

Review Article

Core reporting practices in structural equation modeling

James B. Schreiber, Ph.D.*

Duquesne University, School of Education, 600 Forbes Avenue, Pittsburgh, PA 15282, USA

Abstract

Background: Structural equation modeling (SEM) is a popular analysis technique because of the wide range of questions that it can help answer. There are several pieces of information specific to SEM that should be reported when this technique is used.

Objectives: To demonstrate a basic framework for reporting SEM analyses, to provide definitions of key terms readers will encounter, and to illustrate 2 examples for reporting SEM results.

Methods: Data from 650 participants who completed 3 self-report surveys were used to test a confirmatory factor analysis and a structural model as examples of information to be reported.

Results: The results displayed are requisite information for any SEM analysis.

Conclusions: It is important for investigators to provide this information so that readers can properly evaluate the results and conclusions based on the analyses.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Structural equation modeling; Statistical reporting; Research methods

Introduction

Structural equation modeling (SEM), or latent variable analysis, is a useful technique because a wide variety of data, research designs, and theoretical models can be analyzed. Many social science researchers are interested in measuring and understanding the relationships among latent, unobservable, variables. The purpose of this paper is to provide a basic description of important information that should be provided in the report of any SEM analysis. Included are 2 examples of SEM analysis reports and a brief description of common statistical packages used in SEM.

The language of SEM

A firm understanding of any analysis technique begins with an ability to comprehend the basic terminology used. For SEM, readers must be able to interpret the common graphical representations of theorized models presented. There are 2 basic types of variables: unobserved and observed. *Unobserved variables* called, *latent factors*, *factors*, or *constructs* are graphically depicted with circles or ovals (Fig. 1). *Common factor* is another term used for latent factors because of the shared effects in common with 1 or more observed variables.

* Tel.: +1 412 396 1081; fax: +1 412 396 1681.

E-mail address: schreiberj@duq.edu.

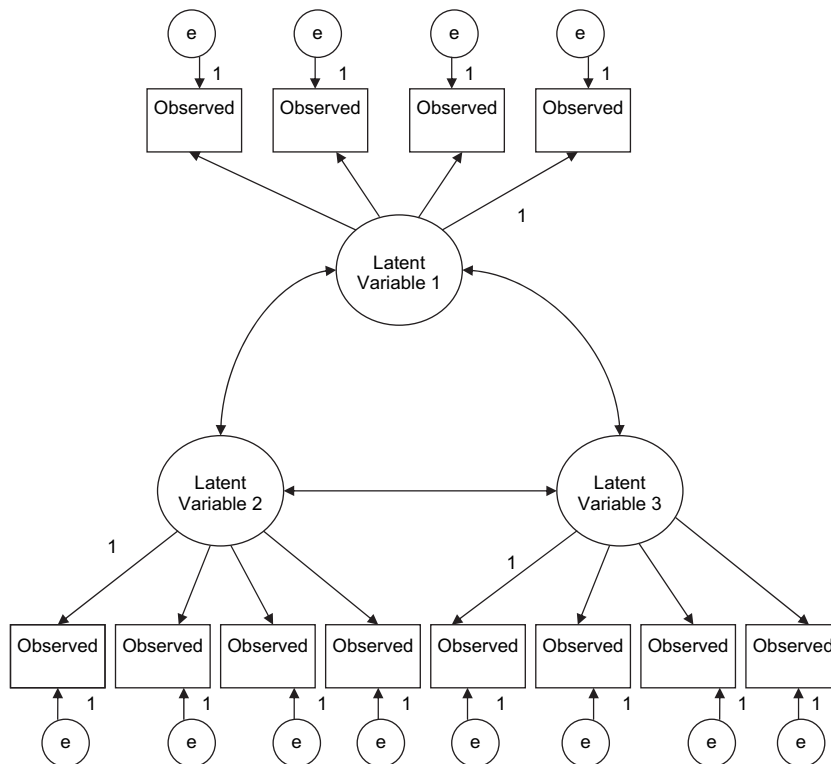


Fig. 1. Generic CFA example of 3 latent variables and 12 observed variables.

Observed variables are also termed *measured*, *indicator*, or *manifest*. A square or rectangle is the traditional graphic for observed variables (Fig. 1). In Fig. 1, the large circles represent the latent constructs and the small circles represent the unique factors—measurement errors—in the observed variable or disturbances in equation, measurement, or both. *Errors* are variance in the responses that are not explained by the latent construct. Single head and dual head arrows are called *paths*. Single-head arrows (paths) represent directional effects from 1 variable (latent or observed) to another and dual-head arrows represent a correlation or relationship. An example of an unobserved variable is Life Satisfaction (LS) and an observed variable is the response to a particular question related to LS, such as “I enjoy spending time with my family.”

In the diagram, the regression coefficient, the direction and magnitude of the relationship, is set at 1 for each path. Because error is unobserved, there is no specific measurement unit. As such, a value must be set, and a value of 1.0 is most common. The value could be set to a different number to scale the measurement error; however,

most software programs are designed so that 1 is the default or easiest setting.

SEM generally encompasses 2 components: a measurement model (ie, confirmatory factor analysis, CFA) and a structural model (Figs. 1 and 2, respectively). In this context, a model is the theoretical relationships among the variables. The measurement model specifies the relationship of the latent to the observed variables, whereas the structural model (directional paths in bold in Fig. 2) identifies specific relationships among the latent variables.¹ Two other terms associated with SEM are *exogenous*, similar to independent variables, and *endogenous*, similar to dependent/outcome variables. Path-analytic models, which can also be thought of as only the structural analysis, are not discussed in this paper. But, the information that should be provided in those reports is the same.

The premise of SEM is to determine if a theoretical model is supported by the data collected. Mathematically, it is the comparison of an observed covariance matrix and a reproduced covariance matrix. The reproduced matrix is based on the mathematical equations derived from the

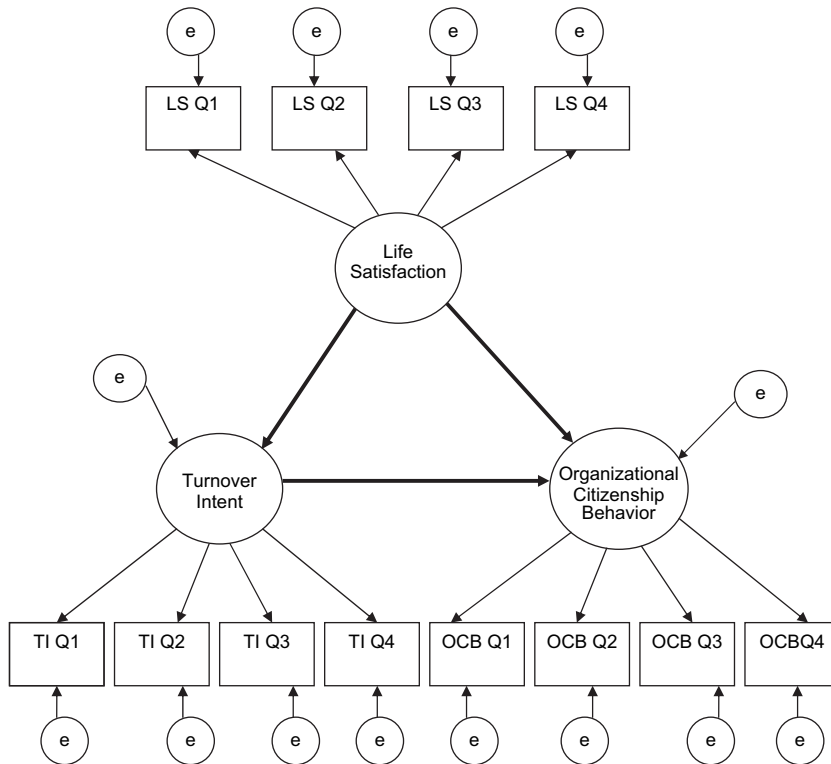


Fig. 2. Generic structural model.

specified theoretical model. The more the reproduced matrix deviates from the observed matrix, the less the theorized model fits the data. Finally, SEM allows the researcher to test hypothesized direct relationships between independent and dependent variables, such as multiple regression, *and* allows for the testing of indirect or mediated relationships between observed and unobserved latent variables while examining the reliability of the items to the latent variables.

Nonanalytic concepts

There are several nonanalytical and analytical concepts related to SEM that should be understood when reading or reporting results (Table 1). Three basic nonanalytic concepts are important for reporting: (1) a well-developed theoretical model, (2) operational definitions of the latent variables, and (3) graphical representation of the theorized model. A theoretical framework is expected in every report that explains and justifies the model to be tested. Within the framework, operational definitions of the latent constructs should be clearly delineated. A graphical

representation of the theorized model should be presented for the readers. The graphic can be used as an advanced organizer for the reader that specifies theorized relationships. Those theorized relationships are the hypotheses that will be tested, and a graphic display consumes less space and is easier to read and interpret than a large list of written hypotheses in the narrative of the report.

Analytical concepts

The analytical concepts are discussed and follow the pattern in the checklist in Table 1. The following discussion does not follow the checklist line by line, but the information on the checklist is discussed.

Sample size

SEM requires large sample sizes, which affects sampling error. The general rule of thumb is 20 cases for each free parameter, but 10 is a more likely target.² A free parameter (value), such as a coefficient, is 1 that is to be estimated in the analysis; that is, it is not specified before analysis

Table 1
Checklists for SEM reporting

<i>Nonanalytic reporting</i>
Well-developed theoretical framework
Theorized model display
Operational definitions
 <i>Analytic reporting</i>
Sample size
Original and final
Missing data
How handled (listwise deletion, imputed means...)
justified?
Specification/identification
Normality
Outliers
Linearity/multicollinearity (are variables too correlated
>0.90)
Software and estimation method stated: justified?
Power
Assessment of fit
Model Chi-square
Multiple fit indices justified
Parameters estimated and significant tests
Squared multiple correlation (CFA) variance
accounted for (SEM)
Standardized and unstandardized estimates
Residual analysis—predicted and actual covariance
matrix examination
Correlation and means tables
Modifications
Rationale for modification
Lagrange test for adding paths
Wald test for dropping paths
Correlation between estimated parameters
(hypothesized & final models)
Equivalent model
Diagram of final model

by the researcher. For example, if the theorized model has 15 free parameters to estimate, then a sample size between 150 and 300 participants is recommended. Power analysis is also an important consideration when selecting sample size for SEM analyses. Power at the coefficient level (path) can be examined.³ Researchers, though, tend to think about power at the model level. Estimated power and the associated minimum sample size have been studied.^{4,5} In general, the smaller the degrees of freedom for the model, the larger the sample size needed to reach a power of 0.80 for the model. For example, with degrees of freedom of 45, power of 0.80, and a model fit index, root mean square error of approximation (RMSEA) = 0.05, the minimum sample size is

252. If the degrees of freedom change to 8, then the minimum sample size increases to 954.⁴

Sample size can be referenced to the number of observations (participants) per items in a survey for example. Rules of thumb such as 5-10 observations per item are common to see, but that range is acceptable as long as the variable distributions are normal or elliptical and the latent variables have multiple indicators (eg, items). For readers, it is important for the authors to discuss sample size in reference to the number of observed variables, free parameters, distribution of the data, and complexity of the model tested.

Missing data

Respondents commonly fail to complete every item on a survey or instrument. Missing data are a serious concern because the results may be biased, or in error. There are 3 different types of missing data: missing completely at random (MCAR), missing at random (MAR), and non-ignorable.⁶ MCAR occurs when the missing data are not related to any group of participants or any other variable. MAR occurs when data may be missing at random, for example, persons with lower “socioeconomic status” may be less likely to provide their current health level, but upon examination of the variable “health level,” it is unrelated to the variable “socioeconomic status.” Nonignorable missing data are most serious, because the data are missing as a function of another variable. For example, more males failed to respond to socioeconomic status items than females.

Several approaches for dealing with the missing data, such as pairwise deletion, listwise deletion, regression, multiple imputation, full information maximum likelihood (FIML), and expectation maximization (EM) algorithm, are available.^{6,7} In general, pairwise deletion occurs when cases, or participants, are deleted from the analysis if they are missing from that part of the analysis, but are included in other analyses if the data are complete. Pairwise deletion is not recommended because sample size can change depending on which variables are being analyzed. This can lead to mathematical problems such as a 0 in the denominator of an inverted matrix. Pairwise also greatly limits the ability to generalize the results of the sample to the population. Listwise is acceptable but is best used when the data are MCAR. It is robust with MAR if the missing values of the independent variables do not depend

on the values of the dependent or outcome variable. If data are MAR, FIML is recommended. If the data are nonignorable, FIML estimates will tend to be the least biased.⁶ The methods and rationale for handling missing data should be provided in any report, along with a brief discussion of the strengths and limitations of that method.⁸

Specification

Specification involves the selection of variables for inclusion in the model. For a measurement model, the problem occurs in determining if the observed variables are correctly “connected” to the latent variables. For a structural model, the failure to include relevant variables will create a specification error. A specification error will cause the estimations to be incorrect and therefore result in inappropriate inferences.

Recursive and nonrecursive models

A model is recursive if all the paths flow one way with no feedback or reciprocal loops and the errors are uncorrelated. Figs. 1 and 2 are recursive models. Nonrecursive models can have feedback loops or correlated errors. Fig. 3a is nonrecursive due to the feedback paths between the endogenous Y variables. There are actually partially recursive models that can be treated as recursive because there is no direct relationship between the endogenous variables, even though the errors are correlated (Fig. 3b). This is important because of increased possibility of an identification problem (see below), which will result in an estimation problem. Fig. 3c is considered nonrecursive because of the directional path between endogenous Y variables.

Identification

As has been stated, “If life were fair, the researcher could proceed from specification to collection of the data to estimation. Unfortunately, ...the analysis...is not always so straightforward.”² A model is identified if there is a theoretically possible unique estimate for each parameter. A larger sample will not solve an identification problem because a general requirement for identification is related to the relationship between the number of free parameters and the number of observed variables, not sheer sample size. The number of free parameters should be equal to or greater than the number of observed variables, that is, the degrees of freedom must be greater than 0. Based on this general requirement, there are 3 general types of model identification: underidentified, where the degrees of freedom are less than 0; overidentified, where the degrees of freedom are greater than 0; and just-identified, where the degrees of freedom equal 0. Most statistical packages provide information about the identification of the analyzed model and whether there is an identification problem by indicating that the solutions provided are not admissible; otherwise, the software program will fail to complete the analysis, altogether.

Estimation methods and normality issues

Estimation concerns the procedure to be used to derive the parameter estimates, such as the coefficients and standard errors. Although the most common is maximum likelihood (ML) and is the default on most programs, the type of data one has (nominal, ordinal, etc) and the distributional qualities (eg, skew and kurtosis) of those data should determine the estimation method to

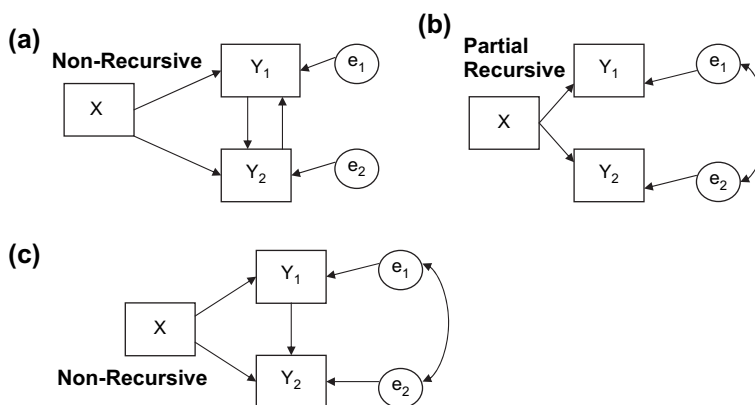


Fig. 3. (a-c) Partial- and nonrecursive examples.

be used. For normally distributed data, ML is a full information method where all parameters are estimated at once.⁹ Normally distributed data are symmetrical about the mean. Generalized least squares (GLS) and unweighted least squares (ULS) are also full information. GLS is part of the weighted least squares (WLS) family, and ULS requires all observed variables to have the same scale.^{2,7}

If the data are continuous but non-normal, there is always the option to transform the data. A corrected normal theory method can be used wherein the parameters are estimated with ML along with robust standard errors or corrected test statistics. When the data are non-normal, do not follow a normal bell curve distribution, and the non-normality is not corrected, the standard errors can be too low, which results in an erroneous rejection of the null hypothesis. Also, the Chi-square value, which is an indicator of how well the data fit the model, can be inflated, which will result in the rejection of true models based on exact fit tests (see below). The Satorra-Bentler method is a correction that adjusts the Chi-square value lower based on the size of observed kurtosis.¹⁰ An asymptotically distribution free (ADF) method is possible, but it requires very large sample sizes, such as 2500 cases. ADF estimation does not assume normal distribution of the variables or multivariate normality; therefore, if the data are skewed, ADF estimation is a technique for dealing with that problem.

The decision rule of thumb, for an individual variable, has been an absolute value of skew greater than 1, transform the data. Simulated data research indicates that ML estimation with robust standard errors can be used with skewness below 2 and kurtosis below 7.¹¹

Likert-type scales, for example, strongly disagree to strongly agree, common in survey-based research are technically ordered categorically, that is, there is an underlying continuum, but each choice is a separate category. Although there is no consensus on how to handle Likert-type scales with 3–5 categories, ordered categorical data can be estimated with the WLS family of estimators. For example, the MPlus program uses WLS with robust standard errors and adjusted Chi-square statistics.⁷ And LISREL 8.80 uses diagonally WLS.¹² If there are different types of data, categorical, ordered categorical, and continuous in the data set, then an appropriate estimation method needs to be chosen.⁷ The data could be transformed, that is, standardized, but with the

available technologies the raw data analysis is possible. Although a technical discussion of these estimation procedures is beyond the scope of this article, it is important that the estimation method is justified and explained with reference to the type of data and the distributional properties (normal/non-normal) of the data.

Fit indices

There are numerous fit indices. A fit index provides a global examination of how well the collected data fit the hypothesized model. Researchers typically have personal fit preferences, but common fit indices for a single analysis are comparative fit index (CFI), Tucker-Lewis index (TLI, or non-normed fit index), and RMSEA.^{2,13} Most statistical packages that run SEM models produce numerous fit indices. Table 2 provides a list of some of the major indices with general cutoff levels for a good fit. The model Chi-square value and degrees of freedom should be reported because some fit indices are based on those values. The model Chi-square is a badness of fit index because in general the higher the Chi-square value the more likely a statistically significant result indicating the worse the data fit the model. Chi-square and the degrees of freedom are expected to be reported, but are not typically used to justify the fit of the data to the model because the Chi-square value is affected by sample size. That is, large sample sizes will cause the Chi-square value to be statistically significantly different from 0 even when the fit of the data to the model is good.

The RMSEA is also a badness of fit index, wherein the higher the value the worse the data fit the model. An overall RMSEA less than or equal to 0.06, and a confidence interval range from 0.00 to 0.08 indicates a close or good fit.¹⁴ For example, an RMSEA of 0.056 with a lower bound of 0.02 and upper bound of 0.08 would be considered a close fit. There are mixed outcomes with RMSEA where the overall value is below 0.06 and the lower value is 0.00 and the upper value is over 0.10. This happens more often with small samples, but is a good reminder that fit indices are sample statistics that are affected by sampling error.²

The CFI is based on a ratio of the Chi-square of the tested model and the independent or null model. The range of values for the CFI is 0 to 1, but a value of 1 does not indicate a perfect fit, just that the Chi-square of the model is less than the degrees of freedom of the model tested. Generally,

Table 2
A list of popular fit indices

	Abbreviation	General rule for acceptable fit if data are continuous	Categorical data	Sivo et al simulation
Absolute/predictive fit indices				
Chi-square	χ^2	Ratio of χ^2 to $df \leq 2$ or 3, useful for nested models/model trimming		
Akaike information criterion	AIC	Smaller the better; good for model comparison (non-nested), not a single model		
Browne-Cudeck criterion	BCC	Smaller the better; good for model comparison, not a single model		
Bayesian information criterion	BIC	Smaller the better; good for model comparison (non-nested), not a single model		
Consistent AIC	CAIC	Smaller the better; good for model comparison (non-nested), not a single model		
Expected cross-validation index	ECVI	Smaller the better; good for model comparison (non-nested), not a single model		
Comparative fit indices				
Normed fit index	NFI	Comparison to a baseline (independence) or other model ≥ 0.95 for acceptance		0.96
Incremental fit index	IFI	≥ 0.95 for acceptance		0.98
Tucker-Lewis index	TLI	≥ 0.95 can be $0 > TLI > 1$ for acceptance	0.96	
Comparative fit index	CFI	≥ 0.95 for acceptance	0.95	0.98
Relative noncentrality fit index	RNI	≥ 0.95 , similar to CFI but can be negative, therefore, CFI better choice		
Parsimonious fit indices				
Parsimony-adjusted NFI	PNFI	Very sensitive to model size		0.77
Parsimony-adjusted CFI	PCFI	Sensitive to model size		
Parsimony-adjusted GFI	PGFI	The closer to 1 the better, though it is typically lower than other indices and sensitive to model size		0.77
Other indices				
Goodness-of-fit index	GFI	≥ 0.95 Not generally recommended		0.96
Adjusted GFI	AGFI	≥ 0.95 Performance has been poor in simulation studies		0.95
Root mean square residual	RMR	The closer to 0 the better		0.11
Standardized RMR	SRMR	≤ 0.08		0.07
Weighted root mean residual	WRMR	< 0.90	< 0.90	
Root mean square error of approximation	RMSEA	< 0.06 - 0.08 with confidence interval	< 0.06	0.03

Sivo et al's fit values are based on a sample size of 500 and indicate the optimal index value without rejecting any correct model (p. 276).

a value equal to or above 0.95 has been considered a good fit.¹⁴ The comparison of the baseline model versus the null has been critiqued as not overly informative.¹⁵ Researchers can choose to constrain different paths to test and use as alternative null models.¹⁵

One recent simulation indicates that fit values are affected by sample size, complexity, and misspecification of the model.¹⁶ For example, for a simple confirmatory model, the RMSEA value for a sample size of 500 should be 0.03 to ensure the rejection of all misspecified models; however, for the standardized root mean square residual (SRMR), the value would be 0.15, which is considered quite high.¹⁶

The fit indices discussed above are used for continuous data but work has been conducted with categorical data. The TLI and CFI values, for categorical data, should be greater than or equal to 0.95 and the RMSEA, the value should be less than or equal to 0.06.¹⁷ Another index, the weighted root mean square residual (WRMR), should have a value less than 0.90 for both continuous and categorical data.¹⁷

Because there is continued research on fit indices, it is important that citations for the fit value are relatively current, for example, within 5–10 years of the current date. Finally, there should be justification for the group of fit indices provided and those values should converge to a conclusion that the data fit the theorized model. It is important to judge the overall fit from multiple indices and not just a single fit index.¹⁴

Nested models

Nested models are modified versions of the original theorized model. Nested models occur when researchers add or remove relationships among observed or latent variables after analysis. The core of this postanalysis should be an examination of the coefficients of hypothesized relationships. Many times, the examination and discussion of the coefficients are secondary to the fit, but they should not be. In other words, the researcher examines the significance of individual structural paths representing the impact of one latent construct on another or the latent construct on the observed variable as is the case with CFA. The statistical significance of path coefficients are established through an examination of the *t*-values or *z*-values (depending on the software) associated with structural coefficients. The authors also could provide the standard errors in combination with the unstandardized estimates.

It is important to note that a standardized coefficient can be greater than 1 and does not indicate that there is something wrong, but there might be a high degree of multicollinearity, 2 variables highly correlated, for example, will bias the results.¹⁸

SEM is an *a priori* method; therefore, the removal or addition of a relationship (direct/indirect or correlated), called a modification, must make theoretical, not just statistical sense. Basing modifications to the hypothesized model only on statistical results may lead to a final model that highlights quirks or unique characteristics of the sample that are not representative of the population. Typically, removing a path will result in a higher or worse Chi-square value and adding a path will lower the Chi-square value. Modification values, which indicate which paths should be added or removed, can be produced by many of the software programs. The value is either the Lagrange multiplier, which examines how much the Chi-square would decrease if a specific path was added, or the Wald statistic, which examines how much the Chi-square would increase if a specific path was removed.²

Once a modification is made (also called model trimming or building), the original model and the new model should be compared. The comparison is a Chi-square difference test, which tests the null hypothesis of identical fit between the 2 models. For example, if the original model Chi-square was 21.25 with 5 degrees of freedom and the new model with a path added has a Chi-square of 13.21 with 4 degrees of freedom, with a resultant statistical significance of $P = .004$, the new model is a better fit. If multiple systematic modifications are made, authors should report fit indices with the first and final model. It should be noted that once modifications have been completed the analysis has moved from confirmatory to exploratory. Researchers often respecify or modify the original model when parameter estimates are statistically nonsignificant.

Abuse of fit indices occurs when the researcher becomes fascinated with improving the fit. This abuse has been described similarly to eating salted peanuts, “one is never enough.”^{19,p750} For the purposes of reporting, it is important that each modification made based on statistical criteria is also justified theoretically.

Alternative models

When testing alternative models with the same data that are not nested or hierarchically related,

the Chi-square test between old and new models is not appropriate. There are 2 good choices for the researcher at this point for fit index comparison: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). In general, the model with the lowest value is best. For example, if the BIC value for alternative *A* is 2311.12, and for *B* the value is 1679.13, alternative *B* would be the preferred model.

Researchers may consider mathematically equivalent models.²⁰ An equivalent model should provide the same predicted correlation or covariances but with different paths among the manifest variables.^{2,20} If examining mathematically equivalent models, the chosen model must be justified over the other mathematically identical models.

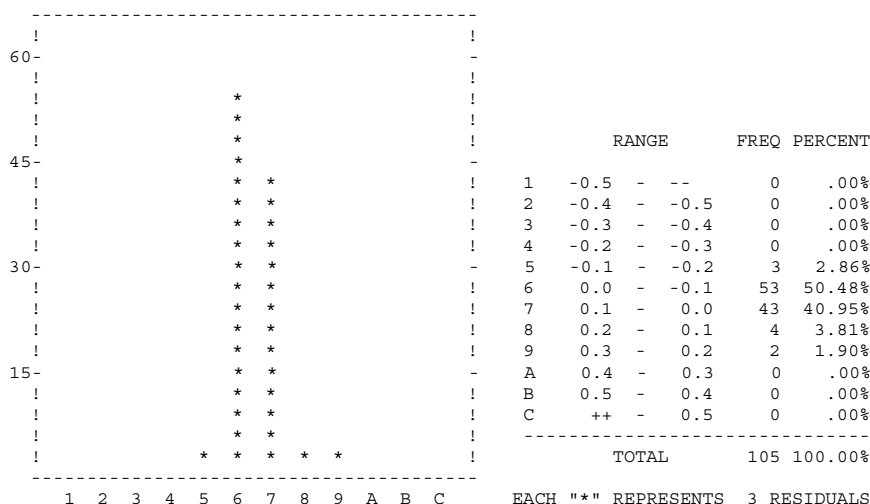
Residuals

The residuals, sometimes labeled correlation residuals, fitted, or standardized, are the difference between the observed score and the predicted score based on the model and should be discussed in any SEM report (Fig. 4). The general rule of thumb is to look for correlation residuals (fitted residuals) that have an absolute value greater than 0.10. The more residuals over 0.10 the less explanatory the model is to the data. In Fig. 4, approximately 8% of the residuals are greater than an absolute value of 0.10. There is no rule for how many residuals over 0.10 indicate a problem, but as the incidence of these larger residuals

increases, the model's explanatory power deteriorates. Software programs provide different types of outputs. Fig. 4 is a histogram and frequency plot from EQS 6.1.²¹ The major statistical packages provide some form of residual matrix that should be discussed.²² A discussion of the standardized form of the residuals is preferred. The standardized residual values enable the researcher to determine the number of standard deviations of observed residuals from 0 residuals that should exist if the causal model fits perfectly.²² A Q-plot, which graphs the standardized residuals in comparison to a 45° line, displays residual values that depart extremely from the Q-plot line, which provides evidence that the model is in some way misspecified.²²

CFA and SEM examples

The measurement model of SEM is a CFA (Fig. 1) and depicts the pattern of observed variables for those latent constructs in the hypothesized model. The objective for CFA is to test the reliability of the observed variables and provide a rigorous test of convergent and discriminant validity.^{2,23} Second, the measurement model examines the extent of interrelationships, covariation (or lack thereof), among the latent constructs, themselves. As part of the process, factor loadings, unique variances, and modification indices,



Note: The more residuals over the absolute value of .10, the less explanatory power the model has.

Fig. 4. Residual histogram and frequency chart. The more residuals over the absolute value of 0.10, the less explanatory power the model has.

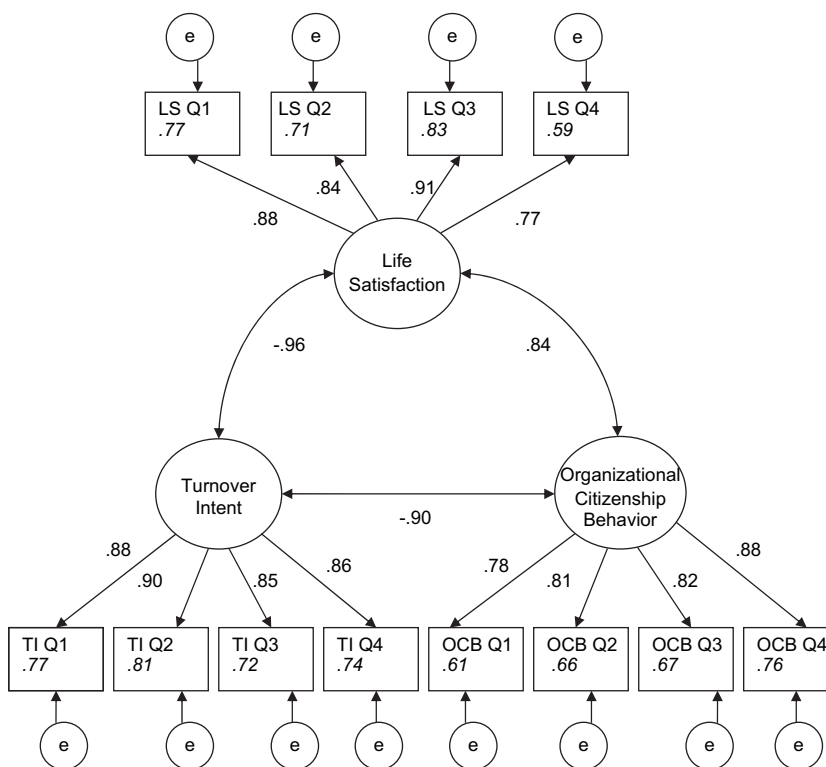
are estimated to derive the best indicators of latent variables before testing a structural model.

CFA is theory driven. Therefore, the planning of the analysis is driven by the theoretical relationships among the observed and unobserved variables. When a CFA is conducted, the researcher uses a hypothesized model to estimate a population covariance matrix which is compared to the observed covariance matrix. Mathematically, the researcher wants to minimize the difference between the estimated and observed matrices.

CFA example

The example for this article is based on employees' relationships and behaviors within an organization. The data used here are not for peer-reviewed empirical results, but rather, only for example purposes. Therefore, a full theoretical framework is not provided here, but is expected in a research report. A CFA with results is depicted in Fig. 5. In this example, there are 3 latent

variables, LS, turnover intent (TI), and organizational citizenship behavior (OCB), with 12 manifest or observed variables. All 3 latent variables are measured with statement-based self-report instruments and a Likert-type scale. The scales for each range from 1 to 6 with 1 being strongly disagree and 6 being strongly agree. If the data being analyzed were composed of different types of data, such as categorical, continuous, or order categorical, it would be important to choose an estimation method appropriate for that set of data.⁷ LS refers to an employee's overall impression of their life and various aspects of their life, such as family. An example item is, "I am happy when I think about my family." TI refers to the overall impression that the employee is seeking new employment with another firm. An example item is, "Each week, I read open position advertisements in my field." OCB refers to an employee assuming extra voluntary roles, or discretionary behavior, that contributes to organizational effectiveness. An example item is, "I commonly put in



Chi-square = 378.26, df = 51 CFI = .97 TLI = .96 RMSEA = .05

Standardized Results are presented with Square Multiple Correlations in Italics

Fig. 5. Results and final model of CFA.

extra time on meetings to make sure the organization runs smoothly.” The data for this study should not be cited as actual results. The data are real, but due to a confounding effect during data collection, cannot be used for peer-reviewed empirical results. Therefore, it is used as an example here.

In Fig. 5, the large circles are the unobserved or latent variables, LS, TI, and OCB. The small circles with the letter “e” are the unique factors, or measurement errors, in the variables. The unique factors differ from the latent factors because their effect is associated only with 1 observed variable. The straight line pointing from a latent variable to the observed variables indicates the “causal” effect of the latent variable on the observed variables. The curved double-head arrow between latent variables indicates a correlation. If the curve was changed to a straight 1-headed arrow, a hypothesized direct relationship between the 2 latent variables would be indicated. A directional path would be considered a structural component of the model. In the example, all of the latent variables—including the errors—are unobserved exogenous and all of the manifest variables are observed endogenous.

The diagram is the hypothesized model that is to be tested to see how well it “fits” the observed data. Mathematical equations describe the pictured relationships. Presentation of those equations is beyond the scope of this article.^{19,24,25}

Results for the CFA

For this example, 650 respondents were asked to complete a series of surveys. Of those 650, 629 completed surveys were collected. A missing data analysis was completed to examine the relationships between missing values on each variable, including demographic characteristics. No statistically significant relationships were observed. Therefore, a listwise deletion procedure was used. The data are skewed but below an absolute value of 2, so ML estimation with robust standard errors was used.¹¹ The software used was MPlus, version 2.14.⁷ Based on the number, sample size, degrees of freedom (51), and the number of free parameters (27), there appears to be enough statistical power. Finally, the correlation table was examined for multicollinearity and no relationships were above >0.90. High bivariate correlations cause operational problems with the matrices involved with CFA and SEM. This topic is beyond the scope of this article.

The fit indices indicate a good fit of the data to the hypothesized structure. The CFI is 0.97, the TLI is 0.96, and the RMSEA is 0.05, with a 90% confidence interval of 0.04–0.08. All 3 of these values indicate a good fit.¹⁴ CFI and TLI may be low and the RMSEA may be high, given the sample size.¹⁶ All items loaded statistically significantly ($P < .05$) on the theorized latent variables, and no modifications were warranted based on the values calculated (Table 3). A random assortment

Table 3
Confirmatory Factor Analysis Example

Standardized and Unstandardized Regression Weights and Squared Multiple Correlations (SMC)

	Standardized			Unstandardized (se)			SMC
	OCB	TI	LS	OCB	TI	LS	
OCB Q1	.78			1.00 (.00)			0.61
OCB Q2	.81			1.07 (.03)			0.66
OCB Q3	.82			1.07 (.03)			0.67
OCB Q4	.88			1.11 (.03)			0.77
LS Q1		.88			1.00 (.00)		0.77
LS Q2		.84			1.10 (.02)		0.71
LS Q3		.91			1.06 (.02)		0.83
LS Q4		.77			.92 (.02)		0.59
TI Q1			.88			1.00 (.00)	0.77
TI Q2			.90			.93 (.02)	0.81
TI Q3			.85			.89 (.02)	0.72
TI Q4			.86			.88 (.02)	0.74

Note: Statistically Significant $p < .05$ in bold.

OCB, Organizational Citizenship Behavior;

LS, Life Satisfaction;

TI, Turnover Intent.

of items led to fit indices of 0.90 for the TLI and 0.09 for the RMSEA. No multivariate outliers or large residuals based on the observed and replicated covariance matrices were observed.

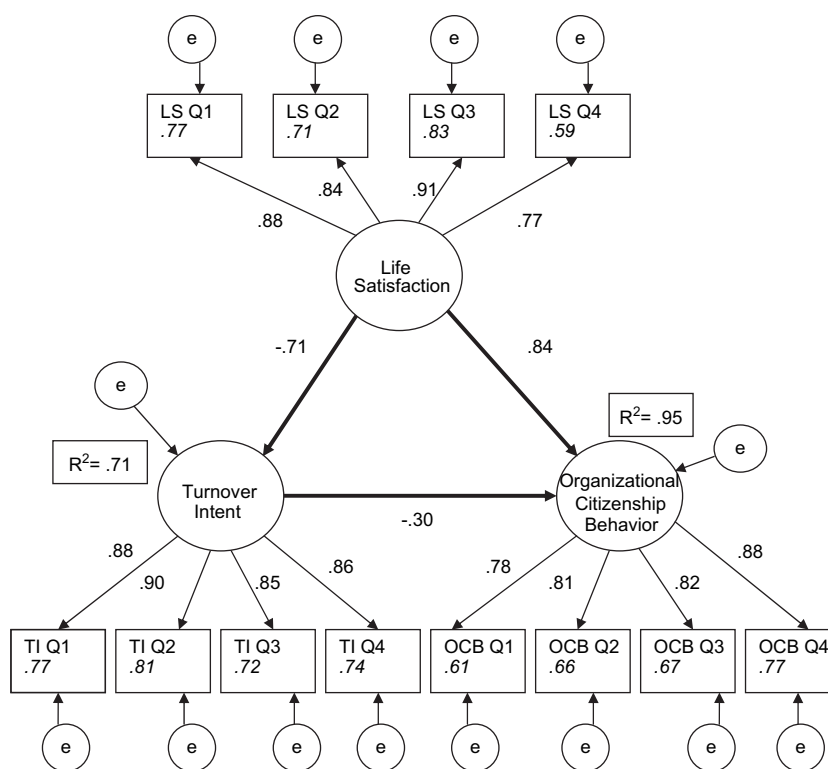
Standardized values are provided in Fig. 5. Unstandardized values with standard errors are in Table 3. In Fig. 5, the italicized values are the squared multiple correlations and are an indicator of lower bound of reliability for that item. It is calculated by squaring the standardized coefficient (eg, $0.78^2 = 0.61$). The observed data appear to fit the hypothesized model and there are large correlations among the latent variables. Finally, because each item only loaded on 1 factor, there is evidence of convergent and discriminant validity.⁷ A second test for discriminant validity is to constrain (fix) the relationships between the latent constructs to 1. In this example, doing so causes the Chi-square value to increase to over 1969 with the *df* equal to 54. Examining the Chi-square from the tested model to this constrained model difference, which is statistically significant, indicates that the relationships are not 1 thus

providing discriminant validity evidence. Even with the Chi-square difference, these data appear to indicate that LS and TI are highly related.

Though it is typical and appropriate in research studies to provide internal consistency values of the scores from the items with the latent constructs once a CFA is run, those values are less of a concern in CFA because 1 can have high internal consistency (>0.80) across a set of responses from a survey, yet still have multiple constructs. The internal consistency values for these data are all above 0.90. Also, reliability and validity are a quality of the responses from instruments, and not the instruments, therefore these values need to be recalculated each time scores are obtained from the responses to the instruments.

SEM example

SEM has been described as a combination of exploratory factor analysis and multiple regression,¹⁹ at least in part because although SEM is a confirmatory technique, it can also be used for



Chi-square 378.26 *df* = 51 CFI = .97 TLI .96 RMSEA = .05

Standardized Results are presented with Square Multiple Correlations in Italics

Fig. 6. Results of SEM.

exploratory purposes. SEM, in comparison to CFA, extends the possibility of relationships among the latent variables and encompasses 2 components: a measurement model (essentially the CFA) and a structural model. In Figs. 2 and 6, the structural component of the model is highlighted by the directional paths in bold. The structural model displays the direct and indirect effects among latent constructs and observable variables in the proposed model as a succession of structural equations—akin to running several regression equations.

There is no consensus on the wording that should be used with structural models, and therefore, readers will see the term “cause” or phrase “causal modeling” or “direct-,” “indirect-,” and “total-” effects. For this example, the direct, indirect, and total language is used to highlight different aspects of the results. A direct effect (Figs. 2 and 6) represents the effect of an independent variable (exogenous) on a dependent variable (endogenous). In the diagrams, LS and TI each have a direct effect on OCB, and LS has an indirect effect on OCB through TI. That is, TI is acting as a mediating variable.²⁶ For this example, TI is both an exogenous and an endogenous variable.

A missing data analysis was completed to examine the relationships between missing values on each variable and other demographic information. No statistically significant relationships were observed. Therefore, a listwise deletion procedure was used. The data are skewed but below an absolute value of 2, so ML estimation with robust standard errors was used.¹¹ The software used was MPlus version 2.14.⁷ Based on the number sample size and degrees of freedom (51) and the number of free parameters (27), there is enough power.

Standardized coefficient values are provided in Fig. 6, and unstandardized values with standard errors are presented in Table 4. Table 5 has the means, standard deviations, and correlations. In Fig. 5, the italicized values are the squared multiple correlations and are an indicator of lower bound of reliability for that item. It is calculated by squaring the standardized coefficient (eg, $0.78^2 = 0.61$). The observed data appear to fit the hypothesized model based on the fit indices.

The structural analysis indicates that LS has a statistically significant direct positive effect on OCB indicating that the higher LS is, the more employees are performing extra role behaviors. LS also has an indirect effect on OCB through TI. The more you are satisfied with your life, the less likely you are to contemplate a job switch, but you

are more likely to perform extra role behaviors. LS and TI account for 95% of the variance in OCB. LS accounts for 71% of the variability in TI scores.

Although the focus of structural modeling is to estimate relationships among hypothesized latent constructs, structural modeling can be used to test experimental data where 1 or more of the variables have been manipulated. Nested data, such as clients nested within physicians and physicians nested within geographical regions, and trajectory models such as different paths of anxiety over time can also be analyzed within the SEM framework. In sum, SEM allows researchers to test theoretical propositions regarding how constructs are theoretically linked and the directionality of theoretically important relationships.

Software programs

Below is a brief description of several of the most common software programs for SEM.

MPlus 4.0

MPlus is a Microsoft Windows program that can be used to analyze a variety of categorical and continuous latent variable models with both continuous and categorical data.⁷ The program can analyze the models presented here, plus growth, multilevel, latent class, and cohort modeling, to name a few. Missing data can be estimated with ML. The interface for programming can be run through a Language Generator or programmed line-by-line by the researcher. A trial, or student, version is available for download at www.statmodel.com.

Table 4
Results from the Sample SEM Analysis

SEM Example					
	Standardized		Unstandardized (se)		R ²
	LS	OCB	LS	OCB	
Direct					
OCB	.84		.94(.02)		0.95
TI	−0.71	−0.30	−.67(0.03)	−.28(.02)	0.71
Indirect					
TI	−0.26				
Total					
OCB	.84				
TI	-.96	-.30			

OCB = Organizational Citizenship Behavior

LS = Life Satisfaction

TI = Turnover Intent.

Table 5

Correlations, means, and standard deviations of individual items for each latent construct

Item	1	2	3	4	5	6	7	8	9	10	11	Mean	SD
1 LS 1												5.16	1.00
2 LS 2	0.77											4.96	1.16
3 LS 3	0.77	0.72										5.18	1.02
4 LS 4	0.60	0.62	0.70									5.12	1.05
5 TI 1	−0.62	−0.70	−0.59	−0.49								5.19	1.04
6 TI 2	−0.65	−0.64	−0.70	−0.61	0.62							5.13	1.07
7 TI 3	−0.61	−0.62	−0.68	−0.64	0.53	0.64						5.07	1.02
8 TI 4	−0.68	−0.67	−0.76	−0.63	0.58	0.70	0.71					5.16	1.04
9 OCB 1	0.61	0.60	0.64	0.53	−0.55	−0.65	−0.63	−0.66				5.09	1.11
10 OCB 2	0.57	0.55	0.60	0.47	−0.50	−0.66	−0.61	−0.66	0.81			5.24	1.05
11 OCB 3	0.62	0.61	0.68	0.55	−0.55	−0.68	−0.63	−0.72	0.75	0.80		5.18	1.09
12 OCB 4	0.59	0.55	0.63	0.48	−0.54	−0.66	−0.55	−0.64	0.72	0.76	0.73	5.26	1.03

N = 629.

MX 6.0

MX is a free downloadable program for a computer with Microsoft Windows and can be used to analyze a variety of models with both continuous and categorical variables.²⁷ MX can also impute missing data with ML. The graphical interface for programming, that is, specifying the model, allows the writing of a program script or drawing the model. More information and the download can be found at www.vcu.edu/mx.

Amos 7.0

Version 7 of Amos is a Microsoft Windows program where a model can be specified by writing a script or drawing the model.²⁸ The program uses an ADF estimator for non-normal data but cannot analyze ordinal data. A new feature allows non-numerical data to be analyzed without having to create numerical values. Amos is licensed from the Smallwaters Company to SPSS. More information can be found at <http://www.spss.com/amos/index.htm>.

EQS 6.1

EQS is another Microsoft Windows program that allows the specification of a variety of models.²¹ The user can specify a model by writing a script, using a programming language generator, or drawing the model. A variety of new additions to the program have expanded its capabilities with version 6.1, such as the multilevel modeling component. EQS also has a data entry and analysis component. More information and a demonstration version can be found at <http://www.mvsoft.com/eqs60.htm>.

CALIS

CALIS is part of the SAS comprehensive statistical package.²⁹ CALIS can analyze a smaller variety of models compared to other SEM related packages. Users familiar with programming in LISREL or EQS will find the transfer to using CALIS easy because the programming commands must be typed into an editor window that describes the data and the model. More information can be obtained at <http://v8doc.sas.com/sashtml/stat/chap19/index.htm>.

LISREL 8.8

LISREL is sometimes called the father of SEM software programs. It has been around for more than 30 years and can analyze a variety of data and models.¹² The program has 3 components, PRELIS, LISREL, and SIMPLIS. PRELIS is a data screening program and it prepares the matrices needed for LISREL. LISREL is the SEM program and the user can write syntax in the LISREL programming language or specify a model by drawing it. Users can also use SIMPLIS to write the syntax, and instead of using the LISREL language the user names the observed and latent variables and then specifies the paths with equation-like statements. A demonstration version and other details can be found at <http://www.ssicentral.com/lisrel/index.html>.

Conclusion

This paper provides a basic outline and example of reporting results for SEM analyses. SEM is a flexible technique, and there are a wide variety of analyses that can be completed within the SEM

framework. The basic reporting of information from the type of data collected to the specific estimation procedure used is similar across all of them. It is important for evaluation purposes that the information related choices made before, during, and after initial analysis is provided and justified.

References

- Loehlin JC. *Latent Variable Models*. Mahwah, NJ: Erlbaum; 1992.
- Kline RB. *Principles and Practices of Structural Equation Modeling*. 2nd ed. New York, NY: The Guilford Press; 2005. p. 105.
- Saris WE, Satorra A. Power evaluation in structural equation models. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models*. Newbury Park, CA: Sage; 1993. p. 181–204.
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1996;1:130–149.
- Hancock GR, Freeman MJ. Power and sample size for the root mean square error of approximation of not close fit in structural equation modeling. *Educ Psychol Measure* 2001;61:741–748.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. New York, NY: Wiley; 2002.
- Muthén B, Muthén L. *MPlus* [computer software]. Los Angeles, CA: Muthén & Muthén; 1998–2004.
- Allison PD. *Missing Data*. Beverly Hills, CA: Sage Publications; 2003.
- Eliason SR. *Maximum Likelihood Estimation: Logic and Practice*. Beverly Hills, CA: Sage Publications; 1993.
- Satorra A, Bentler PM. Corrections to test statistics and stand errors on covariance structure analysis. In: von Eye A, Clogg CC, eds. *Latent Variable Analysis*. Thousand Oaks, CA: Sage; 1994. p. 399–419.
- Curran PJ, West SG, Finch JF. The robustness of test statistics to normality and specification error in confirmatory factor analysis. *Psychol Methods* 1997;1:16–29.
- Jöreskog KG, Sörbom D. *LISREL 8.80 for Windows* [computer software]. Lincolnwood, IL: Scientific Software International; 2006.
- Maruyama GM. *Basics of Structural Equation Modeling*. Thousand Oaks, CA: Sage; 1998.
- Hu L-T, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 1999;6:1–55.
- Widaman KF, Thompson JS. On specifying the null model for incremental fit indexes in structural equation modeling. *Psychol Methods* 2003;8:16–37.
- Sivo SA, Fan X, Witta LE, Willse JT. The search for optimal cutoff properties: fit index criteria in structural equation modeling. *J Exp Educ* 2006;74(3):267–288.
- Yu C-Y. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. MPlus software Web site. Available from: <http://www.statmodel.com>; 2002. Accessed 05.01.05.
- Jöreskog KG. How large can a standardized coefficient be?. SSI Scientific Software International. Available from: <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf>; 1999. Accessed 31.10.06.
- Ullman JB. Structural equation modeling. In: Tabachnick BG, Fidell LS, eds. *Using Multivariate Statistics*. 4th ed. Needham Heights, MA: Allyn and Bacon; 2001. p. 729–801.
- MacCallum RC, Austin JT. Applications of structural equation modeling in psychological research. *Annu Rev Psychol* 2000;51:201–226.
- Bentler P. *EQS 6.1* [software]. Encino, CA: Multivariate Software Inc.; 2005.
- Byrne BM. *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models*. New York, NY: Springer-Verlag; 1989.
- Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. *Psychol Bull* 1988;103(3):411–423.
- Long JS. *Confirmatory Factor Analysis: A Preface to LISREL*. Beverly Hills, CA: Sage Publications; 1983.
- Long JS. *Covariance Structure Models: An Introduction to LISREL*. Beverly Hills, CA: Sage Publications; 1983.
- Baron RM, Kenny DS. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51:1173–1182.
- Neale MC, Noker SM, Xie G, Maes HH. *Mx: Statistical Modeling*. 6th ed. VCU Box 900126, Richmond, VA 232298: Department of Psychiatry; 2003.
- Arbuckle JL. *AMOS 7.0* [software]. Chicago, IL: SPSS Inc.; 2006.
- SAS Institute Inc. *The SAS System of Windows [Release 8.01]* [software]. Cary, NC: SAS Institute Inc.; 2000.