Psychological Methods
2001, Vol. 6, No. 3, 258–269

1082-989X/01/$5.00 DOI: 10.1037//1082-989X.6.3.258

Testing Parameters in Structural Equation Modeling: Every "One" Matters

Richard Gonzalez
University of Michigan

Dale Griffin
University of British Columbia

A problem with standard errors estimated by many structural equation modeling programs is described. In such programs, a parameter's standard error is sensitive to how the model is identified (i.e., how scale is set). Alternative but equivalent ways to identify a model may yield different standard errors, and hence different Z tests for a parameter, even though the identifications produce the same overall model fit. This lack of invariance due to model identification creates the possibility that different analysts may reach different conclusions about a parameter's significance level even though they test equivalent models on the same data. The authors suggest that parameters be tested for statistical significance through the likelihood ratio test, which is invariant to the identification choice.

Structural equations modeling (SEM) is a powerful and flexible tool for data analysis that proceeds in roughly three stages (for introductions to SEM, see Bollen, 1989; Hoyle, 1995). In Stage 1, the analyst postulates a structural model, possibly with latent variables. In Stage 2, the analyst compares the overall fit of the model to the data. In Stage 3, the analyst evaluates the specific parameters of the model. Problems need to be solved at each stage. For example, when postulating the structural model, one should consider alternative but equivalent models (e.g., Breckler, 1990); when testing the overall fit of the model, one should be sensitive to issues such as sample size and the meaning of several fit indices (Bollen, 1989; Hu & Bentler, 1995); and when evaluating the individual parameters, the scale of the input data should be considered (i.e., the correlation matrix problem; see Cudeck, 1989). It is well-known that decisions made in Stage 1 about how to identify the model do not influence the results of Stage 2—equivalent ways to identify a model produce identical fits. What is not as well-known is that the seemingly arbitrary identification decisions made in Stage 1 have important implications for the parameter tests in Stage 3.

In this article, we show that the standard errors estimated in many SEM programs may not be invariant under equivalent ways to identify a model. *Model identification* refers to the scaling restrictions imposed on a subset of parameters to allow estimation (e.g., fixing a path coefficient to 1 or fixing a latent variance to 1). As we show, a parameter's standard error, and hence its corresponding statistical test, can vary dramatically across different identifications, even though all other indices indicate identical model fits. An analogous observation was made in the behavioral genetics literature by Neale and his colleagues (Neale, Heath, Hewitt, Eaves, & Fulker, 1989; Neale & Miller, 1997) and in the econometrics literature by Dagenais and Dufour (1991).

In the first section of the article we illustrate this problem with a specific example. We then clarify why the standard errors and corresponding Z tests of most SEM programs fail to be invariant, and we point to a test that is invariant across different model identifications. The article presents a relatively intuitive description of the problem, designed to be accessible to a nontechnical audience.

The research reported in this article was partially supported by National Science Foundation Grant 98-09261. We thank Richard Bagozzi, Steve Boker, Tony Greenwald, Rowell Huesmann, David Kenny, and Rod McDonald for their helpful discussion and suggestions. William Venables provided geometric insights into the differences between the likelihood ratio test and the Wald test.

Correspondence concerning this article should be addressed to Richard Gonzalez, Department of Psychology, University of Michigan, Ann Arbor, Michigan 48109, or to Dale Griffin, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada. Electronic mail may be sent to gonzo@umich.edu or to dale.griffin@commerce.ubc.ca.

Different Z Tests for Identical Models

To illustrate how the standard error and Z test of a parameter can fluctuate as a function of equivalent identifications, we present a simplified model from Wheaton, Muthen, Alwin, and Summers (1977) that has become a standard example in the SEM literature. The observed covariance matrix is presented in Table 1; sample size $N = 932$.

In this simple model there are two latent variables, each with two indicators; there are also equality constraints on the error variances, $V(E1) = V(E3) = \phi_1$ and $V(E2) = V(E4) = \phi_2$, and the loadings. The model and data are taken from the EQS manual (Bentler, 1995). The critical parameter for our illustration is the covariance between the two latent variables (ϕ_{12}). Restrictions are needed in order to set the scale of the latent variables. The analyst may fix the variance of each of the latent variables to a constant or may fix indicator coefficients to a constant (by convention, this constant is chosen to be 1; e.g., Bollen, 1989). We refer to such restrictions as *model identifications.* Figures 1 and 2 illustrate the model under two different identifications, or scale restrictions. The model in Figure 1 forces one indicator loading to 1 within each latent variable and treats the latent variances as parameters to estimate, whereas the model in Figure 2 forces the latent variances to 1 and treats the indicator loadings as parameters. Except for this difference in identification, the two models are equivalent, and both models produce identical fits to the observed covariance matrix.

The two models represented in Figures 1 and 2 were estimated using the observed covariance matrix with CALIS (SAS Institute, Inc., 1989); the parameter estimates are presented in Table 2. Comparable results are obtained with LISREL (Jöreskog & Sörbom, 1989), SEPATH (StatSoft, 1997), AMOS (Arbuckle & Wothke, 1999), EQS, and RAMONA (Browne & Mels, 1996). Maximum-likelihood estimation was used to compute the estimates in Table 2, but the points made in this section also apply to other estimation procedures such as generalized least squares (GLS).

Table 1
Data From Wheaton et al. (1977)

Variable (V)	V1	V2	V3	V4
V1	11.834	0.660	0.560	0.440
V2	6.947	9.364	0.470	0.520
V3	6.819	5.091	12.532	0.670
V4	4.783	5.028	7.495	9.986

Note. Variances are shown on the diagonal, covariances in the lower half, and correlations in the upper half ($N = 932$). Data are from "Assessing Reliability and Stability in Panel Models," by B. Wheaton, B. Muthen, D. F. Alwin, & G. F. Summers, in *Sociological Metholodology* (Vol. 8, pp. 84–136), by D. R. Heise (Ed.), 1977, Oxford, England: Blackwell. Copyright 1977 by Blackwell Publishers. Reprinted with permission. (Data also presented by Bentler, 1995.)

These two methods of model identification produced identical model fits. The models in Figures 1 and 2 both had a goodness-of-fit index = .998 and $\chi^2(4, N = 932) = 2.969, p = .563$. Both identifications produced the identical model-implied covariance matrix. Thus, any index based on the model-implied covariance matrix will be invariant across these two identifications.[1]

There are two notable differences in the covariance estimate between the two latent variables presented in Table 2. One difference is the scale of the parameter. Obviously, when the identification was implemented by fixing the latent variances to 1 (as in Figure 2), the covariance between the latent variables was identical to the latent correlation. By comparison, when the model identification was implemented by fixing an indicator coefficient to one (as in Figure 1), then the covariance between the two latent variables was estimated. In the latter case, the standardized covariance is identical to the correlation between the two latent variables (as shown in Table 2). The second difference is the topic of this article. Even though both models have identical global fits and the standardized parameter estimates are identical, the Z tests for the covariance between the two latent variables differ.[2]

[1] This result shows that the models under consideration are scale invariant (Browne, 1982). Scale invariance occurs when the model-implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ under parameter vector $\boldsymbol{\theta}$ is related to the model-implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*)$ under parameter vector $\boldsymbol{\theta}^*$ by the equation $\boldsymbol{\Sigma}(\boldsymbol{\theta}^*) = \mathbf{D}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}'$ for some symmetric matrix $\mathbf{D}$ and superscript t representing the transpose. Throughout this article the matrix $\mathbf{D}$ is the identity matrix and the two parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ arise from different identifications of the same model.

[2] There are other ways to identify this model, which also lead to the identical $\chi^2(4) = 2.969$. For example, take the model in Figure 1 and interchange which path within each latent variable is fixed to 1. That is, allow λ_1 to be a free parameter and fix λ_2 to 1. Under this identification, the Z test for the latent variable covariance ϕ_{12} is 11.7, which differs from the two Z tests for the same parameter reported in Table 2 of 12.83 and 26.61.

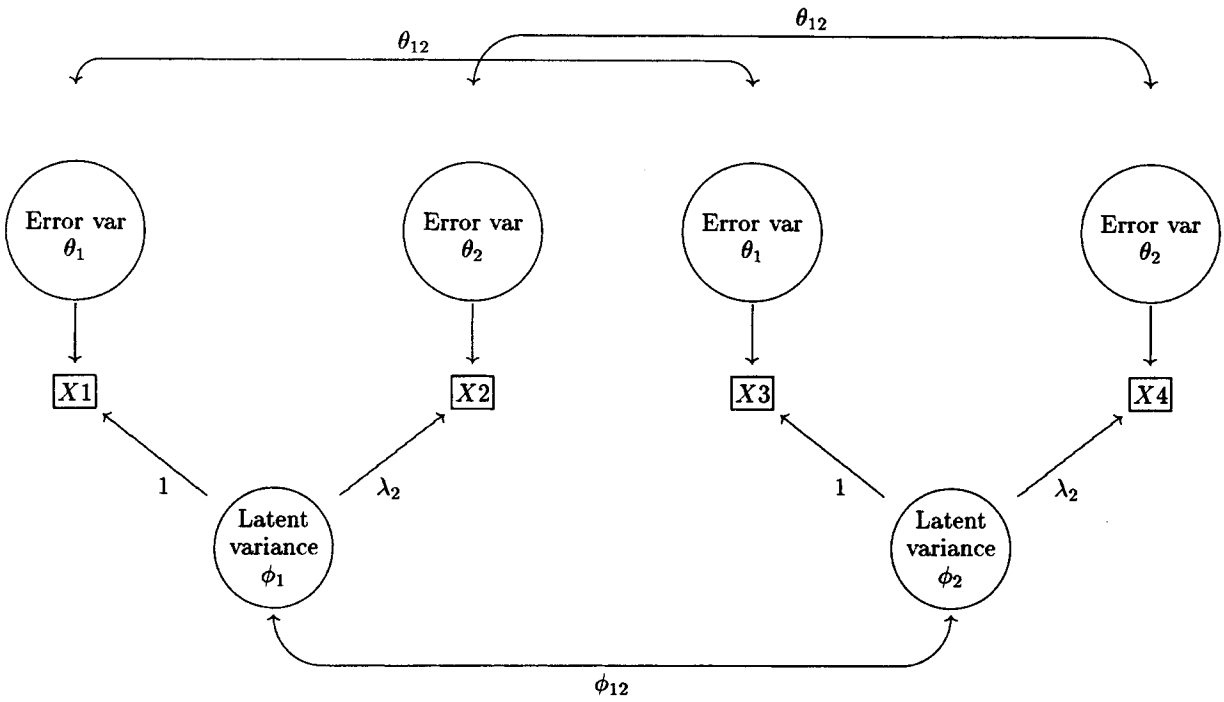

Figure 1. Two latent variables each with two indicators. The scale restriction is imposed in this model by fixing one indicator path (λ) to 1. The variances (var) of each latent variable as well as the covariance between them are estimated. In addition, the model constrains the latent variances $\phi_1 = \phi_2$ and the error covariances θ_{12}.

Although in this example the tests lead to the same action in rejecting the null hypothesis, with such dramatic differences in Z tests (26.61 vs. 12.83) one can imagine cases where two identifications can lead to identical model fits but different conclusions about the significance of a given parameter.

Because one model estimates "standardized" latent variables and the other does not, this may suggest to the reader that we have stumbled on a well-known problem—different results can be observed depending on whether the SEM analysis is performed on the covariance matrix or the correlation matrix (e.g., Cudeck, 1989; Jöreskog & Sörbom, 1989). As noted by McDonald, Parker, and Ishizuka (1993), the issue of model identification is related to the concept of scale invariance, but it is not identical to it. For instance, had the researcher supplied the observed correlation matrix instead of the observed convariance matrix and estimated the models in Figures 1 and 2, different Z tests across identifications would still be observed.

However, intuition gained from the proper analysis of correlation matrices can be applied to the present problem. Cudeck (1989) distinguished between scale invariant models and scale free parameters. In Cudeck's terminology, a parameter is scale free if (a) the model is scale invariant (see Footnote 1) and (b) for all choices of matrix **D** the value of the parameter remains the same. He showed that even for scale invariant models the standard errors for parameters that are not scale free will be incorrect when one uses a typical SEM program with a correlation matrix as input. Cudeck provided an example showing that a scale free parameter yields the same standard error regardless of whether the standard error is computed with the formula that corrects for the correlation matrix as input or the formula that assumes the covariance matrix as input (Cudeck, 1989, p. 323). His example contained six indicators and two latent factors. The variances of the latent variables were fixed to 1. The six indicator coefficients, the six error variances, and the covariance between the two latent variables were free parameters. The Z test for the latent variable covariance, the only scale free parameter in this model, was 8.32, regardless of whether or not the standard error was corrected for entering a correlation matrix.

We now extend Cudeck's (1989) example. When a different identification is used such that the latent variances are free to vary and the indicator coefficients are fixed, the Z test for the latent variable co-

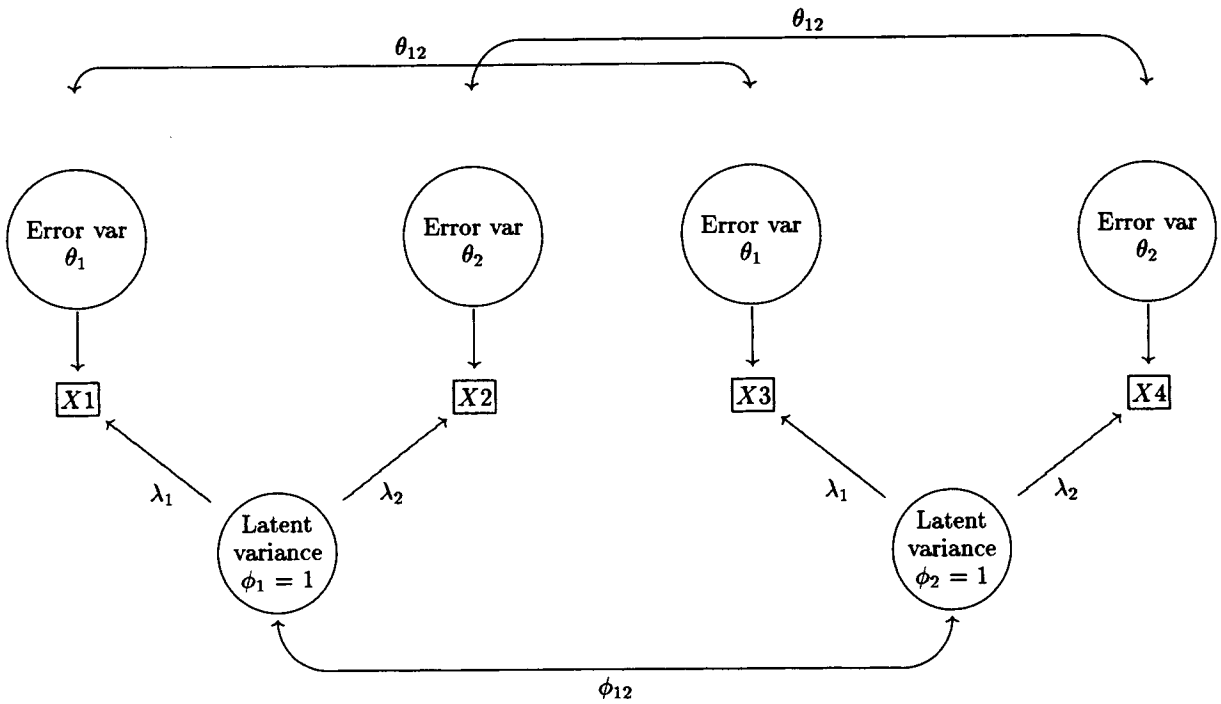

Figure 2. Two latent variables, each with two indicators (λ_1 and λ_2). The scale restriction is imposed in this model by fixing the latent variances (var) to 1. In addition, the model constrains the error covariances θ_{12}. ϕ = latent covariance.

variance is 5.06. Note that the difference in Z tests occurs on a scale free parameter even though both model identifications yield identical χ^2s (7.92) and degrees of freedom (8). Thus, the concern raised in this article is not with which input matrix is used but rather with how the model is identified.

The next section presents an informal, somewhat intuitive analysis of why the standard error differs across different identifications of a model. After presenting this intuitive explanation, we recommend a method for testing parameters that is invariant to model identification. The article ends with a more formal analysis of the problem and some simulation results.

Why Does the Z Test Vary Across Identification? An Intuitive Explanation

The reason Z tests vary across different model identifications is because the standard errors are typically

Table 2
CALIS Results From Applying the Models in Figures 1 and 2 to the Covariances in Table 1

Parameter	Model in Figure 1			Model in Figure 2		
	Estimate	*SE*	*Z*	Estimate	*SE*	*Z*
Loading λ_1	(0.84)			2.94 (0.84)	0.09	32.86
Loading λ_2	0.83 (0.79)	0.03	24.49	2.46 (0.79)	0.08	30.36
Latent variance $\phi_1 = \phi_2$	8.65	0.53	16.43			
Error variance θ_1	3.53	0.27	13.23	3.53	0.27	13.23
Error variance θ_2	3.65	0.21	17.65	3.65	0.21	17.65
Error covariance θ_{12}	0.91	0.12	7.45	0.91	0.12	7.45
Covariance ϕ_{12}	5.91 (0.68)	0.46	12.83	0.68 (0.68)	0.03	26.61

Note. The critical row in this table is the covariance between the two latent variables ϕ_{12}, which have different values for the Z test. Standardized parameter estimates appear in parentheses.

computed using an approximation that is influenced by how the model is identified. Intuitively, the approximation measures the curvature of the fit function evaluated against parameter estimates (Buse, 1982; Edwards, 1972). The fit function is the index the program optimizes, such as the likelihood function or GLS. The standard measure of curvature is based on the second derivative, which is the rate of change of the fit function given a change in the parameter of interest, and leads to an asymptotic estimate of a parameter's standard error (see, e.g., Bollen, 1989; Silvey, 1975).

The identification of the model directly influences this measure of curvature—different identifications can produce different estimates of curvature. To illustrate, we refer to Figure 3, which plots a hypothetical fit function against one parameter; analogous figures appeared in Buse (1982) and in Bollen (1989). Imagine that there are four ways to identify the identical model, and each identification yields the same value for the fit function. That is, the four fit functions attain the same maximum height (for instance, the same χ^2 value) at different parameter values because the four identifications place the parameters on different scales. Note that the curvature at the maximum point differs for the four identification methods. These four fit functions would then produce four different estimates of the standard error. If model identification influences the scale of the parameter and the curvature of the fit function in a nonlinear way, then the Z test will differ across identical identification methods. We show in the formal section below that the impact of model identification can be understood through the "Jacobian matrix"; there we see how the off-diagonal terms of the Jacobian matrix explain the lack of identification invariance.

This intuitive analysis of the problem suggests a solution. Rather than approximating the value of the fit function by a measure of curvature that is sensitive to model identification, one could evaluate the fit function directly both when the parameter is freely estimated (i.e., maximum height) and when the parameter is restricted to the value of the null hypothesis. That is, one could estimate the model twice: once with the parameter of interest free to vary and again with the parameter constrained to the null value. Note that the four identifications cross at the value of the fit function corresponding to the null hypothesis. This yields two values for each fit function that are the same across identifications: the value at the null and the maximum height. Geometrically, this vertical difference between the two model fits remains unchanged when the parameter space is rescaled, even though the curvature is sensitive to the particular parameterization (see Buse, 1982).

In the context of SEM, the test based on curvature

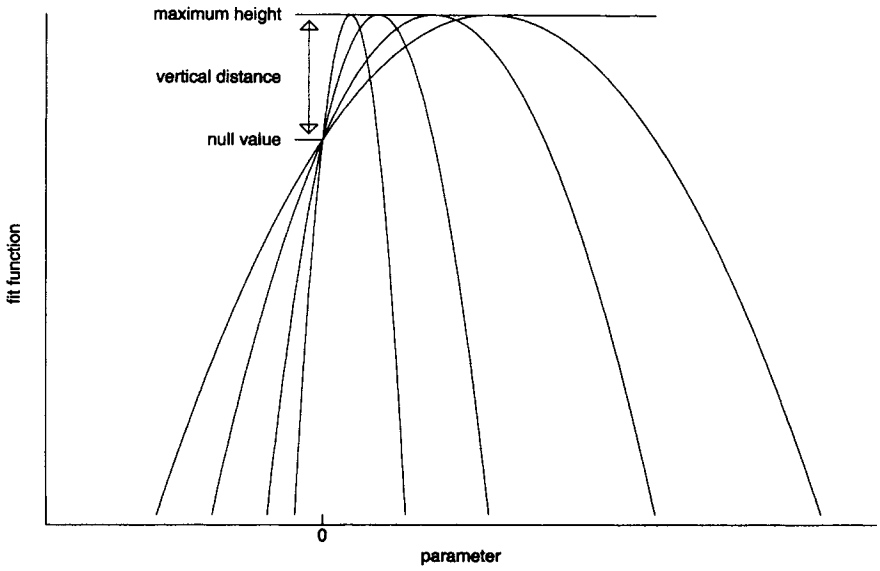

Figure 3. Hypothetical fit function with one parameter under four different model identifications. The four identifications lead to the same maximum height, and they have the same value under the null hypothesis. The identifications differ, however, in their curvature near the maximum height.

is the Wald test (Wald, 1943) and the difference between the two fit functions is the likelihood ratio test. Satorra (1989) proved, among other things, that under the null hypothesis, the Wald test is asymptotically equivalent to the likelihood ratio test. However, it is well-known that the Wald test is not invariant across different model identifications, or parameterizations. Satorra was aware of this point: "Different parameterizations may also lead to substantial changes on the quality of these approximations (and note that in general W [the Wald test] and GW [the generalized Wald test] are not invariant under such reparameterization)" (p. 147).

We suggest that investigators avoid the Z test printed in most SEM outputs. Instead, we recommend the likelihood ratio test as an alternative procedure for testing individual parameters because it is invariant to different methods of imposing scale restrictions for model identification. We now turn to a discussion of the likelihood ratio test.

Illustrating the Likelihood Ratio Test

In SEM models, the likelihood ratio test can be implemented by taking the difference of the χ^2 fit for the "reduced model" and the χ^2 fit for the "full model," and comparing this difference to a χ^2 distribution with degrees of freedom equal to the difference in the number of free parameters. In the present context, the *full model* refers to the model that allows the parameter of interest to be estimated, and the *reduced model* refers to the model that fixes the parameter of interest to the value of the null hypothesis. Dagenais and Dufour (1991) discussed a family of tests based on Neyman's $C(\alpha)$ test as an alternative to the likelihood ratio test that is invariant to model identification under special conditions. In this article we focus on the likelihood ratio test because it is relatively easy for a user to implement in current SEM programs.

We illustrate the invariance of the likelihood ratio test by returning to the two-factor, two-indicator example in Figures 1 and 2. Recall that both models produced identical χ^2 values of 2.969, yet the Z tests for the parameter of interest ϕ_{12}, the covariance between the two latent variables, differed across the two identifications. If both identifications depicted in Figures 1 and 2 are reestimated as reduced models with the value of ϕ_{12} fixed to zero (the value of the null hypothesis), then both reduced models produce the identical χ^2 of 343.25. Thus, the likelihood ratio test for this example is $343.25 - 2.969 = 340.29$, which follows a χ^2 distribution with one degree of freedom. Recall that when there is one degree of freedom, the square root of the χ^2 is normally distributed; thus the Z test corresponding to the null hypothesis test of ϕ_{12} is 18.45 for both identifications. Neither identification produced a Wald Z test for ϕ_{12} (see Table 2) that was close to the Z given by the likelihood ratio test. For an example of an empirical article using the likelihood ratio test to evaluate individual parameters, see Bentler and Speckart (1981).

Unfortunately, the likelihood ratio test will not work in all cases because sometimes the reduced model cannot be fit. One reason is that for some situations setting a parameter to zero produces estimation problems (see Kenny's, 1979, related discussion of *empirical underidentification*). In such cases, one could derive the error term symbolically (see Bollen, 1989, pp. 107–110 and Appendix 4B, for examples) or use nonstandard numerical procedures (Silvey, 1975, Sections 4.7.5 and 7.4.1). These alternatives are not easy to implement in most standard SEM packages and require additional derivation or programming on the part of the user. The likelihood ratio test should not be performed when one parameter is tested at a boundary (such as when testing a variance at 0) because in that situation the statistic can be distributed as a mixture of χ^2s (see, e.g., Self & Liang, 1987).

We now turn to a more formal analysis of the effects of identification. The next section shows that a Jacobian matrix can explain how identification influences the standard errors of individual parameters.

Formal Analysis

To simplify our treatment, we focus on maximum likelihood estimation because it is widely used, but the argument can be extended to GLS as well as frameworks with less stringent distributional assumptions (see Bentler, 1995, for references). To keep the notation simple, we do not distinguish between population parameters and their sample estimates.

We consider two matrices. One is the covariance matrix of parameter estimates—that is, the matrix that has the parameter variances in the diagonal and parameter covariances in the off-diagonal. The parameter correlation matrices for the two identifications on the example data appear in Table 3. Following standard theory, the parameter covariance matrix is proportional to the inverse of the Hessian matrix. The Hessian matrix has in the diagonal the second derivative of the fit function with respect to each parameter,

Table 3
The Parameter Correlation Matrices for the Two Identifications in Figures 1 and 2

Parameter	θ_1	θ_2	θ_{12}	ϕ_1	λ_2	ϕ_{12}
			Figure 1			
θ_1	—	−.45	.22	−.49	.65	−.31
θ_2		—	.27	.31	−.59	.21
θ_{12}			—	−.07	.01	−.16
ϕ_1				—	−.61	.88
λ_2					—	−.42
ϕ_{12}						—
	θ_1	θ_2	θ_{12}	λ_1	λ_2	ϕ_{12}
			Figure 2			
θ_1	—	−.45	.22	−.49	.35	.16
θ_2		—	.27	.31	−.44	−.07
θ_{12}			—	−.07	−.05	−.22
λ_1				—	.17	.21
λ_2					—	.32
ϕ_{12}						—

and the off-diagonal cells contain the partial derivatives of all pairwise combinations of parameters. The definition of the parameter covariance matrix appears in many treatments of SEM (e.g., Bollen, 1989). For an introduction to relevant theory and theorems, see Mood (1950) or Silvey (1975); for a distinction between the observed and expected Hessian matrices, which play an important role below, see Efron and Hinkley (1978); and for a discussion of numerical approximation, see Dolan and Molenaar (1991).

The second matrix that we need is the Jacobian matrix of the mapping that converts one model identification to another. The Jacobian matrix is created by taking the first derivative of a vector of functions with respect to a vector of parameters. In the present application we consider the vector of functions that maps the parameters in one identification to corresponding parameters in a different identification. For instance, the indicator coefficient λ_1 for the identification given in Figure 2 is equivalent to the square root of the variance of the latent variable ϕ_1 for the identification given in Figure 1. In addition, the covariance between the two latent variables in Figure 2 is identical to the standardized covariance in Figure 1, that is,

$$\phi'_{12} = \frac{\phi_{12}}{\phi_1}, \tag{1}$$

where the parameter with a prime is from the identification in Figure 2 and parameters without a prime are from the identification of Figure 1. The mapping functions between each pair of parameters are listed in Table 4.

The Jacobian matrix is, in this example, the 6 × 6 matrix of partial derivatives of each of the six functions in Table 4 with respect to each of the six parameters. The Jacobian matrix for this example is as follows:

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2\sqrt{\phi_1}} & 0 & 0 \\ 0 & 0 & 0 & \frac{\lambda_2}{2\sqrt{\phi_1}} & \sqrt{\phi_1} & 0 \\ 0 & 0 & 0 & -\frac{\phi_{12}}{\phi_1^2} & 0 & \frac{1}{\sqrt{\phi_1}} \end{matrix}$$

The *ij*th entry in this matrix is the partial derivative of the *i*th function from Table 4 with respect to the *j*th parameter. Note that the structure of the Jacobian matrix does not depend on data but on the mapping that relates the two model identifications to each other.

Let **J** represent the Jacobian matrix and **V** represent the covariance matrix of the parameters under the identification depicted in Figure 1 (i.e., the square root of the diagonal of **V** contains the standard errors of the parameters). The covariance matrix for the pa-

Table 4
The Mapping of Parameters From Figure 1 to the Parameters in Figure 2

Parameter in Figure 1	Mapping from Figure 1 to Figure 2
θ_1	$\theta'_1 = f_1(\theta_1,\theta_2,\theta_{12},\phi_1,\phi_2,\phi_{12}) = \theta_1$
θ_2	$\theta'_2 = f_2(\theta_1,\theta_2,\theta_{12},\phi_1,\phi_2,\phi_{12}) = \theta_2$
θ_{12}	$\theta'_{12} = f_3(\theta_1,\theta_2,\theta_{12},\phi_1,\phi_2,\phi_{12}) = \theta_{12}$
ϕ_1	$\lambda'_1 = f_4(\theta_1,\theta_2,\theta_{12},\phi_1,\phi_2,\phi_{12}) = \sqrt{\phi_1}$
ϕ_2	$\lambda'_2 = f_5(\theta_1,\theta_2,\theta_{12},\phi_1,\phi_2,\phi_{12}) = \lambda_2\sqrt{\phi_1}$
ϕ_{12}	$\phi'_{12} = f_6(\theta_1,\theta_2,\theta_{12},\phi_1,\phi_2,\phi_{12}) = \phi_{12}/\phi_1$

Note. Each parameter i has a multivariate function f_i that relates it to the parameter vector in the other identification. The first column lists the six parameters in Figure 1, and the second column gives the mapping of the parameters from the model in Figure 1 to the parameters from the model in Figure 2.

rameters under the identification in Figure 2 is a function of the parameter covariance matrix for the identification in Figure 1 and the Jacobian matrix:

$$\mathbf{V}' = \mathbf{J}\mathbf{V}\mathbf{J}^t, \tag{2}$$

where the superscript t denotes matrix transposition and the prime denotes the new identification. The square root of the diagonal of $\mathbf{V}'$ contains the standard errors of the parameters under Figure 2, and the square root of the diagonal of $\mathbf{V}$ contains the standard errors of the parameters under Figure 1. Thus, intuitively, the Jacobian matrix is what relates the standard errors in one identification to the standard errors in the other. The theory underlying Equation 2 was reviewed in Edwards (1972) and Mood (1950). For a more technical definition of equivalent models under reidentification, see Luijben (1991).

Take, for instance, parameter *i*. If the *i*th row of $\mathbf{J}$ contains zeros everywhere except for a constant at the diagonal position *ii*, then the Z test for parameter i will be invariant across the two model identifications. For example, in Table 4, the parameters θ_1 and θ_2 will have identical Z tests across these two identifications (as well as identical parameter estimates and standard errors). This will be true even when the constant at position *ii* is not 1 because the scaling constant will cancel when forming the Z ratio.

However, if the *i*th row of $\mathbf{J}$ contains nonzero elements anywhere other than at position *ii*, then the Z test for parameter i will not, in general, be invariant across model identifications. For instance, let $\mathbf{V}_{ij}$ refer to row i and column j of parameter covariance matrix $\mathbf{V}$ (i.e., from the identification in Figure 1). The standard error of ϕ_{12} under the identification in Figure 1 is $\sqrt{\mathbf{V}_{66}}$, whereas the standard error of ϕ_{12} under the identification in Figure 2 will be given by $\sqrt{\mathbf{V}'_{66}}$, which can be expressed as

$$\sqrt{\frac{\phi_{12}^2\mathbf{V}_{44} - 2\phi_{12}\phi_1\mathbf{V}_{46} + \phi_1^2\mathbf{V}_{66}}{\phi_1^4}}, \tag{3}$$

where ϕ_1, ϕ_2, ϕ_{12}, and $\mathbf{V}_{ij}$ refer to estimates from the identification in Figure 1. The terms other than $\mathbf{V}_{66}$ are introduced because of the nonzero elements in $\boldsymbol{J}$. Thus, the structure of the Jacobian matrix tells us when a parameter test will be invariant across two parameterizations and when it will not.

Note that under the null hypothesis that the population value of $\phi_{12} = 0$, Equation 3 reduces to

$$\sqrt{\frac{\mathbf{V}_{66}}{\phi_1^2}}. \tag{4}$$

Thus, when the terms in Equation 3 that go to zero under the null hypothesis are manually forced to zero, then the two identifications yield identical standard errors (up to a scaling constant that cancels when constructing the ratio for the Z test). Most SEM programs do not force each parameter to the value of the null hypothesis when computing the standard errors. This means that off-diagonal terms in the Jacobian matrix could influence the standard error and the Z test. The likelihood ratio test is invariant because it compares the model fit when the parameter is free to the model fit when the parameter is restricted to the value of the null hypothesis. This forces all "identification relevant" terms (up to scaling constants) to drop out of the standard error, making the likelihood test invariant to model identification.

The observation that manually forcing the parameter to the value of the null hypothesis yields tests that are invariant to identification is important for another reason. The asymptotic result for the symbolic definition of the standard error is based on the expected Hessian matrix (recall that the inverse of the Hessian matrix is the parameter covariance matrix $\mathbf{V}$). Thus, when distributional assumptions hold, the null hypothesis is true, and sample size is large, the asymptotic result leads to the theoretical standard error. If sample size is small, then observed parameter estimates and their covariances may not necessarily cancel out. For example, in Equation 3 if the null hypothesis that in the population $\phi_{12}^2 = 0$ is true, then as sample size increased, the sample standard error would approach Equation 4 regardless of the model identification that was used.

Simulation Analysis

In their 1982 study, de Pijper and Saris conducted a simulation to study the issue of model identification in a one-factor model with four indicators (see Figure 4). Using LISREL, they compared three models that differed in their identification—fixing the latent variance to 1, fixing λ_1 to 1, and fixing λ_2 to 1. We replicated their simulation using their population parameters (N = 1,000) and estimated the model under four different identifications.

Table 5 presents the population values and maximum-likelihood parameter estimates for one sample of N = 1,000 under four different identifications: fixing the latent variance $\phi_x = 1$ or fixing any one of three indicator coefficients. The effects on the standard errors and Z tests of ϕ_x and λ_4 (the fourth indi-

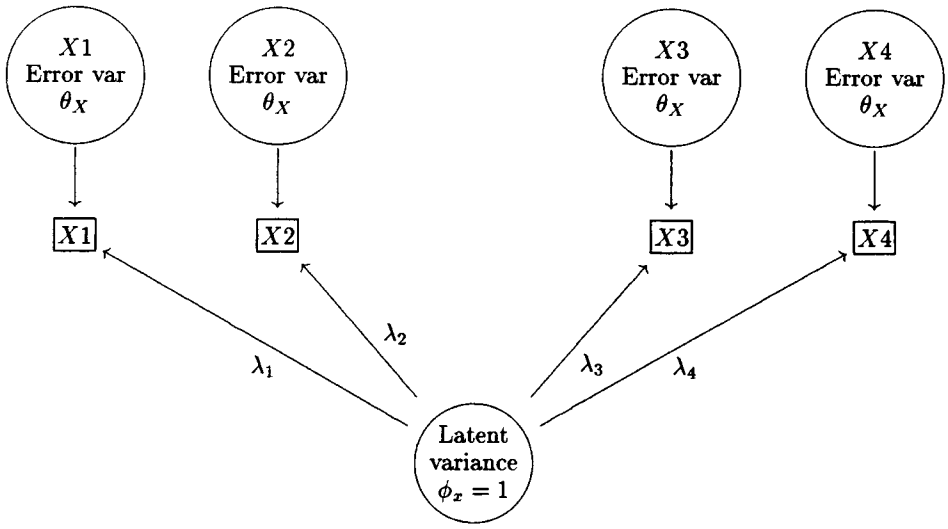

Figure 4. The one-factor model fit by de Pijper and Saris (1982). var = variance; θ = error variance; λ = indicator value; ϕ = latent covariance.

cator coefficient) are also presented. This allows one indicator λ_4 to remain a free parameter across the four example identifications. We verify what de Pijper and Saris (1982) observed in their simulation: *Z* tests can vary across different identifications even though the models have identical fits. Note that identification even influences the indicator coefficient λ_4, which is a free parameter in each of the four identifications. The likelihood ratio test for λ_4 is 15.99 and is identical across the four identifications. The likelihood ratio test for ϕ_x in the three models in which it is a free parameter is 37.39, but this test should be interpreted with caution because the likelihood ratio test is not necessarily distributed as a simple χ^2 when a parameter is tested at the boundary (e.g., a variance tested at 0 is at a boundary because variances cannot be negative; see Self & Liang, 1987).

Our previous analysis in terms of the Jacobian matrix extends the results of the simulation. Arbitrarily taking the model identification that fixes the latent variance $\phi_x = 1$ as the reference, the mapping of these parameters to the parameters from the identification that fixes one of the indicator coefficients $\lambda_i = 1$ and frees the latent variance ϕ_x involves the following three sets of functions: (a) dividing each of the λs in the reference identification by λ_i; (b) dividing the latent variance ϕ_x from the reference identification by λ_i^2; and (c) leaving the error variances, the θs, unchanged. For example, using primers to denote the parameter values under the new identification and

Table 5
Illustration of the One-Factor Model Simulated by de Pijper and Saris (1982): Maximum Likelihood Parameter Estimates

Parameter	Population value	Model 1	Model 2	Model 3	Model 4
λ_1	0.9	0.921 (0.028)	Fixed to 1	1.501 (0.077)	1.125 (0.044)
λ_2	0.6	0.613 (0.031)	0.666 (0.034)	Fixed to 1	0.749 (0.039)
λ_3	0.8	0.819 (0.029)	0.888 (0.035)	1.335 (0.069)	Fixed to 1
λ_4	0.5	0.512 (0.032)	0.555 (0.035)	0.834 (0.061)	0.625 (0.040)
ϕ	1	Fixed to 1	0.848	0.376	0.670
Z test for λ_4		16.16	15.89	13.75	15.79
Z test for ϕ		NA	16.23	9.98	14.03

Note. The model is described in Figure 4; maximum likelihood estimates were computed from one sample with $N = 1{,}000$. The population error variances for the four indicators were .19, .64, .36, and .75, respectively. The *Z* tests for the error variances were invariant across these four identifications and are not presented in this table. Standard errors appear in parentheses. NA = not applicable.

choosing to fix the indicator coefficient λ_1 for indicator X_1, the vector of mapping functions is $\lambda_1' = 1$, $\lambda_2' = \lambda_2/\lambda_1$, $\lambda_3' = \lambda_3/\lambda_1$, $\lambda_4' = \lambda_4/\lambda_1$, $\phi' = \phi/\lambda_1^2$, $\theta_1' = \theta_1$, $\theta_2' = \theta_2$, $\theta_3' = \theta_3$, $\theta_4' = \theta_4$. The Jacobian matrix of this vector is a 9×9 matrix of partial derivatives of each of the mapping functions with respect to each of the parameters in the reference identification. The effect on the standard errors of this new identification is completely determined by this Jacobian matrix.

Conclusions

Testing the individual parameters of a model should be routine in SEM.[3] As Bollen (1989) and others have argued, the assessment of a structural equation model should not stop with global indicators such as goodness of fit, but an analyst should also examine the parameter estimates, test each parameter for statistical significance, and evaluate parameter estimates against theoretical predictions or make comparisons across models.

The goal of this article was to heighten awareness that the Z test printed in the output of most SEM programs is not invariant to how the model was identified. We showed by example that a parameter's standard error, and hence its significance test, can be sensitive to the arbitrary choice of identification. This lack of identification invariance is another example of what is a growing list of problems associated with the usual SEM standard error and Wald test. Previous problems that have been discussed in the literature include the effect of small samples, nonnormal data, the scale of the input matrix, and nonmonotonic power functions (e.g., Chou & Bentler, 1990; Chou, Bentler, & Satorra, 1991; Cudeck, 1989; McDonald et al., 1993; Stone & Sobel, 1990). Together these results should lead SEM users to exercise caution when interpreting the statistical significance of a parameter.

The lack of invariance in the Wald test should not be interpreted as a reason to avoid using SEM. The Wald test's lack of invariance due to scale transformation (which is what identification in the context of SEM involves) is well-known and creates analogous problems in other areas of statistics (e.g., Dagenais & Dufour, 1991). Fortunately, for many structural equation models there is a way to test parameters in a manner that is invariant to the kind of model identification discussed here: the likelihood ratio test. Put simply, the likelihood ratio test proceeds as follows: (a) estimate the χ^2 for the full model with all parameters free, (b) estimate the χ^2 for the reduced model with all parameters free except that the parameter of interest is set to the value of the null hypothesis, and finally (c) test the difference in the χ^2 tests. The likelihood ratio test is quite flexible in that it can also be used to test multiple constraints simultaneously, and when a sequence of likelihood ratio tests is conducted on a sequence of nested models, the tests are asymptotically independent (Steiger, Shapiro, & Browne, 1985).

Although performing likelihood ratio tests is currently cumbersome for the user, we hope that SEM software will soon provide parameter tests that are invariant to model identification. In the meantime, users may want to perform likelihood ratio tests manually on parameters that are most central to their research rather than rely on the Z tests that appear in the output. The parameters that are most susceptible to the problem addressed in this article are covariances between latent variables and paths from one latent variable to another. Such parameters are usually of theoretical interest. The user should be aware that the measurement portion of the model (e.g., which indicator sets the scale of the latent variable) could have an impact on a parameter's standard error elsewhere in the model (e.g., the standard error of the path between the latent variable and a second latent variable).

One alternative that may occur to the reader is that one may perform a sensitivity analysis in the sense of trying different identifications to examine the effects of model identification on a parameter's significance test. If the test of the parameter is relatively insensitive to different parameterizations, then one may feel comfortable about interpreting the results of the Wald test. However, note that once the reader accepts the logic of fitting multiple models to test one parameter, then he or she might as well just perform the likelihood ratio test (which requires one additional model fit per parameter once the full model is fitted).

A large sample size is one way to reduce the invariance problem. However, it should be kept in mind that "large" usually means greater than $N = 500$ for

[3] We believe that a confidence interval approach may have more utility in scientific study. We chose to discuss hypothesis testing in this article because it is commonly used. The analogous invariance problem is also present for confidence intervals. An attempt to solve the identification problem, and also permit asymmetric confidence intervals, was made by Neale and Miller (1997).

relatively simple models. This sample size heuristic is for the case when the distributional assumptions hold, the null hypothesis is true, and the model has 15 or so free parameters. Even larger sample sizes may be required to bypass the lack of identification invariance when distributional assumptions do not hold or the model contains many free parameters. Rather than worrying about whether one's sample size is large enough to avoid the invariance problem given the details of one's model and data, we recommend the likelihood ratio test because it will always yield the identical result across identical identifications regardless of sample size.

Unfortunately, the likelihood ratio test cannot be computed in all cases because sometimes setting a parameter to zero can create a new problem such as empirical underidentification or require testing a parameter at its boundary. There are complicated options in these cases: One can derive standard errors manually, or one can use nonstandard numerical algorithms. These options are not straightforward to implement. Additional research is needed to find identification invariant Z tests for such cases, and these new methods should be implemented in standard software packages so that they are readily accessible to researchers.

References

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide.* Chicago: SmallWaters.

Bentler, P. M. (1995). *EQS structural equations program manual* [Computer manual]. Encino, CA: Multivariate Software.

Bentler, P. M., & Speckart, G. (1981). Attitudes "cause" behaviors: A structural equation analysis. *Journal of Personality and Social Psychology, 40,* 226–238.

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

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin, 107,* 260–273.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, England: Cambridge University Press.

Browne, M. W., & Mels, G. (1996). Path analysis: RAMONA. In L. Wilkinson (Ed.), *SYSTAT 6.0 for Windows* [Software manual]. Chicago: SPSS, Inc.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier tests: An expository note. *American Statistician, 36,* 153–157.

Chou, C., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research, 25,* 115–136.

Chou, C., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347–357.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327.

Dagenais, M. G., & Dufour, J. (1991). Invariance, nonlinear models, and asymptotic effects. *Econometrica, 59,* 1601–1615.

de Pijper, W. M., & Saris, W. E. (1982). The effect of identification restrictions on the test statistics in latent variable models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation* (pp. 175–182). Amsterdam: North-Holland.

Dolan, C. V., & Molenaar, P. C. M. (1991). A comparison of four methods of calculating standard errors of maximum-likelihood estimates in the analysis of covariance structure. *British Journal of Mathematical and Statistical Psychology, 44,* 359–368.

Edwards, A. W. F. (1972). *Likelihood.* Cambridge, England: Cambridge University Press.

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika, 65,* 457–487.

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

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications.* Chicago: SPSS, Inc.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Luijben, T. C. W. (1991). Equivalent models in covariance structure analysis. *Psychometrika, 56,* 653–665.

McDonald, R. P., Parker, P. M., & Ishizuka, T. (1993). A scale-invariant treatment for recursive path models. *Psychometrika, 58,* 431–443.

Mood, A. M. (1950). *Introduction to the theory of statistics.* New York: McGraw-Hill.

Neale, M. C., Heath, A., Hewitt, J., Eaves, L., & Fulker, D.

(1989). Fitting genetic models with LISREL: Hypothesis testing. *Behavior Genetics, 19,* 37–49.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27,* 113–120.

SAS Institute, Inc. (1989). *SAS/STAT user's guide.* Cary, NC: Author.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54,* 131–151.

Self, S. G., & Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82,* 605–610.

Silvey, S. D. (1975). *Statistical inference.* London: Chapman & Hall.

StatSoft, Inc. (1997). STATISTICA (Version 5.1) [Computer software]. Tulsa, OK: Author.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50,* 253–264.

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55,* 337–352.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426–482.

Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology* (Vol. 8, pp. 84–136). San Francisco: Jossey-Bass.

Received January 18, 2000
Revision received May 15, 2001
Accepted May 15, 2001 ■

New Editors Appointed, 2003–2008

The Publications and Communications Board of the American Psychological Association announces the appointment of five new editors for 6-year terms beginning in 2003.

As of January 1, 2002, manuscripts should be directed as follows:

- For the *Journal of Applied Psychology*, submit manuscripts to **Sheldon Zedeck, PhD,** Department of Psychology, University of California, Berkeley, CA 94720-1650.
- For the *Journal of Educational Psychology*, submit manuscripts to **Karen R. Harris, EdD,** Department of Special Education, Benjamin Building, University of Maryland, College Park, MD 20742.
- For the *Journal of Consulting and Clinical Psychology*, submit manuscripts to **Lizette Peterson, PhD,** Department of Psychological Sciences, 210 McAlester Hall, University of Missouri—Columbia, Columbia, MO 65211.
- For the *Journal of Personality and Social Psychology: Interpersonal Relations and Group Processes*, submit manuscripts to **John F. Dovidio, PhD,** Department of Psychology, Colgate University, Hamilton, NY 13346.
- For *Psychological Bulletin*, submit manuscripts to **Harris M. Cooper, PhD,** Department of Psychological Sciences, 210 McAlester Hall, University of Missouri—Columbia, Columbia, MO 65211.

Manuscript submission patterns make the precise date of completion of the 2002 volumes uncertain. Current editors, Kevin R. Murphy, PhD, Michael Pressley, PhD, Philip C. Kendall, PhD, Chester A. Insko, PhD, and Nancy Eisenberg, PhD, respectively, will receive and consider manuscripts through December 31, 2001. Should 2002 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2003 volumes.