An NCME Instructional Module on

Introduction to Structural Equation Modeling: Issues and Practical Considerations

Pui-Wa Lei and Qiong Wu, *The Pennsylvania State University*

Structural equation modeling (SEM) is a versatile statistical modeling tool. Its estimation techniques, modeling capacities, and breadth of applications are expanding rapidly. This module introduces some common terminologies. General steps of SEM are discussed along with important considerations in each step. Simple examples are provided to illustrate some of the ideas for beginners. In addition, several popular specialized SEM software programs are briefly discussed with regard to their features and availability. The intent of this module is to focus on foundational issues to inform readers of the potentials as well as the limitations of SEM. Interested readers are encouraged to consult additional references for advanced model types and more application examples.

Keywords: structural equation modeling, path model, measurement model

Structural equation modeling (SEM) has gained popularity across many disciplines in the past two decades due perhaps to its generality and flexibility. As a statistical modeling tool, its development and expansion are rapid and ongoing. With advances in estimation techniques, basic models, such as measurement models, path models, and their integration into a general covariance structure SEM analysis framework have been expanded to include, but are by no means limited to, the modeling of mean structures, interaction or nonlinear relations, and multilevel problems. The purpose of this module is to introduce the foundations of SEM modeling with the basic covariance structure models to new SEM researchers. Readers are assumed to have basic statistical knowledge in multiple regression and analysis of variance (ANOVA). References and other resources on current developments of more sophisticated models are provided for interested readers.

Pui-Wa Lei is an assistant professor, Department of Educational and School Psychology and Special Education, The Pennsylvania State University, 230 CEDAR Building, University Park, PA 16802; puiwa@psu.edu. Her primary research interests include structural equation modeling and item response theory. Qiong Wu is a doctoral student in the Department of Educational and School Psychology and Special Education, The Pennsylvania State University, 230 CEDAR Building, University Park, PA 16802; qiong@psu.edu. Her interests are measurement, statistical modeling, and high-stakes testing.

Series Information

ITEMS is a series of units designed to facilitate instruction in educational measurement. These units are published by the National Council on Measurement in Education. This module may be photocopied without permission if reproduced in its entirety and used for instructional purposes. Information regarding the development of new ITEMS modules should be addressed to: Dr. Mark Gierl, Canada Research Chair in Educational Measurement and Director, Centre for Research in Applied Measurement and Evaluation, Department of Educational Psychology, 6-110 Education North, University of Alberta, Edmonton, Alberta, Canada T6G 2G5.

What is Structural Equation Modeling?

Structural equation modeling is a general term that has been used to describe a large number of statistical models used to evaluate the validity of substantive theories with empirical data. Statistically, it represents an extension of general linear modeling (GLM) procedures, such as the ANOVA and multiple regression analysis. One of the primary advantages of SEM (vs. other applications of GLM) is that it can be used to study the relationships among latent constructs that are indicated by multiple measures. It is also applicable to both experimental and non-experimental data, as well as cross-sectional and longitudinal data. SEM takes a confirmatory (hypothesis testing) approach to the

multivariate analysis of a structural theory, one that stipulates causal relations among multiple variables. The *causal* pattern of intervariable relations within the theory is specified a priori. The goal is to determine whether a hypothesized theoretical model is consistent with the data collected to reflect this theory. The consistency is evaluated through *model-data fit*, which indicates the extent to which the postulated network of relations among variables is plausible. SEM is a large sample technique (usually $N > 200$; e.g., Kline, 2005, pp. 111, 178) and the sample size required is somewhat dependent on model complexity, the estimation method used, and the distributional characteristics of observed variables (Kline, pp. 14–15). SEM has a number of synonyms and special cases in the literature including path analysis, causal modeling, and covariance structure analysis. In simple terms, SEM involves the evaluation of two models: a *measurement* model and a *path* model. They are described below.

Path Model

Path analysis is an extension of multiple regression in that it involves various multiple regression models or equations that are estimated simultaneously. This provides a more effective and direct way of modeling mediation, indirect effects, and other complex relationship among variables. Path analysis can be considered a special case of SEM in which structural relations among observed (vs. latent) variables are modeled. Structural relations are hypotheses about directional influences or causal relations of multiple variables (e.g., how independent variables affect dependent variables). Hence, path analysis (or the more generalized SEM) is sometimes referred to as causal modeling. Because analyzing interrelations among variables is a major part of SEM and these interrelations are hypothesized to generate specific observed covariance (or correlation) patterns among the variables, SEM is also sometimes called covariance structure analysis.

In SEM, a variable can serve both as a source variable (called an *exogenous* variable, which is analogous to an independent variable) and a result variable (called an *endogenous* variable, which is analogous to a dependent variable) in a chain of causal hypotheses. This kind of variable is often called a *mediator*. As an example, suppose that family environment has a direct impact on learning motivation which, in turn, is hypothesized to affect achievement. In this case motivation is a mediator between family environment and achievement; it is the source variable for achievement and the result variable for family environment. Furthermore, feedback loops among variables (e.g., achievement can in turn affect family environment in the example) are permissible in SEM, as are reciprocal effects (e.g., learning motivation and achievement affect each other).[1]

In path analyses, observed variables are treated as if they are measured without error, which is an assumption that does not likely hold in most social and behavioral sciences. When observed variables contain error, estimates of path coefficients may be biased in unpredictable ways, especially for complex models (e.g., Bollen, 1989, p. 151–178). Estimates of reliability for the measured variables, if available, can be incorporated into the model to fix their error variances (e.g., squared standard error of measurement via classical test theory). Alternatively, if multiple observed variables that are supposed to measure the same latent constructs are available, then a measurement model can be used to separate the common variances of the observed variables from their error variances thus correcting the coefficients in the model for unreliability.[2]

Measurement Model

The measurement of latent variables originated from psychometric theories. Unobserved latent variables cannot be measured directly but are indicated or inferred by responses to a number of observable variables (indicators). Latent constructs such as intelligence or reading ability are often gauged by responses to a battery of items that are designed to tap those constructs. Responses of a study participant to those items are supposed to reflect where the participant stands on the latent variable. Statistical techniques, such as factor analysis, exploratory or confirmatory, have been widely used to examine the number of latent constructs underlying the observed responses and to evaluate the adequacy of individual items or variables as indicators for the latent constructs they are supposed to measure.

The measurement model in SEM is evaluated through confirmatory factor analysis (CFA). CFA differs from exploratory factor analysis (EFA) in that factor structures are hypothesized a priori and verified empirically rather than derived from the data. EFA often allows all indicators to load on all factors and does not permit correlated residuals. Solutions for different number of factors are often examined in EFA and the most sensible solution is interpreted. In contrast, the number of factors in CFA is assumed to be known. In SEM, these factors correspond to the latent constructs represented in the model. CFA allows an indicator to load on multiple factors (if it is believed to measure multiple latent constructs). It also allows residuals or errors to correlate (if these indicators are believed to have common causes other than the latent factors included in the model). Once the measurement model has been specified, structural relations of the latent factors are then modeled essentially the same way as they are in path models. The combination of CFA models with structural path models on the latent constructs represents the general SEM framework in analyzing covariance structures.

Other Models

Current developments in SEM include the modeling of mean structures in addition to covariance structures, the modeling of changes over time (growth models) and latent classes or profiles, the modeling of data having nesting structures (e.g., students are nested within classes which, in turn, are nested with schools; multilevel models), as well as the modeling of nonlinear effects (e.g., interaction). Models can also be different for different groups or populations by analyzing multiple sample-specific models simultaneously (multiple sample analysis). Moreover, sampling weights can be incorporated for complex survey sampling designs. See Marcoulides and Schumacker (2001) and Marcoulides and Moustaki (2002) for more detailed discussions of the new developments in SEM.

How Does SEM Work?

In general, every SEM analysis goes through the steps of model specification, data collection, model estimation, model

evaluation, and (possibly) model modification. Issues pertaining to each of these steps are discussed below.

Model Specification

A sound model is theory based. Theory is based on findings in the literature, knowledge in the field, or one's educated guesses, from which causes and effects among variables within the theory are specified. Models are often easily conceptualized and communicated in graphical forms. In these graphical forms, a directional arrow (→) is universally used to indicate a hypothesized causal direction. The variables to which arrows are pointing are commonly termed endogenous variables (or dependent variables) and the variables having no arrows pointing to them are called exogenous variables (or independent variables). Unexplained covariances among variables are indicated by curved arrows (⟷). Observed variables are commonly enclosed in rectangular boxes and latent constructs are enclosed in circular or elliptical shapes.

For example, suppose a group of researchers have developed a new measure to assess mathematics skills of preschool children and would like to find out (a) whether the skill scores measure a common construct called *math ability* and (b) whether reading readiness (RR) has an influence on math ability when age (measured in month) differences are controlled for. The skill scores available are: counting aloud (CA)—count aloud as high as possible beginning with the number 1; measurement (M)—identify fundamental measurement concepts (e.g., taller, shorter, higher, lower) using basic shapes; counting objects (CO)—count sets of objects and correctly identify the total number of objects in the set; number naming (NN)—read individual numbers (or shapes) in isolation and rapidly identify the specific number (shape) being viewed; and pattern recognition (PR)—identify patterns using short sequences of basic shapes (i.e., circle, square, and triangle). These skill scores (indicators) are hypothesized to indicate the strength of children's latent math ability, with higher scores signaling stronger math ability. Figure 1 presents the conceptual model.

The model in Figure 1 suggests that the five skill scores on the right are supposedly results of latent math ability (enclosed by an oval) and that the two exogenous observed variables on the left (RR and age enclosed by rectangles) are predictors of math ability. The two predictors (connected by ⟷) are allowed to be correlated but their relationship is not explained in the model. The latent "math ability" variable and the five observed skill scores (enclosed by rectangles) are endogenous in this example. The residual of the latent endogenous variable (residuals of structural equations are also called disturbances) and the residuals (or errors) of the skill variables are considered exogenous because their variances and interrelationships are unexplained in the model. The residuals are indicated by arrows without sources in Figure 1. The effects of RR and age on the five skill scores can also be perceived to be mediated by the latent variable (math ability). This model is an example of a multiple-indicator multiple-cause model (or MIMIC for short, a special case of general SEM model) in which the skill scores are the indicators and age as well as RR are the *causes* for the latent variable.

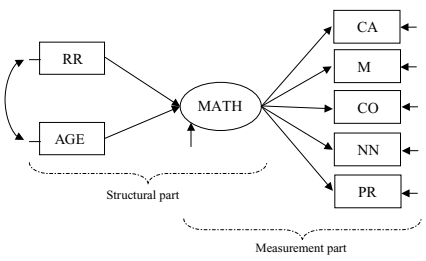

FIGURE 1. Conceptual model of math ability.

Due to the flexibility in model specification, a variety of models can be conceived. However, not all specified models can be identified and estimated. Just like solving equations in algebra where there cannot be more unknowns than knowns, a basic principle of identification is that a model cannot have a larger number of unknown parameters to be estimated than the number of unique pieces of information provided by the data (variances and covariances of observed variables for covariance structure models in which mean structures are not analyzed).[3] Because the scale of a latent variable is arbitrary, another basic principle of identification is that all latent variables must be scaled so that their values can be interpreted. These two principles are necessary for identification but they are not sufficient. The issue of model identification is complex. Fortunately, there are some established rules that can help researchers decide if a particular model of interest is identified or not (e.g., Davis, 1993; Reilly & O'Brien, 1996; Rigdon, 1995).

When a model is identified, every model parameter can be uniquely estimated. A model is said to be over-identified if it contains fewer parameters to be estimated than the number of variances and covariances, just-identified when it contains the same number of parameters as the number of variances and covariances, and under-identified if the number of variances and covariances is less than the number of parameters. Parameter estimates of an over-identified model are unique given a certain estimation criterion (e.g., maximum likelihood). All just-identified models fit the data perfectly and have a unique set of parameter estimates. However, a perfect model-data fit is not necessarily desirable in SEM. First, sample data contain random error and a perfect-fitting model may be fitting sampling errors. Second, because conceptually very different just-identified models produce the same perfect empirical fit, the models cannot be evaluated or compared by conventional means (model fit indices discussed below). When a model cannot be identified, either some model parameters cannot be estimated or numerous sets of parameter values can produce the same level of model fit (as in under-identified models). In any event, results of such models are not interpretable and the models require re-specification.

Data Characteristics

Like conventional statistical techniques, score reliability and validity should be considered in selecting measurement instruments for the constructs of interest and sample size needs to be determined preferably based on power considerations. The sample size required to provide unbiased parameter estimates and accurate model fit information for SEM models depends on model characteristics, such as model size as well as score characteristics of measured variables, such as score scale and distribution. For example, larger models require larger samples to provide stable parameter estimates, and larger samples are required for categorical

or non-normally distributed variables than for continuous or normally distributed variables. Therefore, data collection should come, if possible, after models of interest are specified so that sample size can be determined a priori. Information about variable distributions can be obtained based on a pilot study or one's educated guess.

SEM is a large sample technique. That is, model estimation and statistical inference or hypothesis testing regarding the specified model and individual parameters are appropriate only if sample size is not too small for the estimation method chosen. A general rule of thumb is that the minimum sample size should be no less than 200 (preferably no less than 400 especially when observed variables are not multivariate normally distributed) or 5–20 times the number of parameters to be estimated, whichever is larger (e.g., Kline, 2005, pp. 111, 178). Larger models often contain larger number of model parameters and hence demand larger sample sizes. Sample size for SEM analysis can also be determined based on a priori power considerations. There are different approaches to power estimation in SEM (e.g., MacCallum, Browne, & Sugawara, 1996 on the root mean square error of approximation (RMSEA) method; Satorra & Saris, 1985; Yung & Bentler, 1999 on bootstrapping; Muthén & Muthén, 2002 on Monte Carlo simulation). However, an extended discussion of each is beyond the scope of this module.

Model Estimation[4]

A properly specified structural equation model often has some fixed parameters and some free parameters to be estimated from the data. As an illustration, Figure 2 shows the diagram of a conceptual model that predicts reading (READ) and mathematics (MATH) latent ability from observed scores from two intelligence scales, verbal comprehension (VC) and perceptual organization (PO). The latent READ variable is indicated by basic word reading (BW) and reading comprehension (RC) scores. The latent MATH variable is indicated by calculation (CL) and reasoning (RE) scores. The visible paths denoted by directional arrows (from VC and PO to READ and MATH, from READ to BW and RC, and from MATH to CL and RE) and curved arrows (between VC and PO, and between residuals of READ and MATH) in the diagram are free parameters of the model to be estimated, as are residual variances of endogenous variables (READ, MATH, BW, RC, CL, and RE) and variances of exogenous variables (VC and PO). All other possible paths that are not shown (e.g., direct paths from VC or PO to BW, RC, CL, or RE) are fixed to zero and will not be estimated. As mentioned above, the scale of a latent variable is arbitrary and has to be set. The scale of a latent variable can be standardized by fixing its variance to 1. Alternatively, a latent variable can take the scale of one of its indicator variables by fixing the factor loading (the value of the path from a latent variable to an indicator) of one indicator to 1. In this example, the loading of BW on READ and the loading of CL on MATH are fixed to 1 (i.e., they become fixed parameters). That is, when the parameter value of a visible path is fixed to a constant, the parameter is not estimated from the data.

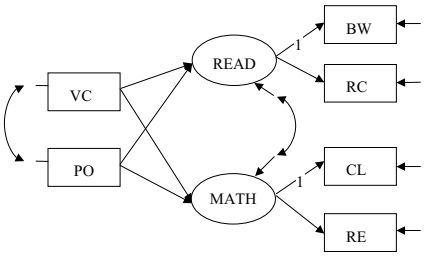

FIGURE 2. Conceptual model of math and reading.

Free parameters are estimated through iterative procedures to minimize a certain discrepancy or fit function between the observed covariance matrix (data) and the model-implied covariance matrix (model). Definitions of the discrepancy function depend on specific methods used to estimate the model parameters. A commonly used normal theory discrepancy function is derived from the maximum likelihood method. This estimation method assumes that the observed variables are multivariate normally distributed or there is no excessive kurtosis (i.e., same kurtosis as the normal distribution) of the variables (Bollen, 1989, p. 417).[5]

The estimation of a model may fail to converge or the solutions provided may be improper. In the former case, SEM software programs generally stop the estimation process and issue an error message or warning. In the latter, parameter estimates are provided but they are not interpretable because some estimates are out of range (e.g., correlation greater than 1, negative variance). These problems may result if a model is ill specified (e.g., the model is not identified), the data are problematic (e.g., sample size is too small, variables are highly correlated, etc.), or both. Multicollinearity occurs when some variables are linearly dependent or strongly correlated (e.g., bivariate correlation $> .85$). It causes similar estimation problems in SEM as in multiple regression. Methods for detecting and solving multicollinearity problems established for multiple regression can also be applied in SEM.

Model Evaluation

Once model parameters have been estimated, one would like to make a dichotomous decision, either to retain or reject the hypothesized model. This is essentially a statistical hypothesis-testing problem, with the null hypothesis being that the model under consideration fits the data. The overall model goodness of fit is reflected by the magnitude of discrepancy between the sample covariance matrix and the covariance matrix implied by the model with the parameter estimates (also referred to as the minimum of the fit function or $F_{\min}$). Most measures of overall model goodness of fit are functionally related to $F_{\min}$. The model test statistic $(N-1)F_{\min}$, where N is the sample size, has a chi-square distribution (i.e., it is a chi-square test) when the model is correctly specified and can be used to test the null hypothesis that the model fits the data. Unfortunately, this test statistic has been found to be extremely sensitive to sample size. That is, the model may fit the data reasonably well but the chi-square test may reject the model because of large sample size.

In reaction to this sample size sensitivity problem, a variety of alternative goodness-of-fit indices have been developed to supplement the chi-square statistic. All of these alternative indices attempt to adjust for the effect of sample size, and many of them also take into account model degrees of freedom, which is a proxy for model size. Two classes of alternative fit indices, incremental and absolute, have been identified (e.g., Bollen, 1989, p. 269; Hu & Bentler, 1999). Incremental fit indices measure the increase in fit relative

to a baseline model (often one in which all observed variables are uncorrelated). Examples of incremental fit indices include normed fit index (NFI; Bentler & Bonett, 1980), Tucker-Lewis index (TLI; Tucker & Lewis, 1973), relative noncentrality index (RNI; McDonald & Marsh, 1990), and comparative fit index (CFI; Bentler, 1989, 1990). Higher values of incremental fit indices indicate larger improvement over the baseline model in fit. Values in the .90s (or more recently $\geq .95$) are generally accepted as indications of good fit.

In contrast, absolute fit indices measure the extent to which the specified model of interest reproduces the sample covariance matrix. Examples of absolute fit indices include Jöreskog and Sörbom's (1986) goodness-of-fit index (GFI) and adjusted GFI (AGFI), standardized root mean square residual (SRMR; Bentler, 1995), and the RMSEA (Steiger & Lind, 1980). Higher values of GFI and AGFI as well as lower values of SRMR and RMSEA indicate better model-data fit.

SEM software programs routinely report a handful of goodness-of-fit indices. Some of these indices work better than others under certain conditions. It is generally recommended that multiple indices be considered simultaneously when overall model fit is evaluated. For instance, Hu and Bentler (1999) proposed a 2-index strategy, that is, reporting SRMR along with one of the fit indices (e.g., RNI, CFI, or RMSEA). The authors also suggested the following criteria for an indication of good model-data fit using those indices: RNI (or CFI) $\geq .95$, SRMR $\leq .08$, and RMSEA $\leq .06$. Despite the sample size sensitivity problem with the chi-square test, reporting the model chi-square value with its degrees of freedom in addition to the other fit indices is recommended.

Because some solutions may be improper, it is prudent for researchers to examine individual parameter estimates as well as their estimated standard errors. Unreasonable magnitude (e.g., correlation > 1) or direction (e.g., negative variance) of parameter estimates or large standard error estimates (relative to others that are on the same scale) are some indications of possible improper solutions.

If a model fits the data well and the estimation solution is deemed proper, individual parameter estimates can be interpreted and examined for statistical significance (whether they are significantly different from zero). The test of individual parameter estimates for statistical significance is based on the ratio of the parameter estimate to its standard error estimate (often called z-value or t-value). As a rough reference, absolute value of this ratio greater than 1.96 may be considered statistically significant at the .05 level. Although the test is proper for unstandardized parameter estimates, standardized estimates are often reported for ease of interpretation. In growth models and multiple-sample analyses in which different variances over time or across samples may be of theoretical interest, unstandardized estimates are preferred.

As an example, Table 1 presents the simple descriptive statistics of the variables for the math ability example (Figure 1), and Table 2 provides the parameter estimates (standardized and unstandardized) and their standard error estimates. This model fit the sample data reasonably well as indicated by the selected overall goodness-of-fit statistics: $\chi^2_{13} = 21.21$, $p = .069$, RMSEA = .056 (<.06), CFI = .99 (>.95), SRMR = .032 (<.08). The model solution is considered proper because there are no out-of-range parameter estimates and standard error estimates are of similar magnitude (see Table 2). All parameter estimates are considered large (not likely zero) because the ratios formed by unstandardized parameter estimates to their standard errors (i.e., z-values or t-values) are greater than |2| (Kline, 2005, p. 41). Standardized factor loadings in measurement models should fall between 0 and 1 with higher values suggesting better indications of the observed variables for the latent variable. All standardized loadings in this example are in the neighborhood of .7, showing that they are satisfactory indicators for the latent construct of math ability. Coefficients for the structural paths are interpreted in the same way as regression coefficients. The standardized coefficient value of .46 for the path from age to math ability suggests that as children grow by one standard deviation of age in months (about 6.7 months), their math ability is expected to increase by .46 standard deviation holding RR constant. The standardized value of .40 for the path from RR to math ability reveals that for every standard deviation increase in RR, math ability is expected to increase by .40 standard deviation, holding age constant. The standardized residual variance of .50 for the latent math variable indicates that approximately 50% of variance in math is unexplained by age and RR. Similarly, standardized residual or error variances of the math

Table 1. Sample Correlation, Mean, and Standard Deviation for the Model of Figure 1

	Variables						
Variables	**AGE**	**RR**	**CA**	**M**	**CO**	**NN**	**PR**
Age	1						
RR	.357	1					
CA	.382	.439	1				
M	.510	.405	.588	1			
CO	.439	.447	.512	.604	1		
NN	.513	.475	.591	.560	.606	1	
PR	.372	.328	.564	.531	.443	.561	1
Mean	50.340	.440	.666	.730	.545	.625	.624
SD	6.706	1.023	1.297	.855	.952	.933	1.196

Note: CA = counting aloud; M = measurement; CO = counting objects; NN = number naming; PR = pattern recognition; RR = reading readiness; age is measured in months.

Table 2. Parameter and Standard Error Estimates for the Model of Figure 1

Model Parameters	Standardized Estimate	Unstandardized Estimate	Standard Error
Loadings/effects on MATH			
CA	.74	1.00[a]	
M	.77	.68	.07
CO	.74	.73	.08
NN	.80	.77	.08
PR	.68	.84	.10
Age	.46	.07	.01
RR	.40	.38	.07
Residual variances			
CA	.45	.75	.10
M	.41	.30	.04
CO	.46	.41	.05
NN	.36	.32	.05
PR	.54	.78	.09
MATH	.50	.46	.09
Covariance			
RR and age	.36	2.45	.56

Note: Table values are maximum likelihood estimates. CA = counting aloud; M = measurement; CO = counting objects; NN = number naming; PR = pattern recognition; RR = reading readiness; age is measured in months.
[a]Fixed parameter to set the scale of the latent variable.

indicator variables are taken as the percentages of their variances unexplained by the latent variable.

Model Modification, Alternative Models, and Equivalent Models

When the hypothesized model is rejected based on goodness-of-fit statistics, SEM researchers are often interested in finding an alternative model that fits the data. Post hoc modifications (or model trimming) of the model are often aided by modification indices, sometimes in conjunction with the expected parameter change statistics. Modification index estimates the magnitude of decrease in model chi-square (for 1 degrees of freedom) whereas expected parameter change approximates the expected size of change in the parameter estimate when a certain fixed parameter is freely estimated. A large modification index (>3.84) suggests that a large improvement in model fit as measured by chi-square can be expected if a certain fixed parameter is freed. The decision of freeing a fixed parameter is less likely affected by chance if it is based on a large modification index as well as a large expected parameter change value.

As an illustration, Table 3 shows the simple descriptive statistics of the variables for the model of Figure 2, and Table 4 provides the parameter estimates (standardized and unstandardized) and their standard error estimates. Had one restricted the residuals of the latent READ and MATH variables to be uncorrelated, the model would not fit the sample data well as suggested by some of the overall model fit indices: $\chi^2_6 = 45.30, p < .01$, RMSEA = .17 (>.10), SRMR = .078 (acceptable because it is < .08). The solution was also

Table 3. Sample Correlation, Mean, and Standard Deviation for the Model of Figure 2

	Variables					
Variables	VC	PO	BW	RC	CL	RE
VC	1					
PO	.704	1				
BW	.536	.354	1			
RC	.682	.515	.781	1		
CL	.669	.585	.560	.657	1	
RE	.743	.618	.532	.688	.781	1
Mean	89.810	92.347	83.448	87.955	85.320	91.160
SD	15.234	18.463	14.546	15.726	15.366	15.092

Note: VC = verbal comprehension; PO = perceptual organization; BW = basic word reading; RC = reading comprehension; CL = calculation; RE = reasoning.

Table 4. Parameter and Standard Error Estimates for the Model of Figure 2

Model Parameters	Standardized Estimate	Unstandardized Estimate	Standard Error
Loadings/effects on READ			
BW	.79	1.00[a]	
RC	.99	1.35	.10
VC	.64	.48	.07
PO	.07	.04	.05
Loadings/effects on MATH			
CL	.85	1.00[a]	
RE	.92	1.05	.07
VC	.64	.55	.06
PO	.23	.16	.05
Residual variances			
BW	.38	79.55	10.30
RC	.02	5.28	11.96
CL	.27	64.22	8.98
RE	.16	36.77	7.88
READ	.52	69.10	10.44
MATH	.33	56.98	9.21
Covariances			
VC and PO	.70	198.05	24.39
Residuals of READ and MATH	.21	31.05	6.31

Note: Table values are maximum likelihood estimates. VC = verbal comprehension; PO = perceptual organization; BW = basic word reading; RC = reading comprehension; CL = calculation; RE = reasoning.
[a]Fixed parameter to set the scale of the latent variable.

improper because there was a negative error variance estimate. The modification index for the covariance between the residuals of READ and MATH was 33.03 with unstandardized expected parameter change of 29.44 (standardized expected change = .20). There were other large modification indices. However, freeing the residual covariance between READ and MATH was deemed most justifiable because the relationship between these two latent variables was not likely fully explained by the two intelligence subtests (VC and PO). The modified model appeared to fit the data quite well ($\chi^2_5 = 8.63, p = .12$, RMSEA = .057, SRMR = .017). The actual chi-square change from 45.30 to 8.63 (i.e., 36.67) was slightly different from the estimated change (33.03), as was the actual parameter change (31.05 vs. 29.44; standardized value = .21 vs. .20). The differences between the actual and estimated changes are slight in this illustration because only one parameter was changed. Because parameter estimates are not independent of each other, the actual and expected changes may be very different if multiple parameters are changed simultaneously, or the order of change may matter if multiple parameters are changed one at a time. In other words, different final models can potentially result when the same initial model is modified by different analysts.

As a result, researchers are warned against making a large number of changes and against making changes that are not supported by strong substantive theories (e.g., Byrne, 1998, p. 126). Changes made based on modification indices may not lead to the "true" model in a large variety of realistic situations (MacCallum, 1986; MacCallum, Roznowski, & Necowitz, 1992). The likelihood of success of post hoc modification depends on several conditions: It is higher if the initial model is close to the "true" model, the search continues even when a statistically plausible model is obtained, the search is restricted to paths that are theoretically justifiable, and the sample size is large (MacCallum, 1986). Unfortunately, whether the initially hypothesized model is close to the "true" model is never known in practice. Therefore, one can never be certain that the modified model is closer to the "true" model.

Moreover, post hoc modification changes the confirmatory approach of SEM. Instead of confirming or disconfirming a theoretical model, modification searches can easily turn modeling into an exploratory expedition. The model that results from such searches often capitalizes on chance idiosyncrasies of sample data and may not generalize to other samples (e.g., Browne & Cudeck, 1989; Tomarken & Waller, 2003). Hence, not only is it important to explicitly account for the specifications made post hoc (e.g., Tomarken & Waller, 2003), but it is also crucial to cross-validate the final model with independent samples (e.g., Browne & Cudeck, 1989).

Rather than data-driven post hoc modifications, it is often more defensible to consider multiple alternative models a priori. That is, multiple models (e.g., based on competing theories or different sides of an argument) should be specified prior to model fitting and the best fitting model is selected among the alternatives. Jöreskog (1993) discussed different modeling strategies more formally and referred to the practice of post hoc modification as *model generating*, the consideration of different models a priori as *alternative models*, and the rejection of the misfit hypothesized model as *strictly confirmatory*.

As models that are just-identified will fit the data perfectly regardless of the particular specifications, different just-identified models (sub-models or the entire model) detailed for the same set of variables are considered equivalent. Equivalent models may be very different in implications but

produce identical model-data fit. For instance, predicting verbal ability from quantitative ability may be equivalent to predicting quantitative ability from verbal ability or to equal strength of reciprocal effects between verbal and quantitative ability. In other words, the direction of causal hypotheses cannot be ruled out (or determined) on empirical grounds using cross-sectional data but on theoretical foundations, experimental control, or time precedence if longitudinal data are available. See MacCallum, Wegener, Uchino, and Fabrigar (1993) and Williams, Bozdogan, and Aiman-Smith (1996) for more detailed discussions of the problems and implications of equivalent models. Researchers are encouraged to consider different models that may be empirically equivalent to their selected final model(s) before they make any substantial claims. See Lee and Hershberger (1990) for ideas on generating equivalent models.

Causal Relations

Although SEM allows the testing of causal hypotheses, a well fitting SEM model does not and cannot prove causal relations without satisfying the necessary conditions for causal inference, partly because of the problems of equivalent models discussed above. The conditions necessary to establish causal relations include time precedence and robust relationship in the presence or absence of other variables (see Kenny, 1979, and Pearl, 2000, for more detailed discussions of causality). A selected well-fitting model in SEM is like a retained null hypothesis in conventional hypothesis testing. It remains plausible among perhaps many other models that are not tested but may produce the same or better level of fit. SEM users are cautioned not to make unwarranted causal claims. Replications of findings with independent samples are essential especially if the models are obtained based on post hoc modifications. Moreover, if the models are intended to be used in predicting future behaviors, their utility should be evaluated in that context.

Software Programs

Most SEM analyses are conducted using one of the specialized SEM software programs. However, there are many options, and the choice is not always easy. Below is a list of the commonly used programs for SEM. Special features of each program are briefly discussed. It is important to note that this list of programs and their associated features is by no means comprehensive. This is a rapidly changing area and new features are regularly added to the programs. Readers are encouraged to consult the web sites of software publishers for more detailed information and current developments.

LISREL

LISREL (linear structural relationships) is one of the earliest SEM programs and perhaps the most frequently referenced program in SEM articles. Its version 8 (Jöreskog & Sörbom, 1996a, 1996b) has three components: PRELIS, SIMPLIS, and LISREL. PRELIS (pre-LISREL) is used in the data preparation stage when raw data are available. Its main functions include checking distributional assumptions, such as univariate and multivariate normality, imputing data for missing observations, and calculating summary statistics, such as Pearson covariances for continuous variables, polychoric or polyserial correlations for categorical variables, means, or asymptotic covariance matrix of variances and covariances (required for asymptotically distribution-free estimator or Satorra and Bentler's scaled chi-square and robust standard errors; see footnote 5). PRELIS can be used as a stand-alone program or in conjunction with other programs. Summary statistics or raw data can be read by SIMPLIS or LISREL for the estimation of SEM models. The LISREL syntax requires the understanding of matrix notation while the SIMPLIS syntax is equation-based and uses variable names defined by users. Both LISREL and SIMPLIS syntax can be built through interactive LISREL by entering information for the model construction wizards. Alternatively, syntax can be built by drawing the models on the Path Diagram screen. LISREL 8.7 allows the analysis of multilevel models for hierarchical data in addition to the core models. A free student version of the program, which has the same features as the full version but limits the number of observed variables to 12, is available from the web site of Scientific Software International, Inc. (http://www.ssicentral.com). This web site also offers a list of illustrative examples of LISREL's basic and new features.

EQS

Version 6 (Bentler, 2002; Bentler & Wu, 2002) of EQS (Equations) provides many general statistical functions including descriptive statistics, *t*-test, ANOVA, multiple regression, nonparametric statistical analysis, and EFA. Various data exploration plots, such as scatter plot, histogram, and matrix plot are readily available in EQS for users to gain intuitive insights into modeling problems. Similar to LISREL, EQS allows different ways of writing syntax for model specification. The program can generate syntax through the available templates under the "Build_EQS" menu, which prompts the user to enter information regarding the model and data for analysis, or through the Diagrammer, which allows the user to draw the model. Unlike LISREL, however, data screening (information about missing pattern and distribution of observed variables) and model estimation are performed in one run in EQS when raw data are available. Model-based imputation that relies on a predictive distribution of the missing data is also available in EQS. Moreover, EQS generates a number of alternative model chi-square statistics for nonnormal or categorical data when raw data are available. The program can also estimate multilevel models for hierarchical data. Visit http://www.mvsoft.com for a comprehensive list of EQS's basic functions and notable features.

Mplus

Version 3 (Muthén & Muthén, 1998–2004) of the Mplus program includes a Base program and three add-on modules. The Mplus Base program can analyze almost all single-level models that can be estimated by other SEM programs. Unlike LISREL or EQS, Mplus version 3 is mostly syntax-driven and does not produce model diagrams. Users can interact with the Mplus Base program through a language generator wizard, which prompts users to enter data information and select the estimation and output options. Mplus then converts the information into its program-specific syntax. However, users have to supply the model specification in Mplus language themselves. Mplus Base also offers a robust option for non-normal data and a special full-information maximum likelihood estimation method for missing data (see footnote 4). With the

add-on modules, Mplus can analyze multilevel models and models with latent categorical variables, such as latent class and latent profile analysis. The modeling of latent categorical variables in Mplus is so far unrivaled by other programs. The official web site of Mplus (http://www.statmodel.com) offers a comprehensive list of resources including basic features of the program, illustrative examples, online training courses, and a discussion forum for users.

Amos

Amos (analysis of moment structure) version 5 (Arbuckle, 2003) is distributed with SPSS (SPSS, Inc., 2006). It has two components: Amos Graphics and Amos Basic. Similar to the LISREL Path Diagram and SIMPLIS syntax, respectively, Amos Graphics permits the specification of models by diagram drawing whereas Amos Basic allows the specification from equation statements. A notable feature of Amos is its capability for producing bootstrapped standard error estimates and confidence intervals for parameter estimates. An alternative full-information maximum likelihood estimation method for missing data is also available in Amos. The program is available at http://www.smallwaters.com or http://www.spss.com/amos/.

Mx

Mx (Matrix) version 6 (Neale, Boker, Xie, & Maes, 2003) is a free program downloadable from http://www.vcu.edu/mx/. The Mx Graph version is for Microsoft Windows users. Users can provide model and data information through the Mx programming language. Alternatively, models can be drawn in the drawing editor of the Mx Graph version and submitted for analysis. Mx Graph can calculate confidence intervals and statistical power for parameter estimates. Like Amos and Mplus, a special form of full-information maximum likelihood estimation is available for missing data in Mx.

Others

In addition to SPSS, several other general statistical software packages offer built-in routines or procedures that are designed for SEM analyses. They include the CALIS (covariance analysis and linear structural equations) procedure of SAS (SAS Institute Inc., 2000; http://www.sas.com/), the RAMONA (reticular action model or near approximation) module of SYSTAT (Systat Software, Inc., 2002; http://www.systat.com/), and SEPATH (structural equation modeling and path analysis) of Statistica (StatSoft, Inc., 2003; http://www.statsoft.com/products/advanced.html).

Summary

This module has provided a cursory tour of SEM. Despite its brevity, most relevant and important considerations in applying SEM have been highlighted. Most specialized SEM software programs have become very user-friendly, which can be either a blessing or a curse. Many SEM novices believe that SEM analysis is nothing more than drawing a diagram and pressing a button. The goal of this module is to alert readers to the complexity of SEM. The journey in SEM can be exciting because of its versatility and yet frustrating because the first ride in SEM analysis is not necessarily smooth for everyone. Some may run into data problems, such as missing data, non-normality of observed variables, or multicollinearity; estimation problems that could be due to data problems or identification problems in model specification; or interpretation problems due to unreasonable estimates. When problems arise, SEM users will need to know how to troubleshoot systematically and ultimately solve the problems. Although individual problems vary, there are some common sources and potential solutions informed by the literature. For a rather comprehensive list of references by topics, visit http://www.upa.pdx.edu/IOA/newsom/semrefs.htm. Serious SEM users should stay abreast of the current developments as SEM is still growing in its estimation techniques and expanding in its applications.

Self-Test

1. List at least two advantages of structural equation models over regular ANOVA and regression models.
2. In a conventional SEM diagram, Y1→Y2 means:

   A. Y2 is an exogenous variable.
   B. Variation in Y2 is affected by variation in Y1.
   C. Variation in Y1 is explained within the model.
   D. Y1 and Y2 are correlated.

3. What is NOT included in a path model?

   A. Exogenous variable
   B. Endogenous variable
   C. Latent variable
   D. Disturbance

4. What is a mediator in a path model?
5. Which of the following statements regarding path models is NOT correct?

   A. It is an extension of multiple regressions.
   B. It can be used to examine mediation effects.
   C. Measurement errors are well accounted for in path models.
   D. It can be used to examine causal hypotheses between independent and dependent variables.

6. A researcher wants to study the effect of SES on students' classroom behavior and would like to find out whether SES affects classroom behavior directly or indirectly by way of its direct effect on academic achievement. Which of the following models is the most appropriate for this study?

   A. Confirmatory factor analysis model
   B. Path model
   C. Growth model
   D. Multiple-sample model

7. What are some possible ways of accounting for measurement errors of observed variables in SEM?
8. What does it mean for a model to be identified?

   A. The model is theory based.
   B. Fit indices for the model are all satisfactory.
   C. An SEM program successfully produces a solution for the model.
   D. There exists a unique solution for every free parameter in the model.

9. In order for a model to be identified, we must set the variances of the latent variables to 1. T/F

10. The model test statistic, $(N-1)F_{\min}$, has a chi-square distribution, where N is the sample size and $F_{\min}$ is minimum of the fit function. T/F
11. Even if a model fits the data quite well, it is still very likely that we will get a significant chi-square value. T/F
12. Which of the following model fit indices is the most sensitive to sample size?

    A. Chi-square
    B. RMSEA
    C. SRMR
    D. CFI

13. Which of the following most likely suggests bad model fit?

    A. High values in incremental fit indices
    B. High value in GFI
    C. High value in model chi-square
    D. High value in RMSEA

14. Which of the following suggests improper solution to the model?

    A. Nonsignificant residual variance
    B. Negative disturbance variance
    C. Negative covariance
    D. Small correlation

15. Which of the following statements about modification index is true?

    A. It indicates the amount of change in parameter estimates if we free a certain fixed parameter.
    B. Model modification decisions must be based on modification index values.
    C. All parameters that have modification index values larger than 3.84 should be freed.
    D. It indicates the amount of change in chi-square if we free a certain fixed parameter.

16. One of the advantages of using SEM is that causal relations among variables can be established using cross-sectional data. T/F

Key to Self-Test

1. SEM can include both observed and latent variables, and relationships among latent constructs can be examined; several dependent variables can be studied in a single SEM analysis; equation residuals can be correlated in SEM.
2. B
3. C
4. A mediator is a variable that serves as both a source variable and a result variable. In other words, it affects, and is also affected by, some other variables within the model.
5. C
6. B
7. There are two possible ways: (1) incorporate reliability coefficients, if applicable, by fixing error variances to squared standard errors of measurement; (2) incorporate measurement models when multiple indicators of the constructs are available.
8. D
9. F (The scale of a latent variable can be set in two ways: standardizing or fixing the factor loading of one indicator to 1.)
10. F (It has a chi-square distribution only when the model is correctly specified and other model assumptions are satisfied.)
11. T (Holding model-data discrepancy constant, model chi-square value increases as sample size increases.)
12. A
13. D (Both high values of model chi-square and RMSEA indicate bad model fit but chi-square is also affected by sample size, so the best option is RMSEA.)
14. B
15. D
16. F (There are certain necessary conditions for causal relations that are not likely satisfied by cross-sectional studies.)

Acknowledgment

We thank James DiPerna, Paul Morgan, and Marley Watkins for sharing the data used in the illustrative examples, as well as Hoi Suen and the anonymous reviewers for their constructive comments on a draft of this article.

Notes

[1]When a model involves feedback or reciprocal relations or correlated residuals, it is said to be nonrecursive; otherwise the model is recursive. The distinction between recursive and nonrecursive models is important for model identification and estimation.

[2]The term error variance is often used interchangeably with unique variance (that which is not common variance). In measurement theory, unique variance consists of both "true unique variance" and "measurement error variance," and only measurement error variance is considered the source of unreliability. Because the two components of unique variance are not separately estimated in measurement models, they are simply called "error" variance.

[3]This principle of identification in SEM is also known as the *t*-rule (Bollen, 1989, p. 93, p. 242). Given the number of p observed variables in any covariance-structure model, the number of variances and covariances is $p(p+1)/2$. The parameters to be estimated include factor loadings of measurement models, path coefficients of structural relations, and variances and covariances of exogenous variables including those of residuals. In the math ability example, the number of observed variances and covariances is $7(8)/2 = 28$ and the number of parameters to be estimated is 15 (5 loadings + 2 path coefficients + 3 variance–covariance among predictors + 6 residual variances – 1 to set the scale of the latent factor). Because 28 is greater than 15, the model satisfies the *t*-rule.

[4]It is not uncommon to have missing observations in any research study. Provided data are missing completely at random, common ways of handling missing data, such as imputation, pairwise deletion, or listwise deletion can be applied. However, pairwise deletion may create estimation problems for SEM because a covariance matrix that is computed based on different numbers of cases may be singular or some estimates may be out-of-bound. Recent versions of some SEM software programs offer a special maximum likelihood estimation method (referred to as full-information maximum likelihood), which uses all available data for estimation and requires no imputation. This option is logically appealing because there is no need to make additional assumptions for imputation and there is no loss of observations. It has also

been found to work better than listwise deletion in simulation studies (Kline, 2005, p. 56).

[5]When this distributional assumption is violated, parameter estimates may still be unbiased (if the proper covariance or correlation matrix is analyzed, that is, Pearson for continuous variables, polychoric, or polyserial correlation when categorical variable is involved) but their estimated standard errors will likely be underestimated and the model chi-square statistic will be inflated. In other words, when the distributional assumption is violated, statistical inference may be incorrect. Other estimation methods that do not make distributional assumptions (e.g., the asymptotically distribution-free estimator or weighed least squares based on the full asymptotic variance–covariance matrix of the estimated variances and covariances) are available but they often require unrealistically large sample sizes to work satisfactorily ($N > 1{,}000$). When the sample size is not that large, a viable alternative is to request robust estimation from some SEM software programs (e.g., LISREL8, EQS6, Mplus3), which provides some adjustment to the chi-square statistic and standard error estimates based on the severity of non-normality (Satorra & Bentler, 1994). Statistical inference based on adjusted statistics has been found to work quite satisfactorily provided sample size is not too small.

References

Arbuckle, J. L. (2003). *Amos 5* (computer software). Chicago, IL: Smallwaters.

Bentler, P. M. (1989). *EQS Structural Equations Program Manual*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bentler, P. M. (2002). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

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. (2002). *EQS 6 for Windows User's Guide*. Encino, CA: Multivariate Software, Inc.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons, Inc.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*, 445–455.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Davis, W. R. (1993). The FC1 rule of identification for confirmatory factor analysis. *Sociological Methods and Research*, *21*, 403–437.

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

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Lang (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage.

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.

Jöreskog, K. G., & Sörbom, D. (1996a). *PRELIS: User's reference guide*. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1996b). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Kenny, D. A. (1979). *Correlation and causality* (chapters 3–5). New York: Wiley.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Lee, S., & Hershberger, S. L. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, *25*, 313–334.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107–120.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185–199.

Marcoulides, G. A., & Moustaki, I. (2002). *Latent variable and latent structure models.* Mahwah, NJ: Lawrence Erlbaum Associates.

Marcoulides, G. A., & Schumacker, R. E. (Eds.) (2001). *New developments and techniques in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247–255.

Muthén, L. K., & Muthén, B. O. (1998–2004). *Mplus user's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *4*, 599–620.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Reilly, T., & O'Brien, R. M. (1996). Identification of confirmatory factor analysis models of arbitrary complexity: The side-by-side rule. *Sociological Methods and Research*, *24*, 473–491.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavior Research*, *30*, 359–383.

SAS Institute, Inc. (2000). *The SAS System for Windows (release 9.01)* (computer software). Cary, NC: SAS Institute, Inc.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.

SPSS, Inc. (2006). SPSS Version 14.0 [Software Program]. Chicago: SPSS Inc.

StatSoft, Inc. (2003). *STATISTICA (version 6.1)* (computer software). Tulsa, OK: StatSoft, Inc.

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

Systat Software, Inc. (2002). *SYSTAT (Version 10.2)* (computer software). Richmond, CA: Systat Software, Inc.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*, 578–598.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Williams, L. J., Bozdogan, H., & Aiman-Smith, L. (1996). Inference problems with equivalent models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 279–314). Mahwah, NJ: Erlbaum.

Yung, Y.-F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.