.* New York: McGraw-Hill Book Company.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18*, 161-169.

Steigler, J.H., & Lind, J.C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA, May.

Walker, H.M., & Lev, J. (1953). *Statistical inference.* New York: Holt, Rinehart and Winston.

Werts, C.E., Rock, D.A., Linn, R.L., & Jöreskog, K.G. (1976). Comparison of correlations, variances, covariances, and regression weights with or without measurement error. *Psychological Bulletin, 83*, 1007-1013.

Wolfle, L.M. (1979). Unmeasured variables in path analysis. *Multiple Linear Regression Viewpoints, 9*(1), 20-56.

Wolfle, L.M. (1982). Causal models with unmeasured variables: An introduction to LISREL. *Multiple Linear Regression Viewpoints, 11*(2), 9-54.

Wonnacott, R.J., & Wonnacott, T.H. (1970). *Econometrics.* New York: John Wiley & Sons.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics, 5*, 161-215.