CHAPTER 23

Structural Equation Modeling

JODIE B. ULLMAN AND PETER M. BENTLER

A FOUR-STAGE GENERAL PROCESS OF MODELING 663
MODEL ESTIMATION TECHNIQUES AND TEST STATISTICS 667
MODEL EVALUATION 671
MODEL MODIFICATION 674
MULTIPLE GROUP MODELS 675
A GUIDE TO SOME RECENT LITERATURE 676
REFERENCES 683

Structural equation modeling (SEM) is a collection of statistical techniques that allow a set of relationships between one or more independent variables (IVs), either continuous or discrete, and one or more dependent variables (DVs), either continuous or discrete, to be examined. Both IVs and DVs can be either factors or measured variables. Structural equation modeling is also referred to as causal modeling, causal analysis, simultaneous equation modeling, analysis of covariance structures, path analysis, or confirmatory factor analysis. The latter two are actually special types of SEM.

SEM allows questions to be answered that involve multiple regression analyses of factors. At the simplest level, a researcher posits a relationship between a single measured variable (perhaps, acceptance of risky behavior) and other measured variables (perhaps, gender, academic achievement, and institutional bonds). This simple model is a multiple regression presented in diagram form in Figure 23.1. All four of the measured variables appear in boxes connected by lines with arrows indicating that gender, academic achievement, and institutional bonds (the IVs) predict acceptance of risky behavior (the DV) in adolescents. Lines with two arrows indicate a covariance among the IVs. The presence of a residual indicates imperfect prediction.

A more complicated model of acceptance of risky behavior appears in Figure 23.2. In this model, Acceptance of Risky Behavior is a latent variable (a factor) that is not directly measured but rather assessed indirectly using two measured variables (okay to drink and okay to smoke). Acceptance of Risky Behavior is, in turn, predicted by gender (a measured variable) and by Weak Institutional Bonds, a second factor that is assessed through two measured variables (bonds to family and bonds to teachers). For clarity in the text, initial capitals are used for names of factors and lowercase letters for names of measured variables.

Figures 23.1 and 23.2 are examples of path *diagrams*. These diagrams are fundamental to SEM because they allow the researcher to diagram the hypothesized set of relationships in the model. The diagrams are helpful in clarifying a researcher's ideas about the relationships among variables and they can be directly translated into the equations needed for the analysis.

Several conventions are used in developing SEM diagrams. Measured variables, also called *observed variables*, indicators, or *manifest variables,* are represented by squares or rectangles. Factors have two or more indicators and are also called *latent variables*, *constructs,* or *unobserved variables*. Factors are represented by circles or ovals in path diagrams. Relationships between variables are indicated by lines; lack of a line connecting variables implies that no direct relationship has been hypothesized. Lines have either one or two arrows. A line with one arrow represents a hypothesized direct relationship between two variables, and the variable with the arrow pointing to it is the DV. A line with a two-headed arrow indicates an unanalyzed relationship, simply a covariance between the two variables with no implied direction of effect.

This chapter was supported, in part, by NIDA grant DA01070–38.

Handbook of Psychology, Second Edition, edited by Irving B. Weiner.

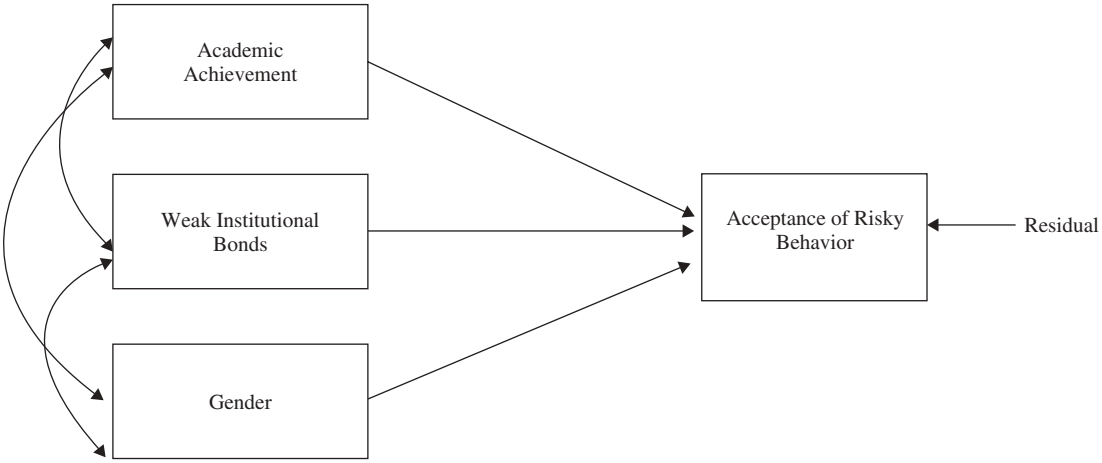

Figure 23.1 Path diagram of a multiple regression model

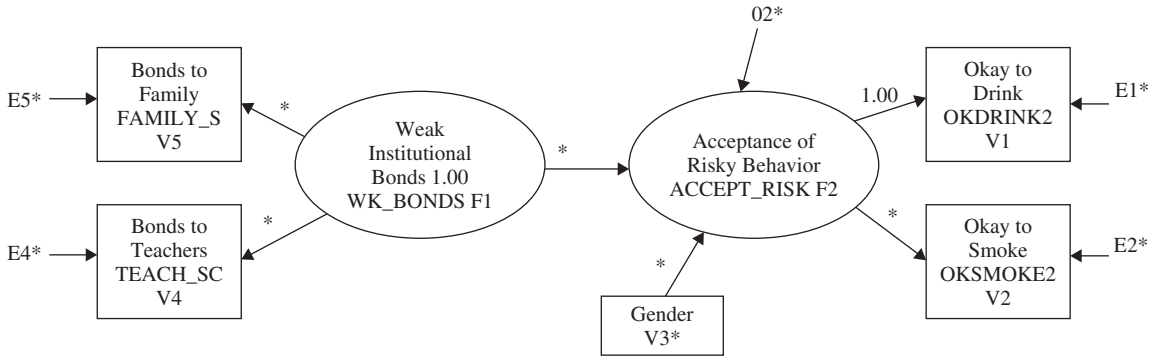

Figure 23.2 Example of a structural equation model

In the model of Figure 23.2, Acceptance of Risky Behavior is a latent variable (factor) that is predicted by gender (a measured variable), and Weak Institutional Bonds (a factor). Notice the line with the arrow at either end connecting Weak Institutional Bonds and gender (no line in the figure, that is, no covariance depicted). This line with an arrow at either end implies that there is a relationship between the variables but makes no prediction regarding the direction of effect. Also notice the direction of the arrows connecting the Acceptance of Risky Behavior construct (factor) to its indicators: The construct *predicts* the measured variables. The implication is that Acceptance of Risky Behavior drives, or creates, "okay to drink" and "okay to smoke." It is impossible to measure this construct directly, so we do the next best thing and measure several indicators of risky behavior. We hope that we are able to tap into adolescents' Acceptance of Risky Behavior by measuring several observable indicators, in this example, two.

In Figure 23.2, bonds to family, bonds to teachers, okay to drink, and okay to smoke, and the latent variable, Acceptance of Risky Behavior, all have one-way arrows pointing to them. These variables are dependent variables in the model. Gender and Weak Institutional Bonds are IVs in the model; as such they have no one-way arrows pointing to them. Notice that all the DVs, both observed and unobserved, have arrows labeled E or D pointing toward them. Es (errors) point to measured variables; Ds (disturbances) point to latent variables (factors). As in multiple regression, nothing is predicted perfectly; there is always residual error. In SEM, the residual variance (the variance unexplained by the IV[s]) is included in the diagram with these paths.

The part of the model that relates the measured variables to the factors is sometimes called the *measurement model*. In this example, the two constructs Weak Institutional Bonds and Acceptance of Risky Behavior and the indicators of these constructs form the *measurement model.* The hypothesized relationships among the constructs, in this example, the one path between Weak Institutional Bonds and Acceptance of Risky Behavior, is called the *structural model.*

Note, both models presented so far include hypotheses about relationships among variables (covariances) but not about means or mean differences. Mean differences associated with group membership can also be tested within the SEM framework.

The first step in a SEM analysis is specification of a model, so this is a *confirmatory* rather than an exploratory technique. The model is estimated, evaluated, and perhaps modified. The goal of the analysis might be to test a model, to test specific hypotheses about a model, to modify an existing model, or to test a set of related models.

There are a number of advantages to use of SEM. When relationships among factors are examined, the relationships are free of measurement error because the error has been estimated and removed, leaving only common variance. Reliability of measurement can be accounted for explicitly within the analysis by estimating and removing the measurement error. Additionally, as was seen in Figure 23.2, complex relationships can be examined. When the phenomena of interest are complex and multidimensional, SEM is the only analysis that allows complete and simultaneous tests of all the relationships. In the social sciences we often pose hypotheses at the level of the construct. With other statistical methods these construct-level hypotheses are tested at the level of a measured variable (an observed variable with measurement error). When the level of the hypothesis and the level of data are mismatched faulty conclusions may occur. This mismatch problem is often overlooked. A distinct advantage of SEM is the ability to test construct-level hypotheses at a construct level.

Three General Types of Research Questions That Can Be Addressed With SEM

The fundamental question that is addressed through the use of SEM techniques involves a comparison between a dataset, an empirical covariance matrix, and an estimated population covariance matrix that is produced as a function of the model parameter estimates. The major question asked by SEM is, "Does the model produce an estimated population covariance matrix that is consistent with the sample (observed) covariance matrix?" If the model is reasonable, the parameter estimates will produce an estimated matrix that is close to the sample covariance matrix. "Closeness" is evaluated primarily with the chi-square test statistics and fit indices. After establishing that the model is adequate we can test hypotheses within the model by evaluating the model parameter estimates. We can also test hypotheses involving statistical comparisons of different models, models that are subsets of one another (nested models).

If the estimated population covariance matrix and the empirical covariance matrix are very close, the model parameters (path coefficients, variances, and covariances) used to estimate the population covariance matrix could be evaluated. Using the example illustrated in Figure 23.2 we could test the hypothesis that increased (weaker) Institutional Bonds predicts greater Acceptance of Risky Behavior. This would be a test of the path coefficient between the two latent variables, Weak Institutional Bonds and Acceptance of Risky Behavior (the null hypothesis for this test would be H_0: $\gamma = 0$, where γ is the symbol for the path coefficient between an independent variable and a dependent variable). This parameter estimate is then evaluated with a z test.

Not only is it possible to test hypotheses about specific parameters within a model, it is also possible to statistically compare nested models to one another. Each model might represent a different theory; SEM provides a strong test for competing theories (models).

A FOUR-STAGE GENERAL PROCESS OF MODELING

The process of modeling could be thought of as a four-stage process: model specification, model estimation, model evaluation, and model modification. In this section each of these stages is discussed and illustrated with a small example based on simulated data.

Model Specification/Hypotheses

The first stage in the modeling process is specifying the model, that is, the specific set of hypotheses to be tested. This is done most frequently through a diagram. This examples has five measured variables: (1) FAMILY_S, a Likert-scale measure of strength of bonds to family; (2) TEACH_SC, a Likert-scale measure of strength of bonds to teachers; (3) OKDRINK1, a Likert-scale measure of endorsement of drinking alcohol; (4) OKSMOKE2, a Likert-scale measure of endorsement of smoking tobacco; and (5) Gender.

The hypothesized model for these data is diagrammed in Figure 23.2. Latent variables are represented with circles and measured variables are represented with squares. A line with an arrow indicates a hypothesized direct relationship between the variables. Absence of a line implies no hypothesized direct relationship. The asterisks indicate

parameters to be estimated. The variances of IVs are parameters of the model and are estimated or fixed to a particular value. The number 1 indicates that a parameter, either a path coefficient or a variance, has been set (fixed) to the value of 1. (The rationale behind "fixing" paths is discussed in the section about identification.)

This example contains two hypothesized latent variables (factors): Weak Institutional Bonds (WK_BONDS), and Acceptance of Risky Behavior (ACCEPT_RISK). The weak institutional bonds (WK_BONDS) factor is hypothesized to have two indicators, bonds to family (FAMILY_S) and bonds to teachers (TEACH_SC). Higher numbers on these measured variables indicate weaker bonds. Weak Institutional Bonds predict both weak family and teacher bonds. Note that the direction of the prediction matches the direction of the arrows. The Acceptance of Risky Behavior factor also has two indicators endorsing acceptance of smoking and drinking (OKSMOKE2, OKDRINK2). Acceptance of Risky Behavior predicts higher scores on both of these behavioral indicators. This model also hypothesizes that both Weak Institutional Bonds and gender predict level of Acceptance of Risky Behavior; weaker Institutional Bonds and being male (higher code for gender) predict higher levels of Acceptance of Risky Behavior. Also notice that no arrow directly connects Institutional Bonds with gender. There is no hypothesized relationship, either predictive or correlational, between these variables. However, we can, and we will, test the hypothesis that there is a correlation between Weak Institutional Bonds and gender.

These relationships are directly translated into equations and the model then estimated. The analysis proceeds by specifying a model as in the diagram and then translating the model into a series of equations or matrices. One method of model specification is the Bentler-Weeks method (Bentler & Weeks, 1980). In this method every variable in the model, latent or measured, is either an IV or a DV. The parameters to be estimated are (a) the regression coefficients, and (b) the variances and the covariances of the independent variables in the model (Bentler, 1989). In Figure 23.2 the regression coefficients and covariances to be estimated are indicated with an asterisk (*).

In the example, FAMILY_S, TEACH_SC, OKDRI NK2, OKSMOKE2 are all DVs because they all have at least one line with a single-headed arrow pointing to them. Notice that ACCEPT_RISK is a latent variable and also a dependent variable. Whether or not a variable is observed makes no difference as to its status as a DV or IV. Although ACCEPT_RISK is a factor, it is also a DV because it has arrows from both WK_BONDS and Gender. The seven IVs in this example are gender, WK_BONDS, and the residuals variances (D2, E1, E2, E4, E5).

Residual variables (errors) of measured variables are labeled *E* and *errors of latent variables* (called *disturbances*) are labeled *D*. It may seem odd that a residual variable is considered an IV but remember the familiar regression equation:

$$Y = X\beta + e \tag{1}$$

where Y is the DV and X and *e* are both IVs.

In fact the Bentler-Weeks model *is* a regression model, expressed in matrix algebra:

$$\eta = \beta\eta + \gamma\xi \tag{2}$$

where, if q is the number of DVs and r is the number of IVs, then η (eta) is a $q \times 1$ vector of DVs, β (beta) is a $q \times q$ matrix of regression coefficients between DVs, γ (gamma) is a $q \times r$ matrix of regression coefficients between DVs and IVs, and ξ (xi) is an $r \times 1$ vector of IVs.

What makes this model different from ordinary regression is the possibility of having latent variables as DVs and predictors, as well as the possibility of DVs predicting other DVs.

The syntax for this model estimated in EQS (a popular SEM computer package) is presented in Table 23.1. As seen in Table 23.1, the model is specified in EQS using a series of regression equations. In the /EQUATIONS section, as in ordinary regression, the DV appears on the left side of the equation, and its predictors are on the right-hand side. But unlike regression, the predictors may be IVs or other DVs. Measured variables are referred to by the letter *V* and the number corresponding to the variable given in the /LABELS section. Errors associated with measured variables are indicated by the letter E and the number of the variable. Factors are referred to with the letter F and a number given in the /LABELS section. The errors, or disturbances, associated with factors are referred to by the letter D and the number corresponding to the factor. An asterisk indicates a parameter to be estimated. Variables included in the equation without asterisks are considered parameters fixed to the value 1. In this example start values are not specified and are estimated automatically by the program through simply including an asterisk. If specific start values were required, a numerical starting value would be included in front of the asterisk. The variances of IVs are parameters of the model and are indicated in the /VAR paragraph. In the /PRINT paragraph, FIT = ALL requests

TABLE 23.1 EQS 6.1 Syntax for SEM Model of Predictors of Acceptance of Risky Behavior Presented in Figure 23.2

```
/TITLE
Acceptance of Risky Behavior
/SPECIFICATIONS
DATA='c:\data.ess';
VARIABLES=5; CASES=4578;
METHOD=ML, ROBUST; ANALYSIS=COVARIANCE;
MATRIX=RAW;
/LABELS
V1=OKDRINK2; V2=OKSMOKE2; V3=GENDER2;
V4=TEACH_SC; V5=FAMILY_S;
F1=WK_BONDS; F2=ACCEPT_RISK;
/EQUATIONS
V1 = 1F2 + E1;
V2 = *F2 + E2;
V4 = *F1 + E4;
V5 = *F1 + E5;
F2 = *F1 + *V3 + D2;
/VARIANCES
V3 = *;
F1 = 1;
E1, E2 = *;
E4, E5, = *;
D2 = *;
/COVARIANCES
/PRINT
FIT=ALL;
TABLE=EQUATION;
/LMTEST
/WTEST
/END
```

all goodness-of-fit indices available. Take a moment to confirm that the diagram relationships exactly match the regression equations given in the syntax file.

Identification. In SEM a model is specified, parameters for the model are estimated using sample data, and the parameters are used to produce the estimated population covariance matrix. But only models that are identified can be estimated. A model is said to be identified if there is a unique numerical solution for each of the parameters in the model. For example, say that the variance of $y = 10$ and that the variance of $y = \alpha + \beta$. Any two values can be substituted for α and β as long as they sum to 10. There is no unique numerical solution for either α or β; that is, there are an infinite number of combinations of two numbers that would sum to 10. Therefore this single equation model is not identified. However, if we fix α to 0, then there is a unique solution for β, 10, and the equation is identified. It is possible to use covariance algebra to calculate equations and assess identification in very simple models; however, in large models this procedure quickly becomes unwieldy. For a detailed, technical discussion of identification, see Bollen (1989). The following guidelines are rough, but may suffice for many models.

The first step is to count the numbers of data points and the number of parameters that are to be estimated. *The data in SEM are the variances and covariances in the sample covariance matrix*. The number of data points is the number of nonredundant sample variances and covariances,

$$\text{Number of data points} = \frac{p\,(p+1)}{2}, \qquad (3)$$

where p equals the number of measured variables.

The number of parameters is found by adding together the number of regression coefficients, variances, and covariances that are to be estimated (i.e., the number of asterisks in a diagram).

If there are more data points than parameters to be estimated, the model is said to be overidentified, a necessary condition for proceeding with the analysis. If there are the same numbers of data points as parameters to be estimated, the model is said to be just-identified. In this case, the estimated parameters perfectly reproduce the sample covariance matrix, chi-square and degrees of freedom are equal to zero, and the analysis is uninteresting because hypotheses about adequacy of the model cannot be tested. However, hypotheses about specific paths in the model can be tested. If there are fewer data points than parameters to be estimated, the model is said to be underidentified and parameters cannot be estimated. The number of parameters needs to be reduced by fixing, constraining, or deleting some of them. A parameter may be fixed by setting it to a specific value or constrained by setting the parameter equal to another parameter.

In the acceptance of risky behavior example of Figure 23.2, there are five measured variables, so there are 15 data points: $5(5 + 1)/2 = 15$ (5 variances and 10 covariances). There are 11 parameters to be estimated in the hypothesized model: five regression coefficients and six variances. The hypothesized model has four fewer parameters than data points, so the model may be identified.

The second step in determining model identifiability is to examine the measurement portion of the model. The measurement part of the model deals with the relationship

between the measured indicators and the factors. It is necessary both to establish the scale of each factor and to assess the identifiability of this portion of the model.

To establish the scale of a factor, either the variance of the factor is set to 1, or one of the regression coefficients from the factor to a measured variable is fixed to 1. Fixing the regression coefficient to 1 gives the factor the same variance as the measured variable. If the factor is an IV, either alternative is acceptable. If the factor is a DV, most researchers fix the regression coefficient to 1. In the example, the variance of the Weak Institutional Bonds factor was set to 1 (normalized) while the scale of the Acceptance of Risky Behavior factor was set equal to the scale of okay to drink.

To establish the identifiability of the measurement portion of the model look at the number of factors and the number of measured variables (indicators) loading on each factor. If there is only one factor, the model may be identified if the factor has at least three indicators with nonzero loading and the errors (residuals) are uncorrelated with one another. If there are two or more factors, again consider the number of indicators for each factor. If each factor has three or more indicators, the model may be identified if errors associated with the indicators are not correlated, each indicator loads on only one factor, and the factors are allowed to covary. If there are only two indicators for a factor, the model may be identified if there are no correlated errors, each indicator loads on only one factor, and none of the covariances among factors is equal to zero.

In the example, there are two indicators for each factor. The errors are uncorrelated and each indicator loads on only one factor. Additionally, the covariance between the factors is not zero. Therefore, this part of the model may be identified. Please note that identification may still be possible if errors are correlated or variables load on more than one factor, but it is more complicated.

The third step in establishing model identifiability is to examine the structural portion of the model, looking only at the relationships among the latent variables (factors). Ignore the measured variables for a moment; consider only the structural portion of the model that deals with the regression coefficients relating latent variables to one another. If none of the latent DVs predict each other (the beta matrix is all zeros), the structural part of the model may be identified. This example has only one latent DV, so that part of the model may be identified. If the latent DVs do predict one another, look at the latent DVs in the model and ask if they are recursive or nonrecursive. If the latent DVs are recursive, there are no feedback loops among them, and there are no correlated disturbances (errors) among them. (In a feedback loop, DV1 predicts DV2 and DV2 predicts DV1. That is, there are two lines linking the factors, one with an arrow in one direction and the other line with an arrow in the other direction. Correlated disturbances are linked by single curved lines with double-headed arrows.) If the structural part of the model is recursive, it may be identifiable. These rules also apply to path analysis models with only measured variables. The acceptance of risky behavior example is a recursive model and therefore may be identified.

If a model is nonrecursive, either there are feedback loops among the DVs or there are correlated disturbances among the DVs, or both. Two additional conditions are necessary for identification of nonrecursive models, each applying to each equation in the model separately. Look at each equation separately; for identification it is necessary that each equation not contain all of the latent DVs. One latent DV must be excluded from each equation. The second condition is that the *information matrix* (a matrix necessary for calculating standard errors) is full rank and can be inverted. The inverted information matrix can be examined in the output from most SEM programs. If, after examining the model, the number of data points exceeds the number of parameters estimated and both the measurement and structural parts of the model are identified, there is good evidence that the whole model is identified.

Sample size. Covariances are less stable when estimated from small samples. SEM is based on covariances. Parameter estimates and chi-square tests of fit are also sensitive to sample size. Therefore SEM is a large sample technique. Velicer and Fava (1998) and MacCallum, Widaman, Preacher, and Hong (1999) found, in exploratory factor analysis models, that the size of the factor loadings, the number of variables, and the size of the sample were important elements in obtaining a good factor model. This can reasonably be generalized to SEM models. Models with strong expected parameter estimates, reliable measured variables, and well-defined constructs may require less data (Ullman, 2007). Interestingly, although SEM is a large data technique new test statistics have been developed that allow for estimation of small models with as few as 60 respondents (Yuan & Bentler, 1999).

Power. Two general approaches are available for power estimation in SEM. The MacCallum, Browne, and Sugawara (1996) approach estimates power relative to an alternative hypothesis specified in terms of lack of fit. In the MacCallum et al. approach power is estimated based on the degrees of freedom (*df*s) of the model and the

root mean square error of approximation (RMSEA). This approach allows power estimation for the fit (or lack of fit for the model). The Satorra–Saris (1985) approach to power estimates the power to reject specific hypotheses about parameters of the models and employs comparisons of nested models (models that are subsets of one another).

Missing data. Problems of missing data are often magnified in SEM due to the large number of measured variables employed (Allison, 2003; Enders, 2010, Little & Rubin, 2002; Schafer & Graham, 2002). The researcher who relies on using complete cases only is often left with an inadequate number of complete cases to estimate a model and potentially biased estimated parameters. Therefore missing data imputation is particularly important in SEM models. When there is evidence that the data are missing at random (MAR, missingness may depend on observed data) or missing completely at random (MCAR, missingness is unrelated to observed data or the missing data mechanism), a preferred method of imputing missing data, the EM algorithm to obtain maximum likelihood (ML) estimates, is appropriate (Little & Rubin). A full discussion of the EM algorithm is outside the scope of this chapter but the general idea behind the EM approach is that, with respect to the likelihood function, missing values are replaced with expectations given the likelihood function and parameters are estimated iteratively. Using this iterative process yields missing data estimates that have a statistically unbiased mean and variance. Software packages routinely include procedures for estimating missing data. EQS 6.1 (Bentler, 2008) produces the EM-based maximum likelihood solution automatically based on the Jamshidian-Bentler (1999) computations. It should be noted that, if the data are not normally distributed, maximum likelihood test statistics—including those based on the EM algorithm—may be quite inaccurate.

Additionally, a missing data mechanism can be explicitly modeled within the SEM framework. Treatment of missing data patterns through SEM is not demonstrated in this chapter but the interested reader is referred to Allison (1987) and Muthén, Kaplan, and Hollis (1987). Multiple imputation (MI) is also a viable solution when the data meet normality assumptions. However, when the data violate normality the parameter estimates from the MI approach have more bias than those from the ML approach (Yuan, Wallentin, & Bentler, 2011).

Normality is a restrictive assumption in practice. The more general case on how to deal with missing data when the parent distribution is possibly non-normal is discussed in Yuan and Bentler (2000a). They provide a means for accepting the EM-based estimates of parameters, but correcting standard errors and test statistics for non-normality in an approach reminiscent of Satorra-Bentler (1994). Their approach has been uniquely incorporated into the EQS 6.1 program (Bentler, 2008).

Multivariate normality and outliers. Most of the estimation techniques used in SEM assume multivariate normality. To determine the extent and shape of non-normally distributed data, examine the data for evidence of outliers, both univariate and multivariate, and evaluate the skewness and kurtosis of the distributions for the measured variables. If significant skewness is found, transformations can be attempted; however, often variables are still highly skewed or highly kurtotic even after transformation. Some variables, such as drug-use variables, are not expected to be normally distributed in the population. If transformations do not restore normality, or a variable is not expected to be normally distributed in the population, an estimation method can be selected that addresses the non-normality.

Residuals. After model estimation, the residuals should be small and centered around zero. The frequency distribution of the residual covariances should be symmetric. Residuals in the context of SEM are residual *covariances,* not residual *scores,* differences between sample covariances and those reproduced by the model. Nonsymmetrically distributed residuals in the frequency distribution may signal a poorly fitting model; the model is estimating some of the covariances well and others poorly. It sometimes happens that one or two residuals remain quite large, although the model fits reasonably well and the residuals appear to be symmetrically distributed and centered around zero. Typically, more informative than the ordinary residuals are the residuals obtained after standardizing the sample covariance matrix to a correlation matrix and similarly transforming the model matrix. In this metric, it is correlations that are being reproduced, and it is easy to see whether a residual is small and meaningless or too large for comfort. For example, if a sample correlation is .75 and the corresponding residual is .05, the correlation is largely explained by the model. In fact, an average of these standardized root mean square residuals (SRMS) has been shown to provide one of the most informative guides to model adequacy (Hu & Bentler, 1998, 1999).

MODEL ESTIMATION TECHNIQUES AND TEST STATISTICS

After a model is specified, population parameters are estimated with the goal of minimizing the difference between

the observed and estimated population covariance matrices. To accomplish this goal, a function, *F,* is minimized where

$$F = (\mathbf{s} - \sigma(\Theta))' \mathbf{W} (s - \sigma(\Theta)), \quad (4)$$

s is the vector of data (the observed sample covariance matrix stacked into a vector); σ is the vector of the estimated population covariance matrix (again, stacked into a vector), and **(Θ)** indicates that σ is derived from the parameters (the regression coefficients, variances, and covariances) of the model. **W** is the matrix that weights the squared differences between the sample and estimated population covariance matrix.

In factor analysis the observed and reproduced correlation matrices are compared. This idea is extended in SEM to include a statistical test of the differences between the observed covariance matrix and the covariance matrix that is produced as a function of the model. If the weight matrix, **W**, is chosen correctly, at the minimum with the optimal $\hat{\Theta}$, F multiplied by $(N - 1)$ yields a chi-square test statistic.

The trick is to select **W** so that the sum of weighted squared differences between observed and estimated population covariance matrices has a statistical interpretation. In an ordinary chi-square, the weights are the set of expected frequencies in the denominators of the cells. If we use some other numbers instead of the expected frequencies, the result might be some sort of test statistic, but it would not be a $\chi 2$ statistic; that is, the weight matrix would be wrong.

In SEM, estimation techniques vary by the choice of **W.** Unweighted least squares estimation (ULS) does not standardly yield a $\chi 2$ statistic or standard errors, though these are provided in EQS. ULS estimation does not usually provide the best estimates, in the sense of having the smallest possible standard errors, and hence is not discussed further (see Bollen, 1989, for further discussion of ULS).

Maximum likelihood (ML) is usually the default method in most programs because it yields the most precise (smallest variance) estimates when the data are normal. GLS (generalized least squares) has the same optimal properties as ML under normality. When data are symmetrically distributed but normal, an option is EDT (elliptical distribution theory, Shapiro & Browne, 1987). The ADF (asymptotically distribution free) method has no distributional assumptions and hence is most general (Browne, 1984), but it is impractical with many variables and inaccurate without large sample sizes. Satorra and Bentler (1994, 2001) and Satorra (2000) have also developed an adjustment for non-normality that can be applied to the ML, GLS, or EDT chi-square test statistics. Briefly, Satorra-Bentler Scaled χ^2 is a correction to the χ^2 test statistic.[1] EQS also corrects the standard errors for parameter estimates to adjust for the extent of non-normality (Bentler & Dijstra, 1985).

The performance of the χ^2 test statistic derived from these different estimation procedures is affected by several factors, among them (1) sample size, (2) non-normality of the distribution of errors, of factors, and of errors and factors, and (3) violation of the assumption of independence of factors and errors. The goal is to select an estimation procedure that, in Monte Carlo studies, produces a test statistic that neither rejects nor accepts the true model too many times. Several studies provide guidelines for selection of appropriate estimation method and test statistics. The following sections summarize the performance of estimation procedures examined in Monte Carlo studies by Hu, Bentler, and Kano (1992) and Bentler and Yuan (1999). Hu et al. varied sample size from 150 to 5,000 and Bentler and Yuan examined samples sizes ranging from 60 to 120. Both studies examined the performance of test statistics derived from several estimation methods when the assumptions of normality and independence of factors were violated.

Estimation methods/test statistics and sample size. Hu and colleagues found that when the normality assumption was reasonable, both the ML and the Scaled ML performed well with sample sizes more than 500. When the sample size was less than 500, GLS performed slightly better. Interestingly the EDT test statistic performed a little better than ML at small sample sizes. It should be noted that the elliptical distribution theory estimator (EDT) considers the kurtosis of the variables and assumes that all variables have the same kurtosis, although the variables need not be normally distributed. (If the distribution is normal, there is no excess kurtosis.) Finally, the ADF estimator was poor with sample sizes less than 2,500.

In small samples in the range of 60 to 120, when the number of subjects was greater than the number (p^*) of nonredundant variances and covariances in the sample covariance matrix (i.e., $p^* = [p(p + 1)]/2$ where p is the number of variables), Bentler and Yuan found that a

[1]The test statistic is adjusted for by degrees of freedom in the model/estimate of the sum of the nonzero eigenvalues of the product residual weight matrix under the model and the weight matrix used in the estimation and the asymptotic covariance matrix of the differences between the sample covariance matrix and the estimated population covariance matrix.

test statistic based on an adjustment of the ADF estimator and evaluated as an F statistic was best. This test statistic (Yuan-Bentler, 1999) adjusts the chi-square test statistic derived from the ADF estimator as,

$$T_l = \frac{\left[N - (p^* - q)\right] T_{ADF}}{\left[(N-1)(p^* - q)\right]}, \quad (5)$$

where N is the number of subjects, q is the number of parameters to be estimated, and TADF is the test statistic based on the ADF estimator.

Estimation Methods and Non-Normality

When the normality assumption was violated, Hu et al. (1992) found that the ML and GLS estimators worked well with sample sizes of 2,500 and greater. The GLS estimator was a little better with smaller sample sizes but led to acceptance of too many models. The EDT estimator accepted far too many models. The ADF estimator was poor with sample sizes less than 2,500. Finally, the scaled ML performed about the same as the ML and GLS estimators and better than the ADF estimator at all but the largest sample sizes.[2] With small samples sizes the Yuan-Bentler test statistic performed best.

Estimation Methods and Dependence

The assumption that errors are independent underlies SEM and other multivariate techniques. Hu, Bentler, and Kano (1992) also investigated estimation methods and test statistic performance when the errors and factors were dependent but uncorrelated.[3] ML and GLS performed poorly, always rejecting the true model. ADF was poor unless the sample size was greater than 2,500. EDT was better than ML, GLS, and ADF, but still rejected too many true models. The Scaled ML was better than the ADF at all but the largest sample sizes. The Scaled ML performed best overall with medium to larger sample sizes; the Yuan-Bentler performed best with small samples.

[2]This is interesting in that the ADF estimator has no distributional assumptions and, theoretically, should perform quite well under conditions of non-normality.

[3]Factors were dependent but uncorrelated by creating a curvilinear relationship between the factors and the errors. Correlation coefficients examine only linear relationships; therefore, although the correlation is zero between factors and errors, they are dependent.

Some Recommendations for Choice of Estimation Method/Test Statistic

Sample size and plausibility of the normality and independence assumptions need to be considered in selection of the appropriate estimation technique. ML, the Scaled ML, or GLS estimators may be good choices with medium to large samples and evidence of the plausibility of the normality assumptions. The independence assumption cannot be routinely evaluated. ML estimation is currently the most frequently used estimation method in SEM. In medium to large samples the Scaled ML test statistic is a good choice with non-normality or suspected dependence among factors and errors. Because the scaled ML is computer intensive and many model estimations may be required, it is often reasonable to use ML during model estimation and then the scaled ML for the final estimation. In small samples the Yuan-Bentler test statistic seems best. The test statistic based on the ADF estimator (without adjustment) seems like a poor choice under all conditions unless the sample size is very large (>2,500). Similar conclusions were found in studies by Fouladi (2000), Hoogland (1999), and Satorra (1992).

Computer procedure and interpretation. The data used in this example are from a large evaluation of the D.A.R.E. program in Colorado Springs (N = 4,578 students). Details about these data can be found in Dukes, Stein, and Ullman (1997). The model in Figure 23.2 is estimated using ML estimation and evaluated with the Satorra-Bentler scaled chi-square because there was evidence of violation of multivariate normality (Mardia's normalized coefficient = 238.65, $p < .001$). This normalized coefficient is distributed as a z test; therefore, in large samples normalized coefficients greater than 3.3 may indicate violations for normality. In Table 23.1 the estimation method is indicated after ME =. Output for the Mardia's coefficient, model estimation, and chi-square test statistic is given in Table 23.2.

The output first presents model information given normality. Scanning down the table the model information appropriate for this model given the normality violation begins with GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST. Several chi-square test statistics are given in this model estimation and evaluation section presented in Table 23.2. The ROBUST INDEPENCENCE MODEL CHI-SQUARE = 730.858, with 10 *df*s, tests the hypothesis that the measured variables are orthogonal. Therefore, the probability associated with this chi-square should be small, typically less than .05. The model chi-square test statistic is labeled SATORRA-BENTLER

TABLE 23.2 Selected Output From EQS 6.1 for Model Estimation of SEM Model of Acceptance of Risky Behavior Presented in Figures 23.2 and 23.3

MULTIVARIATE KURTOSIS

———————

MARDIA'S COEFFICIENT (G2,P) = 61.0270

NORMALIZED ESTIMATE = 238.6527

——– much output omitted ———

GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE = 1661.460 ON 10 DEGREES OF FREEDOM

INDEPENDENCE AIC = 1641.460 INDEPENDENCE CAIC = 1567.838

MODEL AIC = 23.078 MODEL CAIC = −6.371

CHI-SQUARE = 31.078 BASED ON 4 DEGREES OF FREEDOM

PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00000

THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS 31.364.

FIT INDICES

————–

BENTLER-BONETT NORMED FIT INDEX = 0.981

BENTLER-BONETT NON-NORMED FIT INDEX = 0.959

COMPARATIVE FIT INDEX (CFI) = 0.984

BOLLEN'S (IFI) FIT INDEX = 0.984

MCDONALD'S (MFI) FIT INDEX = 0.997

JöRESKOG-SöRBOM'S GFI FIT INDEX = 0.997

JöRESKOG-SöRBOM'S AGFI FIT INDEX = 0.989

ROOT MEAN-SQUARE RESIDUAL (RMR) = 0.009

STANDARDIZED RMR = 0.021

ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.040

90% CONFIDENCE INTERVAL OF RMSEA (0.027, 0.053)

RELIABILITY COEFFICIENTS

———————————

CRONBACH'S ALPHA = 0.472

GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

ROBUST INDEPENDENCE MODEL CHI-SQUARE = 730.858 ON 10 DEGREES OF FREEDOM

INDEPENDENCE AIC = 710.858 INDEPENDENCE CAIC = 637.237

MODEL AIC = 13.644 MODEL CAIC = −15.805

SATORRA-BENTLER SCALED CHI-SQUARE = 21.6436 ON 4 DEGREES OF FREEDOM

PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00024

MEAN- AND VARIANCE-ADJUSTED CHI-SQUARE = 19.035 ON 4 D.F.

PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00077

RESIDUAL-BASED TEST STATISTIC = 27.386

PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00002

YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC = 27.212

PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.00002

YUAN-BENTLER RESIDUAL-BASED F-STATISTIC = 6.842

DEGREES OF FREEDOM = 4, 4278

PROBABILITY VALUE FOR THE F-STATISTIC IS 0.00002

FIT INDICES

————–

BENTLER-BONETT NORMED FIT INDEX = 0.970

BENTLER-BONETT NON-NORMED FIT INDEX = 0.939

COMPARATIVE FIT INDEX (CFI) = 0.976

BOLLEN'S (IFI) FIT INDEX = 0.976

MCDONALD'S (MFI) FIT INDEX = 0.998

ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.032

90% CONFIDENCE INTERVAL OF RMSEA (0.020, 0.046)

CHI-SQUARE = 19.035 BASED ON 4 DEGREES OF FREEDOM. This information tests the hypothesis that the difference between the estimated population covariance matrix and the sample covariance matrix is not significant. Ideally the probability associated with this chi-square should be large, greater than .05. In Table 23.2 the probability associated with the model chi-square is .00077. This significance indicates that the model does not fit the data. However this is a large sample and small, trivial differences often create significant chi-squares. Recall that the model chi-square is calculated as N^* fmin. For this reason model evaluation relies heavily on other fit indices.

MODEL EVALUATION

Two general aspects of a model are evaluated: (1) the overall fit of the model, and (2) significance of particular parameters of the model (regression coefficients and variances and covariances of independent variables).

Evaluating the overall fit of the model. The model chi-square is highly dependent on sample size; that is, the model chi-square is $(N - 1)$Fmin where N is the sample size and Fmin is the value of Fmin, Equation 4, at the function minimum. Therefore, the fit of models estimated with large samples is often difficult to assess. Fit indices have been developed to address this problem. There are five general classes of fit indices: comparative fit, absolute fit, proportion of variance accounted for, parsimony adjusted proportion of variance accounted for, and residual-based fit indices. A complete discussion of model fit is outside the scope of this chapter; therefore we focus on two of the most popular fit indices: the Comparative Fit Index (Bentler, 1990) and a residual-based fit index, the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993). Ullman (2007), Bentler and Raykov (2000), and Hu and Bentler (1999) offer more detailed discussions of fit indices.

Nested models are models that are subsets of one another. At one end of the continuum is the uncorrelated variables or independence model: the model that corresponds to completely unrelated variables. This model would have degrees of freedom equal to the number of data points minus the variances that are estimated. At the other end of the continuum is the saturated (full or perfect) model with zero degrees of freedom. Fit indices that employ a comparative fit approach place the estimated model somewhere along this continuum, with 0.00 indicating awful fit and 1.00 indicating perfect fit.

The comparative fit index (CFI; Bentler, 1990) also assesses fit relative to other models as the name implies, but uses a different approach. The CFI employs the noncentral $\chi 2$ distribution with noncentrality parameters, τ_i. If the estimated model is perfect, $\tau_i = 0$, therefore, the larger the value of τ_i, the greater the model misspecification.

$$\text{CFI} = 1 - \frac{\tau_{\text{est. model}}}{\tau_{\text{indep. model}}} \tag{6}$$

So, clearly, the smaller the noncentrality parameter, τ_i, for the estimated model relative to the τ_i, for the independence model, the larger the CFI and the better the fit. The τ value for a model can be estimated by

$$\begin{aligned} \hat{\tau}_{\text{indep. model}} &= \chi^2_{\text{indep. model}} - \text{df}_{\text{indep. model}} \\ \hat{\tau}_{\text{est. model}} &= \chi^2_{\text{est. model}} - \text{df}_{\text{est. model}}, \end{aligned} \tag{7}$$

where $\hat{\tau}_{\text{est. model}}$ is set to zero if negative.

For the example,

$$\begin{aligned} \tau_{\text{indep. model}} &= 730.858 - 10 = 720.858 \text{ and} \\ \tau_{\text{est. model}} &= 21.6436 - 4 = 17.6436 \text{ so that} \\ CFI &= 1 - \frac{17.6436}{720.858} = .976. \end{aligned}$$

CFI values greater than .95 are often indicative of good fitting models (Hu & Bentler, 1999). The CFI is normed to the 0 – 1 range, and does a good job of estimating model fit even in small samples (Hu & Bentler, 1998, 1999).

The root mean square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger, 2000) estimates the lack of fit in a model compared to a perfect or saturated model by

$$\text{estimated RMSEA} = \sqrt{\frac{\hat{\tau}}{Ndf_{\text{model}}}} \tag{8}$$

where $\hat{\tau} = \hat{\tau}_{\text{est. model}}$ as defined in Equation 7. As noted above, when the model is perfect, $\hat{\tau} = 0$, and the greater the model misspecification, the larger $\hat{\tau}$. Hence RMSEA is a measure of noncentrality relative to sample size and degrees of freedom. For a given noncentrality, large N and df imply a better fitting model, that is, a smaller RMSEA. Values of .06 or less indicate a close-fitting model (Hu & Bentler, 1999). Values larger than .10 are indicative of poor-fitting models (Browne & Cudeck, 1993). Hu and Bentler (1999) found that in small samples the RMSEA overrejected the true model; that is, its value

was too large. Because of this problem, this index may be less preferable with small samples. As with the CFI, the choice of estimation method affects the size of the RMSEA.

For the example, $\hat{\tau} = 17.6436$, therefore

$$\text{RMSEA} = \sqrt{\frac{17.6436}{17140}} = .032.$$

Both the CFI and RMSEA exceed cut-off values of .95 and .06, respectively, and we may conclude that despite the significant chi-square the model fits.

Interpreting Parameter Estimates—Direct Effects

Given the fit indices there is clear evidence that the model fits well, but what does it mean? The hypothesis that the observed covariances among the measured variables arose because of the linkages between variables specified in the model is supported by fit indices. Note that the chi-square is significant, so in absence of fit indices we would conclude that we should reject our hypothesized model. However, the chi-square test, especially with a large sample such as this, is a notoriously bad measure of fit. The model chi-square is calculated as $(N - 1)$*minimum of the function. Therefore, trivial differences between the sample covariance matrix and the estimated population covariance matrix can force the chi-square to exceed the threshold for significance. Although chi-square test statistics are still routinely reported, more emphasis is placed on fit indices particularly with large samples.

Next, researchers usually examine the statistically significant relationships within the model. Table 23.3 contains edited EQS output for evaluation of the regression coefficients for the example. If the unstandardized parameter estimates are divided by their respective standard errors, a z score is obtained for each estimated parameter that is evaluated in the usual manner,[4]

$$z = \frac{\text{parameter estimate}}{\text{std error for estimate}}. \quad (9)$$

Because of differences in scales, it is sometimes difficult to interpret unstandardized regression coefficients. Therefore, researchers often examine standardized coefficients. Both the standardized and unstandardized regression coefficients for the final model are in Table 23.3 and Figure 23.3. In Figure 23.3 the standardized coefficients are in parentheses. Looking at Table 23.3 in the section labeled MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS, for each dependent variable there are four pieces of information: The unstandardized coefficient is given on the first line, the standard error of the coefficient given normality is given on the second line, the standard error of the coefficient adjusted to the degree of the non-normality is given on the third line, and the test statistic (z score) for

TABLE 23.3 Parameter Estimates, Standard Errors, and Test Statistics for Hypothetical Example

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

(ROBUST STATISTICS IN PARENTHESES)

OKDRINK2=V1 = 1.000 F2 + 1.000 E1

OKSMOKE2=V2 = 1.563*F2 + 1.000 E2

.108

14.535@

(.157)

(9.945@

TEACH_SC=V4 = .482*F1 + 1.000 E4

.021

22.676@

(.027)

(18.069@

FAMILY_S=V5 = .412*F1 + 1.000 E5

.019

22.072@

(.024)

(17.446@

CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

(ROBUST STATISTICS IN PARENTHESES)

ACCEPT_R=F2 = −.020*V3 + .185*F1 + 1.000 D2

.011 .013

−1.775 14.016@

(.013) (.018)

(-1.544) (10.189@

[4]The standard errors are derived from the inverse of the information matrix.

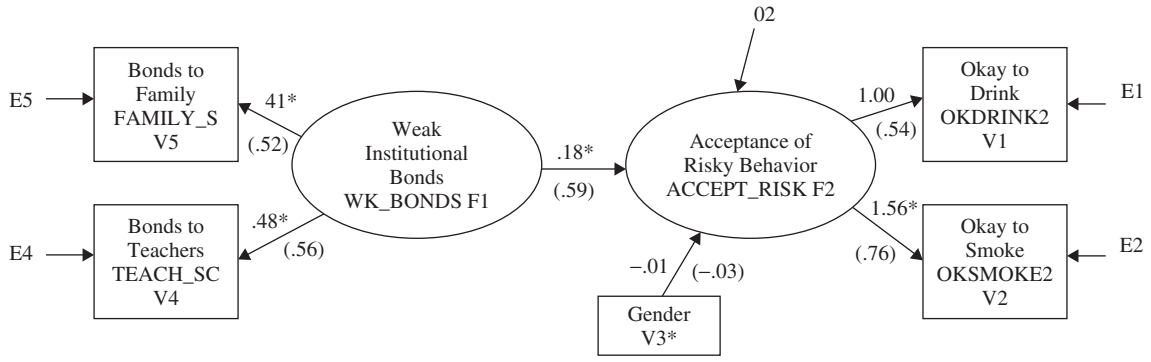

Figure 23.3 Example with unstandardized and standardized coefficients (standardized coefficients in parentheses)

the coefficient is given on the last line. For example, for FAMILY_S predicted from WK_BONDS,

$$\frac{.41}{.024} = 17.45,\ p < .05.$$

It could be concluded that bonds to family (FAMILY_S) is a significant indicator of Weak Institutional Bonds (WK_BONDS); the weaker the Institutional Bonds the weaker the bonds to family. Bonds to teachers (TEACH_SC) is also a significant indicator of Weak Institutional Bonds. Endorsement of smoking (OKSMOKE2) is a significant indicator of Acceptance of Risky Behavior (ACCEPT_RISK); greater acceptance of risky behavior predicts stronger endorsement of smoking (unstandardized coefficient = 1.56, $z = 9.45$, $p < .05$). Because the path from ACCEPT_RISK to OKDRINK is fixed to 1 for identification, a standard error is not calculated.

As seen in Table 23.3, the relationships between the constructs appear in the EQS section labeled, CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS. Weak Institutional Bonds (WK_BONDS) significantly predicts greater Acceptance of Risky Behavior (unstandardized coefficient = .185, standard error = .018, $z = 10.19$, $p < .05$). Gender does not significantly predict Acceptance of Risky Behavior.

Indirect effects. A particularly strong feature of SEM is the ability to test not only direct effects between variables but also indirect effects. Mediational hypotheses are not illustrated in this example, but a simple example is shown in Figure 23.4. Imagine that students are assigned to one of two teaching methods for a statistics class (coded 0 and 1). Final exam scores are recorded at the end of the quarter. The direct effect of teaching method on exam score is path *a.* But is it reasonable to suggest that mere assignment to a teaching method creates the change? Perhaps not. Maybe, instead, the teaching method increases a student's motivational level and higher motivation leads to a higher grade. The relationship between the treatment and the exam score is *mediated* by motivation level. That is to say that type of teaching method indirectly affects final exam score through level of motivation. Or, level of motivation serves as an intervening variable between teaching method and final exam score. Note that this is a different question than is posed with a direct effect: "Is there a difference between the treatment and control group on exam score?" The indirect effect can be tested by testing the product of paths *b* and *c.* This example uses only

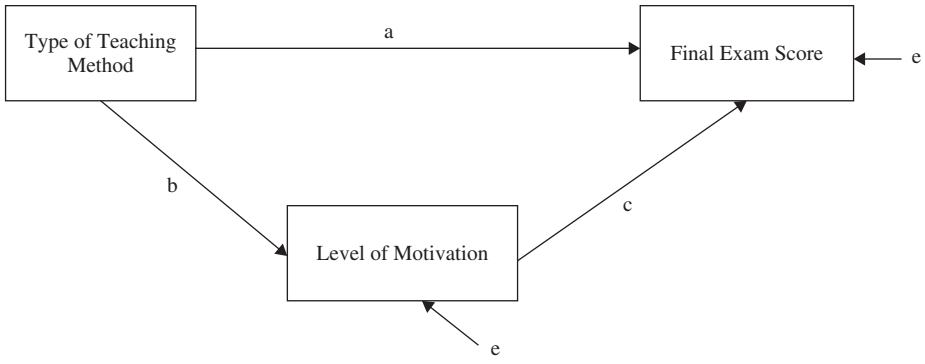

Figure 23.4 Path analysis model with indirect effect

measured variables and is called *path analysis*; however, mediational hypotheses can be tested using both latent and observed variables. A more detailed discussion of indirect effects can be found in MacKinnon, Lockwood, Hoffman, West, and Sheets (2002), MacKinnon, Fairchild, and Fritz, (2007), and MacKinnon (2008). Indirect effects are readily obtainable in the EQS 6.1 program by specifying "effects = yes" in the /PRINT section.

MODEL MODIFICATION

There are at least two reasons for modifying a SEM model: to improve fit (especially in exploratory work) and to test hypotheses (in theoretical work). The three basic methods of model modification are the chi-square difference, Lagrange multiplier (LM), and Wald test. All are asymptotically equivalent under the null hypothesis but approach model modification differently. Because of the relationship between sample size and χ^2, it is hard to detect a difference between models when sample sizes are small.

Chi-square difference test. If models are nested (models are subsets of each other), the χ^2 value for the larger model is subtracted from the χ^2 value for the smaller nested model and the difference, also a χ^2, is evaluated with degrees of freedom equal to the difference between the degrees of freedom in the two models.

Recall in Figure 23.3 the covariance between gender and Institutional Bonds was fixed to zero. We might allow these IVs to correlate and ask, "Does adding (estimating) this covariance improve the fit of the model?" Although our "theory" is that these variables are uncorrelated, is this aspect of theory supported by the data? To examine these questions, a second model is estimated in which Institutional Bonds and gender are allowed to correlate. The resulting $\chi^2 = 10.83$, $df = 3$. In the original model the Satorra-Bentler $\chi^2 = 21.64$, $df = 4$. The χ^2 difference test (or likelihood ratio test for maximum likelihood) is evaluated with *df*s equal to the difference between the models, $df = 4 - 3 = 1$, $p < .05$. Had the data been normally distributed the chi-squares could have simply been subtracted. However, due to the non-normality, the Satorra-Bentler scaled chi-square was employed. When using the S-B chi-square an adjustment to the chi-square difference test is needed (Satorra & Bentler, 2001). After applying the adjustment, $S - B\chi^2_{\text{difference}}$ ($N = 4{,}282$, $df = 1$) $= 14.06$, $p < .01$ and we concluded that model is significantly improved with the addition of this covariance. Although the theory specifies independence between gender and Institutional Bonds, the data support the notion that, indeed, these variables are correlated. Note: In the absence of strong theory to the contrary, it is probably a good idea to always allow the independent measured variables and factors to correlate. When a DV is repeatedly measured such as in a longitudinal study, it may also be a good idea to correlate its associated residual errors.

There is a disadvantage to the χ^2 difference test. Two models need to be estimated to get the χ^2 difference value, and estimating two models for each parameter is time consuming with large models and/or a slow computer.

Lagrange Multiplier Test (LM). The LM test also compares nested models but requires estimation of only one model. The LM test asks if the model would be improved if one or more of the parameters in the model that are currently fixed are estimated. Or, equivalently, What parameters should be added to the model to improve the fit?

The LM test applied to the example indicates that if we add a covariance between gender and Institutional Bonds, the expected drop in χ^2 value is 13.67. This is one path, so the χ^2 value of 13.67 is evaluated with 1 *df*. The *p* level of this difference is $p < .01$, implying that keeping the covariance at zero is not appropriate in the population. If the decision is made to add the path, the model is reestimated. When the path is added, the actual χ^2 drop is slightly larger, 14.06, but yields the same result.

The LM test can be examined either univariately or multivariately. There is a danger in examining only the results of univariate LM tests because overlapping variance between parameter estimates may make several parameters appear as if their addition would significantly improve the model. All significant parameters are candidates for inclusion by the results of univariate LM tests, but the multivariate LM test identifies the single parameter that would lead to the largest drop in model χ^2 and calculates the expected change in χ^2. After this variance is removed, the next parameter that accounts for the largest drop in model χ^2 is assessed, similarly. After a few candidates for parameter additions are identified, it is best to add these parameters to the model and repeat the process with a new LM test, if necessary.

Wald test. The LM test asks which parameters, if any, should be added to a model, but the Wald test asks which, if any, could be deleted. Are there any parameters that are currently being estimated that could, instead, be fixed to zero? Or, equivalently, which parameters are not necessary in the model? The Wald test is analogous to backward deletion of variables in stepwise regression, where one seeks a nonsignificant change in R^2 when variables are left out.

When the Wald test is applied to the example, the only candidate for deletion is the path predicting Acceptance of Risky Behavior from gender. If this parameter is dropped, the χ^2 value increases by 2.384, a nonsignificant change ($p = .123$). The model is not significantly degraded by deletion of this parameter. However, because this was a key hypothesized path, the path is kept. Notice that unlike the LM test, *nonsignificance* is desired when using the Wald test. This illustrates an important point. Both the LM and Wald tests are based on statistical, not substantive, criteria. If there is conflict between these two criteria, substantive criteria are more important.

Some caveats and hints on model modification. Because both the LM test and Wald test are stepwise procedures, Type I error rates are inflated but there are, as yet, no available adjustments as in ANOVA. A simple approach is to use a conservative probability value (say, $p < .01$) for adding parameters with the LM test. Cross validation with another sample is also highly recommended if modifications are made. If numerous modifications are made and new data are not available for cross-validation, compute the correlation between the estimated parameters from the original, hypothesized, model and the estimated parameters from the final model using only parameters common to both models. If this correlation is high (>.90), relationships within the model have been retained despite the modifications.

Unfortunately, the order that parameters are freed or estimated can affect the significance of the remaining parameters. MacCallum (1986) suggests adding all necessary parameters before deleting unnecessary parameters. In other words, do the LM test before the Wald test.

A more subtle limitation is that tests leading to model modification examine overall changes in χ^2, not changes in individual parameter estimates. Large changes in χ^2 are sometimes associated with small changes in parameter estimates. A missing parameter may be statistically needed but the estimated coefficient may have an uninterpretable sign. If this happens, it may be best not to add the parameter, although the unexpected result may help to pinpoint problems with one's theory. Finally, if the hypothesized model is wrong, tests of model modification, by themselves, may be insufficient to reveal the true model. In fact, the "trueness" of any model is never tested directly, although cross validation does add evidence that the model is correct. Like other statistics, these tests must be used thoughtfully.

If model modifications are done in hopes of developing a good-fitting model, the fewer modifications the better, especially if a cross-validation sample is not available. If the LM test and Wald tests are used to test specific hypotheses, the hypothesis will dictate the number of necessary tests.

MULTIPLE GROUP MODELS

The example shown in this chapter uses data from a single sample. It is also possible to estimate and compare models that come from two or more samples, called multiple group models (Jöreskog, 1971; Sörbom, 1974). The general null hypothesis tested in multiple group models is that the data from each group are from the same population. For example, if data are drawn from a sample of boys and a sample for girls for the Acceptance of Risky Behavior model, the general null hypothesis tested is that the two groups are drawn from the same population. If such a restrictive model was acceptable, a single model and model reproduced covariance matrix would approximate the two sample covariance matrices for girls and boys. Typically, identical models do not quite fit, and some differences between models must be allowed.

The analysis begins by developing good-fitting models in separate analyses for each group. The models are then tested in one run with none of the parameters across models constrained to be equal. This unconstrained multiple-group model serves as the baseline against which to judge more restricted models. Following baseline model estimation, progressively more stringent constraints are specified by constraining various parameters across all groups. When parameters are constrained they are forced to be equal to one another. In EQS, an LM test is available to evaluate whether the constraint is acceptable or needs to be rejected. The same result can be obtained by a chi-square difference test. The goal is to not degrade the models by constraining parameters across the groups; therefore, you want a *nonsignificant* χ^2. If a significant difference in χ^2 is found between the models at any stage, the LM test can be examined to locate the specific parameters that are different in the groups. Such parameters should remain estimated separately in each group, that is, the specific *across group* parameter constraints are released.

Hypotheses are tested in a specific order. The first step is usually to constrain the factor loadings (regression coefficients) between factors and their indices to equality across groups. This step tests the hypothesis that the factor structure is the same in the different groups. If these constraints are reasonable, the $\chi 2$ difference test between the restricted model and the baseline model will be nonsignificant for both groups. If the difference between the

restricted and nonrestricted models is significant, we need not throw in the towel immediately; rather results of the LM test can be examined and some equality constraints across the groups can be released. Naturally, the more parameters that differ across groups, the less alike the groups are. Consult Byrne, Shavelson, and Muthén (1989) for a technical discussion of issues concerning partial measurement invariance.

If equality of the factor structure is established, the second step is to ask if the factor variances and covariances are equal. If these constraints are feasible, the third step examines equality of the factor regression coefficients. If all of these constraints are reasonable, the last step is to examine the equality of residual variances across groups, an extremely stringent hypothesis not often tested. If all the regression coefficients, variances, and covariances are the same across groups, it is concluded that the two samples arise from the same population. An example of multiple-group modeling of program evaluation that utilizes a Solomon Four Group design can be found in Ullman, Stein, and Dukes (2000).

A completely different type of multiple-group model is called a *multilevel* model. In this type of modeling analysis, separate models are developed for different levels of a nested hierarchy. For example, researchers might be interested in evaluating an intervention given to several classrooms of students. In these models the dependent variable is measured at the level of the person and predictor variables are included at the individual level and/or at higher levels, say, the classroom. Of particular interest in these models are tests of variability in slopes and intercepts across groups. When there is variability, it is possible to test interesting hypotheses about the moderating effects of level-two variables (say, class size) on level-one relationships (math achievement as a function of gender). An example of a multilevel latent variable model is Stein, Nyamathi, Ullman, and Bentler (2007). Stein et al. examined the effect of marriage (a level-two) variable on risky behaviors (level-one, individual-level behavior) in homeless adults.

Incorporating a mean and covariance structure. Modeling means in addition to variances and covariances requires no modification of the Bentler-Weeks model. Instead a constant, a vector of 1s (labeled V999 in EQS) is included in the model as an independent variable. As a constant, this independent "variable" has no variance and no covariances with other variables in the model. Regressing a variable (either latent or measured) on this constant yields an intercept parameter. The model-reproduced mean of a variable is equal to the sum of the direct and indirect effects for that variable. Therefore if a variable is predicted only from the constant, the intercept is equal to the mean; otherwise, the mean is a function of path coefficients. The inclusion of intercepts allows for tests of latent mean differences across groups and across time. An example of tests of latent means in the context of a Solomon Four Group design evaluating D.A.R.E can be found in Ullman, Stein, and Dukes (2000). Another type of model that incorporates a mean structure is a latent growth curve model. These are outside the scope of this chapter but the interested reader may want to read Biesanz, Deeb-Sossa, Papadakis, Bollen, and Curran (2004), Curran (2000), Curran, Obeidat, and Losardo (2010), Duncan, Duncan, Strycker, Li, and Alpert (1999), Khoo and Muthén, (2000), McArdle and Epstein (1987), and Mehta and West (2000).

A GUIDE TO SOME RECENT LITERATURE

SEM continues to be an ever-expanding field, both in terms of methodology and in terms of applications (Hershberger, 2003). The growth of SEM has mirrored an increase in methodological sophistication of a variety of fields; see Jaffe and Bentler (2009) for the example of drug abuse. SEM research in specific fields can easily be found through search sites such as Bing or Google. Compact overviews of the field are given in books that provide generic insight into SEM concepts and practices, such as those of Byrne (2006), Kline (2010), and Mulaik (2009), and in a very different domain, Grace (2006). Bollen, Bauer, Christ, and Edwards (2010) provide a general overview. Lee (2007), Yuan and Bentler (2007c), and Hayashi, Bentler, and Yuan (2008) provide statistical overviews. General but technical formulations that handle a wide variety of modeling situations can be found in Bartholomew, Knott, and Moustaki (2011) and Skrondal and Rabe-Hesketh (2011). In order to provide a guide to some recent literature on specific topics, and to alert the reader to issues and developments that might become relevant to their own research, this section provides a selective guide to a number of recent methodological publications. We devote a few paragraphs to a half dozen general topics, followed by literature referrals on another dozen topics listed alphabetically.

Conceptions of latent variables. It is clear from our introduction that the vast majority of structural equation models use unmeasured constructs or latent variables. The best conceptual overview of different approaches to defining latent variables is given by Bollen (2002). We cannot review all these viewpoints here, but to give a

flavor, one approach involves the use of true scores of classical test theory. For practical SEM the main conceptual disagreement on latent variables over the last several decades has been in terms of the direction of the arrows in a path diagram and their equation and model testing consequences. On the one hand is the approach emphasized in this review that equates latent variables with common factors. Common factors generate variation in and explain the correlations among the dependent variables that they predict; for example, first-order factors explain correlations of observed variables while second-order factors explain the correlations among first-order factors. If observed variables are not correlated, there can be no factor underlying them. Recently this 100-year-old tradition of measurement (see Cudeck & MacCallum, 2007) has also been called "reflective" measurement. In factor analysis, the arrows in a path diagram go from the latent to the observed variables. When included in a SEM, along with unmeasured residuals (unique or specific factors), the result is a latent variable model because the dimensionality of the space of independent variables is larger than that of the observed variables (Bentler, 1982).

In contrast to this position is the viewpoint that, in many circumstances, the model should be specified differently with the arrows going from the observed to the latent variables; that is, the observed variables are meant to create, and create meaning for, the latent variables: "The indicators determine the latent variable" (Bollen & Lennox, 1991, p. 306). These latent variables presumably are not common or unique factors. They are sometimes called *formative* factors, and their indicator variables, formative or causal indicators. Although formative indicators also have a long history in partial least squares (PLS; see further on), they were introduced into psychology by Bollen and Lennox (1991) and MacCallum and Browne (1993).

The reality is that formative latent variables cannot be identified without somehow requiring the presence of ordinary common factors. In fact, they actually derive their meaning from those factors. Some background and details on this issue, as well as a resolution on how formative factors can be created as ordinary factors and indeed be used in SEM is given in Treiblmaier, Bentler, and Mair (2011). A previous lively discussion is given by Bagozzi (2007), Bollen (2007), and Howell, Breivik, and Wilcox (2007a, 2007b). Additional recent references are Bollen and Davis (2009a, 2009b), Franke, Preacher, and Rigdon (2008), and Hardin, Chang, and Fuller (2011).

Exploratory Factor Analysis

As just noted, latent variables are just the factors of exploratory and confirmatory factor analysis (EFA, CFA). At the preliminary stages of research, EFA is typically an essential methodology to help reduce a variable set, to determine the key dimensions of interest, and to provide evidence on the quality of potential indicators of factors. Preacher and MacCallum (2003) and Costello and Osborne (2005) provide short useful guides on such issues as number of factors and choice of rotation method; a comprehensive overview of both EFA and CFA is given by Mulaik (2010). For technical analysis of alternative factor models such as image factor analysis that seem to have fallen out of favor, see Hayashi and Bentler (2000). A typical controversy revolves around whether principal component analysis can be used to substitute for factor analysis. According to Bentler and Kano (1990) the answer is no, because to do so adequately would require the use of far more measured variables than typically can be accommodated in SEM. That is, factors and components become equivalent only if the number of variables that indicate a given factor becomes large. A new approach to the mathematical relation between factor analysis and component analysis in any given model—not only as the number of indicators gets large—is given in Bentler and de Leeuw (2011), who also provide a new factor analysis estimation methodology.

Extracting too many factors can be a serious problem (Hayashi, Bentler, & Yuan, 2007). If factors are weak in CFA, and surely equally so in EFA, least squares may be a better option than maximum likelihood (ML) (Ximénez, 2009). The usual problems of missing or nonnormal data, or existence of outliers, affects not only SEM but also EFA (Yuan, Marshall, & Bentler, 2002). Recent studies of alternative rotation criteria are given in Sass and Schmitt (2010) and Schmitt and Sass (2011). Inspired by item response theory, Reise, Moore, and Haviland (2010) propose that the bifactor model, a model with one large general factor and several group factors, may provide a more satisfactory structure than existing alternatives in situations where a general dimension makes sense. Jennrich and Bentler (2011) provide a new rotation method for EFA to help find bifactor solutions. EFA is being incorporated into EQS (Bentler, 2008).

Confirmatory factor analysis. As we discussed earlier, CFA is a fundamental component of SEM. Indeed CFA often is the model to use to verify the appropriateness of a measurement model prior to being concerned with regressions among the latent variables. Good basic sources on

CFA are Brown (2006) and Harrington (2009). Some technical issues relating to factor loadings and standard errors are discussed in Yuan, Cheng, and Zhang (2010). Yuan and Chan (2008) discuss how to handle near-singular covariance matrices. A good discussion of reporting practices is given in Jackson, Gillaspy, and Purc-Stephenson (2009).

Confirmatory factor analysis in multiple groups with restrictions across groups is one of the main methods for evaluating measurement invariance, a key issue in assuring that instruments are used fairly and are not biased against certain groups. Bauer (2005a) points out that use of the usual CFA indicators that are generated linearly from factors, when the relation is really nonlinear, can be a serious problem. The equality of factor loadings across groups is a widely known to be a key requirement for invariance, but Wicherts and Dolan (2010) show that equality of intercepts is also critical. The most thorough and modern overview of measurement invariance in a variety of model and data types is given by Millsap (2011).

Exploratory SEM

Although CFA methods integrated into SEM allow some model modification to correct gross misspecifications in the measurement model for the latent variables, developments in SEM across the past four decades have been based on the assumption that the factor structure of the variables is largely known. Indeed, the limitations on number of variables built into current SEM methodology preclude using all the dozens or even hundreds of variables that may exist in a survey or testing context. As a result, some preliminary use of EFA, CFA, and possibly creating composite parcel variables (Little, Cunningham, Shahar, & Widaman, 2002; Yuan, Bentler, & Kano, 1997) to reduce an initial variable set to a manageable one is essential. Thus, by the time a structural model for a selected set of variables and factors becomes relevant to SEM, the measurement model is largely understood. Such a measurement model is usually a cluster-structure model that does not encourage more than one factor to influence any given observed variable.

A different viewpoint is given by Asparouhov and Muthén (2009). Assuming that any large variable set already has been reduced to the key ones for use in SEM, they propose that the measurement model ought to be developed at the same time as the complete SEM. Their exploratory SEM (ESEM) replaces the typical CFA measurement model with an EFA model, but allows latent regressions. The factors that define the latent variables are determined during the EFA with rotations and are subsequently entered into the regressions among factors. As a consequence, the measurement model is rarely a simple cluster structure and correlations between factors and/or latent regression effects are lower. Marsh et al. (2009) illustrate this approach in data on student teaching, suggesting that a simple cluster structure is not an appropriate measurement model and that ESEM allows the full factorial complexity to appear. Similarly, Marsh et al. (2010) apply this approach to the NEO five-factor inventory and report a better fit as compared to CFA with fewer correlated factors; see also Rosellini and Brown (2011).

SEM With Binary, Ordinal, and Continuous Variables

Real datasets are likely to contain a mixture of response formats for variables, including dichotomous (yes/no, etc.) responses, ordinal categorical or polytomous (Likert-type etc.), and continuous variables. At the present time, the polychoric/polyserial methodology (e.g., Lee, Poon, & Bentler, 1995) remains one of the best ways to deal with such varied response formats. In this approach, the categorical responses are viewed as emanating from cuts on an underlying normally distributed continuum with a joint bivariate normality assumed to be underlying the joint categorical contingency table. Even in item response theory, historically a unidimensional but recently also a multidimensional measurement model for categorical variables (e.g., de Ayala, 2009; Reckase, 2009; Wu & Bentler, 2011), there is growing recognition that limited information methods may provide more accurate parameter estimates and model evaluation, and provide more power, except at extremely huge sample sizes (e.g., Joe & Maydeu-Olivares, 2010; Maydeu-Olivares & Joe, 2005). Although modeling all possible response patterns remains an important goal, and spectacular computing improvements to achieve this are being made (see Item Factor Analysis earlier), the approach remains unlikely to be successful in the SEM field where data from only a few hundred subjects may be available but there could be thousands upon thousands of possible response patterns to model. The data are just too sparse.

The important question thus is how well this methodology performs and how to assure that it is appropriately applied. The evidence indicates that polychorics perform very well under a variety of estimation methods (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Yang-Wallentin, Jöreskog, & Luo, 2010), although robust standard errors need to be used. Polychorics have been shown to provide more accurate estimates of a model structure

when compared to Pearson correlations (Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010). Although the assumptions of polychoric correlations are strong, they can be evaluated if desired (Maydeu-Olivares, García-Forero, Gallardo-Pujol, &Renom, 2009). Furthermore, Flora and Curran (2004) find that the methodology was fairly robust to violation of distributional assumptions. See also Bollen and Maydeu-Olivares (2007).

Alternatives also exist, but they are not well studied. Liu (2007) and Liu and Bentler (2009) developed an approach based on a pairwise likelihood that maximizes an objective function based on the product of bivariate probabilities to estimate thresholds as well as polychoric and polyserial correlations simultaneously. The asymptotic distribution of the maximum pairwise likelihood estimators is used to develop a methodology for SEM models. Although it has not been developed for SEM, it is possible that an approach that corrects ordinary correlations based on binary and ordinal data for the coarseness of the response categories (i.e., to minimize the consequences of reducing a continuous variable to one with a few categories) (see Aguinis, Pierce, & Culpepper, 2009) could be developed to produce a useful SEM methodology.

One of the persistent problems with these types of methodologies is that the correlation matrices computed from pairwise information, as is typical with polychorics, may not represent the correlations among real-valued variables; that is, the matrix may have zero or negative eigenvalues or be so badly conditioned that model estimation breaks down. For example, Timmerman and Lorenzo-Seva (2011) report, "The convergence problems of the polychoric approach prevent its general application to empirical data" (p. 218). However, two different approaches were recently developed that can deal with this problem. Bentler and Yuan (2011) developed a way to scale indefinite matrices to assure that the resulting matrix is positive definite. Even better, Yuan, Wu, and Bentler (2011) developed a method for using a ridge correction during estimation with appropriate adjustments to assure that the resulting statistics are correct. The latter two approaches are being incorporated into EQS.

Missing data. Although missing data may be planned as part of design (Graham, Taylor, Olchowski, & Cumsille, 2006), unexpected missing data is inevitable in real data and hence SEM cannot escape dealing with it (Allison, 2003; Enders, 2010; Little & Rubin, 2002; Schafer & Graham, 2002). Peugh and Enders (2004) report that listwise deletion or pairwise present computations are used almost universally. Omitting any subjects that show any missing data can reduce the sample size to the point of instability of estimates and tests, not to speak of bias that often will result. If sample size is not an issue, listwise deletion is acceptable if the missing data mechanism is missing completely at random (MCAR), meaning roughly that the missingness does not depend on either observed or missing data. Tests for MCAR are given by Little (1988) and Kim and Bentler (2002) and further developed for non-normal data by Jamshidian and Jalal (2010) and incorporated into EQS. Although less efficient, pairwise present methods are now also statistically justified (Savalei & Bentler, 2005) and available in EQS.

However, even if MCAR is rejected, data may be missing at random (MAR), meaning roughly that missingness may depend on observed data. Then case-wise or direct ML, computed in EQS via Jamshidian and Bentler (1999), provides an optimal solution for normal data as well as a consistent solution for non-normal data (Yuan & Bentler, 2010a) with robust statistics from Yuan and Bentler (2000a). The Satorra-Bentler (1994) adjusted (mean/variance corrected) statistic performs best under a variety of conditions including small sample sizes and is to be recommended (Savalei, 2010; Yuan & Bentler 2010b). Although the direct ML approach is in principle the best possible, and performs well in practice (Gold, Bentler, & Kim, 2003), the new two-stage ML method (Savalei & Bentler, 2009) is probably better in small samples (see also Cai, 2008; Yuan & Lu, 2008). In the first stage, an unstructured covariance matrix is computed; then the SEM is fit to that matrix using appropriate statistical corrections including for non-normality. An important advantage of this approach is that auxiliary variables can be incorporated into the first stage to reduce bias and variance. Unlike the approach of Graham (2003), they are not used in the SEM of interest that is estimated in the second stage. A technical development of ML under distributional violation with missing data is given by Yuan (2009b). Multiple imputation (MI) is sometimes recommended. It is no doubt a fine method when the data are normal. However, Yuan, Wallentin, and Bentler (2011) compared MI to ML on bias and efficiency of parameter estimates and standard error estimates under non-normality, and found that MI parameter estimates are less efficient and have more biases than those of ML. All of the ML and robust ML methods mentioned here are in EQS. An adapted model-based bootstrap may work well (Savalei & Yuan, 2009).

Research is only beginning on how to handle missing-not-at-random data mechanisms (MNAR). Yuan (2009b) shows how to identify variables that might be responsible for this. Enders (2011) discusses the growth curve context. Kano and Takai (2011) allow the missing-data mechanism

to depend on the latent variables without any need to specify its functional form, and propose a new estimation method based on multi-sample analysis. Surprisingly, complete-case analysis can produce consistent estimators for some important parameters in the model. Song and Lee (2007) developed a Bayesian approach to nonignorable missing data, while Cai, Song, and Lee (2008) provided a more general approach that also handles ordinal as well as continuous data. See also Jamshidian, Yuan, and Le (in press).

Other Important Topics

Case-robust and distribution-robust methods. Statistics in SEM that hold under distributional violations now have a long history (Bentler & Dijkstra, 1985; Browne, 1984; Satorra & Bentler, 1994). Problems of inference in SEM that result from skew and kurtosis are becoming known (e.g., Yuan, Bentler, & Zhang, 2005), and corrections have made it into SEM programs. However, "robust" methods that correct for skew and kurtosis are based on the assumption that the distributions are smooth even if they are not normal. They accept the sample covariance matrix as an appropriate matrix to be modeled. An alternative viewpoint is that outliers or influential cases may make the covariance matrix badly behaved and lead to anomalous estimates (e.g., Bollen, 1987; Yuan & Bentler, 2001). The idea that subjects or cases need to be differentially weighted to better estimate the population covariance of the majority of cases was proposed quite early (Huba & Harlow, 1987). It is still largely ignored, even though Yuan and colleagues have worked out various justified statistical approaches for SEM (Yuan & Bentler, 1998a, 1998b, 2000b; Yuan, Bentler, & Chan, 2004; Yuan, Chan, & Bentler, 2000). Reviews are provided by Yuan and Bentler (2007b) and Yuan and Zhong (2008). Classical data on smoking and cancer, reanalyzed by case-robust SEM, illustrate one approach (Bentler, Satorra, & Yuan, 2009).

Correlation structures. Over the past 100 years, many interesting psychological theories have been phrased in terms of correlation coefficients (standardized covariances) and quantities derived from them, not in terms of covariances. For example, in a typical application, CFA is concerned with the correlational structure of variables, and variances are not really important. Because a statistical theory based on the distribution of correlations was not so easy, the main statistical rationale for SEM over the past 40 years has been based on the asymptotic distribution of covariances. Hence the typical name *covariance structure analysis.* However, the statistical theory now exists for the correct analysis of correlations. See Bentler (2007a) and Bentler and Savalei (2010), or the EQS program.

Diagnostics. The field is still struggling with indices for the evaluation of model adequacy as well as diagnostics for possible problems within an otherwise acceptable model. Overall model test statistics remain important, and attempts to improve them continue (Lin & Bentler, 2010). The relative roles of test statistics versus fit indices is discussed in Yuan (2005), Barrett (2007) with various replies (e.g., Bentler, 2007b; Steiger, 2007), and Saris, Satorra, and van der Veld (2009). Among many studies, Sharma, Mukherjee, Kumar, and Dillon (2005) and Chen, Curran, Bollen, Kirby, and Paxton (2008) provide evidence and caution on the use of standard cutoffs for fit indices. Hancock and Mueller (2011), McDonald (2010), O'Boyle and Williams (2011), and Williams and O'Boyle (2011) discuss the importance of evaluating fit of measurement versus structural models. Various useful model diagnostics are provided by Yuan, Kouros, and Kelley (2008) and Yuan and Hayashi (2011).

Growth curve models. SEM structures that evaluate the means as well as measurement and regression relations have become an important part of structural modeling. A specialized data setup is that of repeated measurement of the same individuals on a given variable across time, where one is interested in a specialized mean structure resulting from the trends across time of individuals on this variable: Some cases may be increasing in level of a trait, others may be staying even, and still others declining. Although the individual trends are of interest, it is the summary statistics such as the mean starting point and the variance around that point, or the mean increment across time and its variance, that can actually be estimated. Luckily, when several sets of such variables are evaluated, along with precursor and consequent variables, quite complicated latent curve models can be motivated and analyzed meaningfully. Short overviews are given by Bentler (2005) and T. Duncan and Duncan (2009), while good texts are Bollen and Curran (2006) and Duncan, Duncan, and Stryker (2006). Interesting modeling issues include: combining model types (Bollen & Curran, 2004), discovering misspecification (Wu & West, 2010), ordinal indicators (Mehta, Neale, & Flay, 2004), power (Hertzog, van Oertzen, Ghisletta, & Linderberger, 2008), multilevel and multiple-group analyses (Hung, 2010), structured models of change (Blozis, 2004), and residual structures (Grimm & Widaman, 2010). A few examples are Bentler, Newcomb, and Zimmerman (2002), Benyamini, Ein-Dor, Ginzburg, and Solomon (2009), Byrne, Lam, and Fielding (2008), and Rudolph, Troop-Gordon, Hessel, and Schmidt (2011).

Interactions and nonlinear effects. In this chapter we emphasize the Bentler-Weeks model. Like all basic SEM approaches, its equations are linear specifications. Unfortunately, chi-square tests in standard SEM may not be able to detect violations of linearity (Mooijaart & Satorra, 2009). Non-normal distributions in variables that are indicators of dependent factors can provide a clue that nonlinear effects may be needed. Methods to allow latent variable interactions and nonlinear relations have been expanding rapidly since Kenny and Judd (1984). Recent examples include Bauer (2005b), Coenders, Batista-Foguet, and Saris (2008), Cudeck, Harring, and du Toit (2009), Klein and Muthén (2007), Lee, Song, and Tang (2007), Marsh, Wen, and Hau (2004), and Wall and Amemiya (2003). Mooijaart and Bentler (2010) developed an approach that includes use of third-order moments. This method seems to be the only one that is insensitive to the standard assumption that the factors involved in nonlinear relations are normally distributed (Mooijaart & Satorra, 2011). Mooijaart and Satorra (in press) show how to optimally select the necessary moments. This method is becoming available in EQS.

Item factor analysis. Conceptually, the factor analysis of responses to individual items that make up larger inventories is just a branch of EFA or CFA depending on the goal and the method. However, factor analysis of individual items usually implies analysis of dichotomous or ordinal responses for which a special set of methodologies has been developing that make modern full information ML methods possible in a reasonable amount of computing time. These developments are described in the review of Wirth and Edwards (2007), and especially in the more recent approaches of An and Bentler (2011a, 2011b), Cai (2010a, 2010b, 2010c), and Edwards (2010). An important application is to the bifactor model in multiple groups (Cai, Yang, & Hansen, 2011) that allows evaluating variable means and variances across groups.

Mediation. Traditional regression emphasizes direct effects of predictor variables on their dependent variables. SEM, of course, has widely expanded the ability to evaluate not only the existence of such effects, but potentially their mechanism of action via intermediary variables. Hence models with mediational paths such as $X \rightarrow Y \rightarrow Z$ have exploded in SEM. The most complete overview is given by MacKinnon (2008). Some other overviews and discussions of theoretical, practical, and technical issues are MacKinnon et al. (2007), Fairchild and MacKinnon (2009), MacKinnon and Fairchild (2009), Zu and Yuan (2010), Preacher and Kelley (2011), Macho and Ledermann (2011), and Wang and Zhang (2011). However, the SEM approach to mediation has come under criticism, e.g., for its reliance on linear equations (Imai, Keele, & Tingley, 2010).

Mixture models. Latent class models for categorical variables have always been an attractive methodology. They remain important with the extension to latent transition analysis that allows for modeling stage-sequential change (Collins & Lanza, 2009). But to SEM researchers, models that attempt to disaggregate a sample with continuous observed variables into subsamples or latent classes that may require different SEM structures is especially interesting. Muthén (1989) was influential in noting that aggregate models may distort paths and effects that may be occurring in subgroups (classes) and proposing the need to disaggregate. This is done with finite mixture SEM. Multiple-group models are specified, one for as many groups as needed, even though the groups are unknown and have to be discovered as part of the analysis. The basic idea is that individuals may come from one or more latent classes and that non-normal distributions may arise from mixing of only a few normal distributions.

Illustrative early papers are Yung (1997) and Muthén and Shedden (1999). Developments and uses have grown rapidly in the last decade. The most important subsets of models are factor mixture models (e.g., Lubke & Muthén, 2005; Lubke & Neale, 2008) and growth mixture models (e.g., Grimm & Ram, 2009; Wu, Witkiewitz, McMahon, & Dodge, 2010). An overview of approaches and applications is given in Hancock and Samuelsen (2008) and Lubke (2010). Overextraction of number of classes, local minima, and other technical problems occur (Bauer, 2007; Bauer & Curran, 2003; Hipp & Bauer, 2006; Nylund, Asparouhov, & Muthén, 2007; Tueller, Drotar, & Lubke, 2011; Tueller & Lubke, 2010). New approaches include multilevel mixture regression (Muthén & Asparouhov, 2009a), fitting multiple conventional SEM models (Yuan & Bentler, 2010c), and allowing covariates for mixed binary and continuous responses (An & Bentler, 2011c; Muthén & Asparouhov, 2009b).

Model comparison. It is often important to compare the fit of two nested models. This process occurs even when only one model is fit, since fit indices like the comparative fit index explicitly compare the current model to the model of uncorrelated variables. As noted by Widaman and Thompson (2003), for resulting fit indices to be meaningful, the models have to be nested, that is, one model must be obtainable from the other by adding restrictions. It is not always obvious whether two models are nested. Bentler and Satorra (2010) provide a simple NET (nesting and equivalence) procedure to evaluate nesting. Their

method also can answer the important question of whether two models, for example, $X \rightarrow Y \rightarrow Z$ and $X \leftarrow Y, \leftarrow Z$, might be equivalent (they are). Equivalent models fit identically and cannot be distinguished statistically, but their interpretations may be quite different.

Most nested model comparisons are done with chi-square difference tests. Yuan and Bentler (2004a) discuss the performance of this test when the base model is misspecified. When robust statistics such as the Satorra-Bentler (1994) SB-scaled test are used, this requires a somewhat involved hand computation (Satorra & Bentler, 2001) and can result in a negative SB chi-square value. Satorra and Bentler (2010) show how to modify the SB difference computations to avoid negative chi-squares; EQS is automating these computations. Bryant and Satorra (in press) show that different programs compute slightly different robust quantities, and hence require tweaks to the basic methodology. MacCallum, Browne, and Cai (2006) propose a method based on RMSEAs to compare and compute power for evaluating small differences between two models. Li and Bentler (2011a) show that their procedure can be improved by a judicious use of a single RMSEA value for these model differences.

Model misspecification. Model evaluation remains a critical issue, both statistically and with fit indices as noted above. From both points of view, everything ever written about model evaluation could be cited here, but we emphasize only a few additional recent publications. Yuan, Marshall, and Bentler (2003) trace the effects of misspecification on parameter estimates. Saris et al., (2009) propose using modification indices (LM test) and expected parameter change to judge misspecifications. Kim (2005) relates fit indices and power, MacCallum, Lee, and Browne (2010) discuss the role of isopower in power analysis, and von Oertzen (2010) discusses how study design can be improved without changing power. Culpepper and Aguinis (2011) discuss analysis of covariance with covariates measured with error, and Yuan and Bentler (2006a) discuss power in latent versus manifest mean structure models. The bootstrap is also useful (Yuan & Hayashi, 2006; Yuan, Hayashi, & Yanagihara, 2007; Savalei & Yuan, 2009).

The noncentral chi-square distribution and the associated noncentrality parameter provide key information for power analysis, confidence intervals in RMSEA, and so on. Curran, Bollen, Paxton, Kirby, and Chen (2002) find empirically that this is acceptable with small misspecifications; Olsson, Foss, and Breivik (2004) agree, but also find that the noncentral chi-square distribution may be inappropriate. Yuan (2008) and Yuan, Hayashi, and Bentler (2007) propose that when model errors are more than minor, the normal distribution may be more appropriate as a reference distribution. See also Shapiro (2009). Chun and Shapiro (2009) marshal simulation evidence to disagree. Raykov (2005) suggests use of a bias-corrected estimator of noncentrality, while Herzog and Boomsma (2009) propose that a correction due to Swain can improve estimation in small samples.

Multilevel models. Data often has a hierarchical, or multilevel, structure so that standard assumptions of independence of subjects break down. For example, students are nested within schools, and many schools exist. In such cases, hierarchical linear regression models (HLM) and multilevel latent variable models (MLM) analyze variation into within Level-1 units (student) and between Level-2 units (schools). Additional levels may also exist (repeated measures within individuals; schools in districts). Some HLM models can be estimated as standard SEM models (Bauer, 2003). Recent overviews include de Leeuw and Meijer (2008), Hox (2010), Hox and Roberts (2011), and Snijders and Bosker (2011). The HLM approach is not SEM, so we concentrate on recent advances in MLM developments. In MLM, as in multiple groups, two or more model matrices are required and the models for these may be identical, similar, or completely unrelated depending on theory.

As in ordinary SEM, there are some lucky circumstances—hard to count on in practice—where MLM statistics are robust to violation of normality assumptions as sample size gets very large (Yuan & Bentler, 2005, 2006b). It is usually sample size at the highest level that is critical to acceptable performance of MLM statistics. Normal theory maximum likelihood is now the default estimation method (Bentler et al., 2011; Liang & Bentler, 2004). More generally, robust statistics have to be used with non-normal distributions. These are provided by Yuan and Bentler (2002a, 2003, 2004b). Bentler and Liang (2008) related MLM and linear mixed effect models, permitting SEM statistics to become relevant to the latter. Yuan and Bentler (2007c) propose fitting multiple single-level models, making the models similar to standard SEM. Culpepper (2009) discusses a multilevel approach to profile analysis for binary data. Grilli and Rampichini (2007) discuss multilevel models for ordinal variables. Rabe-Hesketh, Skrondal, and Pickles (2004) present a generalized linear latent and mixed model framework with a response model and a structural model for the latent variables that allows continuous, ordered categorical, and count responses and a wide range of latent variable structures.

Partial least squares. As in the case of formative measurement, PLS allows estimation of latent factors from observed variables; that is, proxys for the true latent variables are used in the model. However, a consequence is that as of now, "in general, not all parameters will be estimated consistently" (Dijkstra, 2010, p. 37). This means that while the PLS procedure can always be implemented, and may perform quite well in practice (e.g., Reinartz, Haenlein, & Henseler, 2009), even today the properties of the solution remain unknown. An alternative approach was developed by Skrondal and Laake (2001) based on factor score estimates, but it is limited to three groups of factors and allows no higher-order factors. Hoshino and Bentler (2011) developed an extension to Skrondal and Laake's methodology.

Reliability. Internal consistency reliability is estimated almost universally by coefficient alpha. This is not always the best idea, because SEM-based methods lead to superior estimates (Sijtsma, 2009). A review of old and new coefficients from the SEM viewpoint is given by Bentler (2009). The greatest lower bound (GLB) to reliability does not assume a specific SEM model, simply a factor model with an unspecified number of factors. Li and Bentler (2011b) propose a bias reduction method that improves estimation of the GLB. For any given SEM model, the coefficient defined by Bentler (2007c) yields the maximal reliability of a unit-weighted sum. Both of these are computed in EQS. Revelle and Zinbarg (2009) recommend a coefficient based on an SEM model with a general factor. To measure unidimensionality, ten Berge and Sočan (2004) propose use of the proportion of common variance due to a single factor. Raykov has worked extensively on reliability. Illustrative articles on this topic are on reliability for multilevel models (Raykov & Penev, 2010) and for binary measures (Raykov, Dimitrov, & Asparouhov, 2010), on the relation between maximal reliability and maximal validity (Penev & Raykov, 2006), and how to compute generalizability coefficients using SEM (Raykov & Marcoulides, 2006). Statistical issues related to some reliability coefficients are included in Maydeu-Olivares, Coffman, and Hartmann (2007), Maydeu-Olivares, Coffman, García-Forero, and Gallardo-Pujol (2010), Shapiro and ten Berge (2000), and Yuan and Bentler (2002b).

Simulation. Chun and Shapiro (2010) develop a new numerical procedure that can construct covariance matrices with the property that, for a given SEM and a discrepancy function, the corresponding minimizer of the discrepancy function has a specified value. Their method achieves a wider range of covariance matrices than the method of Cudeck and Browne (1992). Headrick (2010) develops power method polynomials and other transformations to non-normality of variables while maintaining correlation structures. Mair, Satorra, and Bentler (2011) describe a procedure based on multivariate copulas for simulating multivariate non-normal data that satisfies a specified covariance matrix that can be based on a general SEM model. This method provides a new way to generate data for Monte Carlo studies. Mair, Wu, and Bentler (2010) provide an interface between the statistical package R and EQS.

REFERENCES

Aguinis, H., Pierce, C. A., & Culpepper, S. A. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods, 12*(4), 623–652.

Allison, P. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545–557.

An, X., & Bentler, P. M. (2011a). *Nesting Monte Carlo EM for high dimensional item factor analysis*. Under editorial review.

An, X., & Bentler, P. M. (2011b). Efficient direct sampling MCEM algorithm for latent variable models with binary responses. *Computational Statistics and Data Analysis*.

An, X., & Bentler, P. M. (2011c). Extended mixture factor analysis model with covariates for mixed binary and continuous responses. *Statistics in Medicine.* doi: 10.1002/sim.4310

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*(3), 397–438.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*(2), 229–237.

Barrett, P, (2007). Structural equation modeling: Adjudging model fit. *Personality and Individual Differences, 42,* 815–824.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). Chichester, UK: Wiley.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28*(2), 135–167.

Bauer, D. J. (2005a). The role of nonlinear factor-to-indicator relationships in tests of measurement equivalence. *Psychological Methods, 10*(3), 305–316.

Bauer, D. J. (2005b). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling, 4,* 513–535.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*(4), 757–786.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Benyamini, Y. Ein-Dor, T., Ginzburg, K., & Solomon, Z. (2009). Trajectories of self-rated health among veterans: A latent growth curve analysis of the impact of posttraumatic symptoms. *Psychosomatic Medicine, 71*(3), 345–352.

Bentler, P. M. (1982). Linear systems with multiple levels and types of latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 101–130). Amsterdam, The Netherlands: North-Holland.

Bentler, P. M. (1990). Fit indexes, Lagrange multipliers, constraint changes and incomplete data in structural models. *Multivariate Behavioral Research, 25*(2), 163–172.

Bentler, P. M. (2005). Latent growth curves. In J. Werner (Ed.), *Zeitreihenanalysen* (pp. 13–36). Berlin, Germany: Logos.

Bentler, P. M. (2007a). Can scientifically useful hypotheses be tested with correlations? *American Psychologist, 62,* 772–782.

Bentler, P. M. (2007b). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42,* 825–829.

Bentler, P. M. (2007c). Covariance structure models for maximal reliability of unit-weighted composites. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 1–19). Amsterdam, The Netherlands: North-Holland.

Bentler, P. M. (2008). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software. www.mvsoft.com

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74,* 137–143.

Bentler, P. M., & de Leeuw, J. (2011). Factor analysis via components analysis. *Psychometrika, 76*(3), 461–470. doi: 10.1007/S11336-011-9217-5

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9–42). Amsterdam, The Netherlands: North-Holland.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25,* 67–74.

Bentler, P. M., & Liang, J. (2008). A unified approach to two-level structural equation models and linear mixed effects models. In D. Dunson (Ed.), *Random effects and latent variable model selection* (pp. 95–119). New York, NY: Springer.

Bentler, P. M., Liang, J., Tang, M.-L., & Yuan, K.-H. (2011). Constrained maximum likelihood estimation for two-level mean and covariance structure models. *Educational and Psychological Measurement, 71*(2), 325–345.

Bentler, P. M., Newcomb, M. D., & Zimmerman, M. A. (2002). Cigarette use and drug use progression: Growth trajectory and lagged effect hypotheses. In D. B. Kandel (Ed.), *Examining the gateway hypothesis: Stages and pathways of drug involvement* (pp. 223–253). New York, NY: Cambridge University Press.

Bentler, P. M., & Raykov, T. (2000). On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology, 85,* 125–131.

Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods, 15,* 111–123.

Bentler, P. M., Satorra, A., & Yuan, K.-H. (2009). Smoking and cancers: Case-robust analysis of a classic data set. *Structural Equation Modeling, 16,* 382–390.

Bentler, P. M., & Savalei, V. (2010). Analysis of correlation structures: Current status and open problems. In S. Kolenikov, D. Steinley, & L. Thombs (Eds.), *Statistics in the social sciences: Current methodological developments* (pp. 1–36). Hoboken NJ: Wiley.

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research, 34*(2), 181–197.

Bentler, P. M., & Yuan, K.-H. (2011). Positive definiteness via off-diagonal scaling of a symmetric indefinite matrix. *Psychometrika, 76*(1), 119–123.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equation with latent variables. *Psychometrika, 45,* 289–308.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*(1), 30–52. doi:10.1037/1082–989X.9.1.30

Blozis, S. A. (2004). Structured latent curve models for the study of change in multivariate repeated measures. *Psychological Methods, 9*(3), 334–353.

Bollen, K. A. (1987). Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods & Research, 15,* 375–84.

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

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*(1), 605–634.

Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*(2), 219–228.

Bollen, K. A., Bauer, D. J., Christ, S. L., & Edwards, M. C. (2010). An overview of structural equations models and recent extensions. In S. Kolenikov, D. Steinley, & L. Thombs (Eds.), *Recent developments in social science statistics* (pp. 37–80). Hoboken, NJ: Wiley.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research, 32,* 336–383.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* Hoboken, NJ: Wiley.

Bollen, K. A., & Davis, W. R. (2009a). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling, 16*(3), 498–522.

Bollen, K. A., & Davis, W. R. (2009b). Two rules of identification for structural equation modeling. *Structural Equation Modeling, 16*(3), 523–536.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305–314.

Bollen, K. A., & Maydeu-Oliveres, A. (2007). A polychoric instrumental variable (PIV) estimator for structural equation models with categorical variables. *Psychometrika, 72,* 309–326.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

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

Browne, M. W., & Cudeck R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural models*. Newbury Park, CA: Sage.

Bryant, F. B., & Satorra, A. (in press). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling*.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Erlbaum.

Byrne, B. M., Lam, W. W. T., & Fielding, R. (2008). Measuring patterns of change in personality assessments: An annotated application of latent growth curve modeling. *Journal of Personality Assessment, 90*(6), 536–546.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Cai, J.-H., Song, X.-Y., & Lee, S.-Y. (2008). Bayesian analysis of nonlinear structural equation models with mixed continuous, ordered and unordered categorical, and nonignorable missing data, *Statistics and its Interface, 1,* 99–114.

Cai, L. (2008). SEM of another flavor: Two new application of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology, 61*(2), 309–329.

Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings-Robbins-Monro algorithm. *Psychometrika, 75*(1), 33–57.

Cai, L. (2010b). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics, 35*(3), 307–335.

Cai, L. (2010c). A two-tier full-information item factor analysis model with applications. *Psychometrika, 75*(4), 581–612.

Cai, L., Yang, L. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*. doi: 10.1037/a0023350

Chen, F., Curran P. J., Bollen K. A., Kirby J. B., & Paxton, P. M. (2008). An empirical evaluation of the use of fixed cutoff points in rmsea test statistics in structural equation models. *Sociological Methods & Research, 36,* 462–494.

Chun, S. Y., & Shapiro, A. (2009). Normal versus noncentral chi-square asymptotics of misspecified models. *Multivariate Behavioral Research, 44,* 803–827.

Chun, S. Y., & Shapiro, A. (2010). Construction of covariance matrices with a specified discrepancy function minimizer, with application to factor analysis. *SIAM Journal of Matrix Analysis and Applications, 31,* 1570–1583.

Coenders, G. Batista-Foguet, J. M., & Saris, W. E. (2008). Simple, efficient and distribution-free approach to interaction effects in complex structural equation models. *Quality & Quantity, 42*(3), 369–396.

Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis*. Hoboken, NJ: Wiley.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1–9.

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika, 57,* 357–369.

Cudeck, R., Harring, J. R., & du Toit, S. H. C. (2009). Marginal maximum likelihood estimation of a latent variable model with interaction. *Journal of Educational and Behavioral Statistics, 34*(1), 131–144.

Cudeck, R., & MacCallum, R. C. (Eds.). (2007). *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Erlbaum.

Culpepper, S. A. (2009). A multilevel nonlinear profile analysis model for dichotomous data. *Multivariate Behavioral Research, 44,* 646–667.

Culpepper, S. A., & Aguinis, H. (2011). Using analysis of covariance (ancova) with fallible covariates. *Psychological Methods*. doi: 10.1037/a0023355

Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J. R. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 1–42). Mahwah, NJ: Erlbaum.

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research, 37,* 1–36.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121–136. doi:10.1080/15248371003699969

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

de Leeuw, J., & Meijer, E. (Eds.). (2008). *Handbook of multilevel modeling*. New York, NY: Springer.

Dijkstra, T. K. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang, (Eds.), *Handbook of partial least squares: Concepts, methods, and applications, computational statistics, vol. II* (pp. 23–46). Heidelberg, Germany: Springer.

Dukes, R. L., Stein, J. A., & Ullman, J. B. (1997). Long-term impact of drug abuse resistance education (D.A.R.E.): Results of a six-year follow-up. *Evaluation Review, 21,* 483–500.

Duncan, T. E., & Duncan, S. C. (2009). The abc's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Compass, 3*(6), 979–991.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling* (2nd ed.). Mahwah, NJ: Erlbaum.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.

Edwards, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika, 75*(3), 474–497.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*(1), 1–16.

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science, 10*(2), 87–99.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling, 7,* 356–410.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*(4), 625–641.

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research, 61*(12), 1229–1237.

Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete nonnormal data. *Structural Equation Modeling, 10,* 47–79.

Grace, J. B. (2006). *Structural equation modeling and natural systems.* New York, NY: Cambridge University Press.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10,* 80–100.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11,* 323–343.

Grilli, L., & Rampichini, C. (2007). Multilevel factor models for ordinal variables. *Structural Equation Modeling, 14*(1), 1–25.

Grimm, K. J., & Ram, N. (2009). A second-order growth mixture model for developmental research. *Research in Human Development, 6*(2–3), 121–143.

Grimm, K. J., & Widaman, K. F. (2010). Residual structures in latent growth curve modeling. *Structural Equation Modeling, 17*(3), 424–442.

Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement, 71*(2), 306–324.

Hancock, G. R., & Samuelsen, K. M. (2008). *Advances in latent variable mixture models*. Charlotte, NC: Information Age.

Hardin, A. M., Chang, J. C., & Fuller, M. A. (2011). Formative measurement and academic research: In search of measurement theory. *Educational and Psychological Measurement, 71*(2), 270–284.

Harrington, D. (2009). *Confirmatory factor analysis*. New York, NY: Oxford University Press.

Hayashi, K., & Bentler, P. M. (2000). On the relations among regular, equal unique variances, and image factor analysis models. *Psychometrika, 65,* 59–72.

Hayashi, K., Bentler, P. M., & Yuan, K.-H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling, 14,* 505–526.

Hayashi, K., Bentler, P. M., & Yuan, K.-H. (2008). Structural equation modeling. In C. R. Rao, J. Miller, & D. C. Rao (Eds.), *Handbook of statistics 27 Epidemiology & medical statistics* (pp. 395–428). Amsterdam, The Netherlands: North-Holland.

Headrick, T. C. (2010). *Statistical simulation: Power method polynomials and other transformations*. New York, NY: Chapman & Hall/CRC.

Hershberger, S. (2003). The growth of structural equation modeling. *Structural Equation Modeling, 10* (1), 35–46.

Hertzog, C., van Oertzen, T., Ghisletta, P., & Linderberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling, 15* (4), 541–563.

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling, 16,* 1–27.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11* (1), 36–53 (erratum: p. 305).

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity, 44* (1), 153–166.

Hoogland, J. J. (1999). *The robustness of estimation methods for covariance structure analysis.* Unpublished PhD dissertation, Rijksuniversiteit Groningen.

Hoshino, T., & Bentler, P. M. (2011). *Bias in factor score regression and a simple solution.* UCLA Statistics Preprint #621. http://preprints.stat.ucla.edu

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Reconsidering formative measurement. *Psychological Methods, 12* (2), 205–218.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007b). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods, 12* (2), 238–245.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Hox, J. J., & Roberts, J. K. (Eds.). (2011). *Handbook of advanced multilevel analysis.* New York, NY: Routledge.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structural equation modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

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

Hu, L.-T., Bentler, P. M., & Kano Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351–362.

Huba, G. J., & Harlow, L. L. (1987). Robust structural equation models: Implications for developmental psychology. *Child Development, 58,* 147–166.

Hung, L.-F. (2010). The multigroup multilevel categorical latent growth curve models. *Multivariate Behavioral Research, 45* (2), 359–392.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15* (4), 309–334.

Jackson, D., Gillaspy, J., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14* (1), 6–23.

Jaffe, A., & Bentler, P. M. (2009). Structural equation modeling and drug abuse etiology: A historical perspective. In L. M. Scheier (Ed.), *Handbook of drug use etiology: Theory, methods and empirical findings* (pp. 547–562). Washington DC: American Psychological Association.

Jamshidian, M., & Bentler, P. M. (1999). Using complete data routines for ML estimation of mean and covariance structures with missing data. *Journal of Educational and Behavioral Statistics, 23,* 21–41.

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika, 75* (4), 1–26.

Jamshidian, M., Yuan, K., & Le, P. (in press). Using confidence intervals to test for non-ignorable non-response.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*. doi: 10.1007/S11336-011-9218-4

Joe, H., & Maydeu-Olivares, A. (2010). A general family of limited information goodness-of-fit statistics for multinomial data. *Psychometrika, 75* (3), 393–419.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 57,* 409–442.

Kano, Y., & Takai, K. (2011). Analysis of NMAR missing data without specifying missing-data mechanisms in a linear latent variate model. *Journal of Multivariate Analysis, 102,* 1241–1255.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201–210.

Khoo, S.-T., & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. In J. S. Rose, L. Chassin, C. C. Preston, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 43–78). Mahwah, NJ: Erlbaum.

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling, 12* (3), 368–390.

Kim, K. H., & Bentler, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika, 67,* 609–624.

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42* (4), 647–673.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Lee, S.-Y. (Ed.) (2007). *Handbook of latent variable and related models*. Amsterdam, The Netherlands: North-Holland.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology, 48,* 339–358.

Lee, S.-Y., Song, X.-Y, & Tang, N.-S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling, 14* (3), 404–434.

Li, L., & Bentler, P. M. (2011a). Quantified choice of root-mean-square errors of approximation for evaluation and power analysis of small differences between structural equation models. *Psychological Methods, 16* (2), 116–126.

Li, L., & Bentler, P. M. (2011b). *The greatest lower bound to reliability: Corrected and resampling estimators*. Under editorial review.

Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika, 69,* 101–122.

Lin, J., & Bentler, P. M. (2010). *A new goodness of fit test statistic for covariance structure analysis and its comparative performance*. Under review.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9* (2), 151–173.

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198–1202.

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data*. New York, NY: Wiley.

Liu, J. (2007). *Multivariate ordinal data analysis with pairwise likelihood and its extension to SEM.* Unpublished dissertation, UCLA.

Liu, J., & Bentler, P. M. (2009). *Latent variable approach to high dimensional ordinal data using composite likelihood.* Technical Report, Department of Psychology, UCLA.

Lubke, G. H. (2010). Latent variable mixture modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences.* New York, NY: Routledge.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10* (1), 21–39.

Lubke, G. H., & Neale, M. C. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research, 43,* 592–620

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

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114* (3), 533–541.

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 11,* 19–35.

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

MacCallum, R. C., Lee, T., & Browne, M. W. (2010). The issue of isopower in power analysis for tests of structural equation models. *Structural Equation Modeling, 17,* 23–41.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99.

Macho, S., & Ledermann, T. (2011). Estimating, testing, and comparing specific effects in structural equation modles: The phantom model approach. *Psychological Methods, 16* (1), 34–43.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis.

MacKinnon, D. P., & Fairchild, A. J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science, 18* (1), 16–20.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593–614.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7* (1), 83–104. doi:10.1037/1082-989X.7.1.83

Mair, P., Satorra, A., & Bentler, P. M. (2011). *Nonnormal multivariate data from copulas: Applications to SEM*. Under editorial review.

Mair, P., Wu, E., & Bentler, P. M. (2010). EQS goes R: Simulations for SEM using the package REQS. *Structural Equation Modeling, 17:2,* 333–349.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16* (3), 439–476.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22* (3), 471–491.

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9,* 275–300.

Maydeu-Olivares, A., Coffman, D. L., García-Forero, C., & Gallardo-Pujol, D. (2010). Hypothesis testing for coefficient alpha: An SEM approach. *Behavior Research Methods, 42* (2), 618–625.

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods, 12* (2), 157–176 (errata: p. 433).

Maydeu-Olivares, A., García-Forero, C., Gallardo-Pujol, D., & Renom, J. (2009). Testing categorized bivariate normality with two-stage polychoric correlation estimates. *Methodology, 5* (4), 131–136.

Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association, 100,* 1009–1020.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58,* 110–133.

McDonald, R. P. (2010). Structural models and the art of approximation. *Perspectives on Psychological Science, 5* (6), 675–686.

Mehta, P. D., Neale, M. C., & Flay, B. R. (2004). Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes. *Psychological Methods, 9* (3), 301–333.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5,* 23–43.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Mooijaart, A., & Bentler, P. M. (2010). An alternative approach for nonlinear latent variable models. *Structural Equation Modeling, 17* (3), 357–373.

Mooijaart, A., & Satorra, A. (2009). On insensitivity of the chi-square model test to non-linear misspecification in structural equation models. *Psychometrika, 74,* 443–455.

Mooijaart, A., & Satorra, A. (2011). *MM versus ML estimates of structural equation models with interaction terms: Robustness to nonnormality of the consistency property*. Preprint #620, UCLA Department of Statistics.

Mooijaart, A., & Satorra, A. (in press). Moment testing for interaction terms in structural equation modeling. *Psychometrika*.

Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. Boca Raton, FL: Chapman & Hall/CRC.

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54* (4), 557–585.

Muthén, B., & Asparouhov, T. (2009a). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society, Series A, 172,* 639–657.

Muthén, B., & Asparouhov, T. (2009b). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431–462.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55* (2), 463–469.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14* (4), 535–569.

O'Boyle, E. H. Jr., & Williams, L. J. (2011). Decomposing model fit: Measurement vs theory inorganizational research using latent variables. *Journal of Applied Psychology, 96* (1), 1–12.

Olsson, U. H., Foss, T., & Breivik, E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a non-central chi-square distribution under model misspecification? *Sociological Methods &* Research, *32*(4), 453–500.

Penev, S., & Raykov, T. (2006). On the relationship between maximal reliability and maximal validity of linear composites. *Multivariate Behavioral Research, 41*(2), 105–126.

Peugh, J., & Enders, C. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525–556.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93–115.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*(1), 13–43.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*(2), 167–190.

Raykov, T. (2005). Bias-corrected estimation of noncentrality parameters of covariance structure models. *Structural Equation Modeling, 12,* 120–129.

Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling, 17*(2), 265–279.

Raykov, T., & Marcoulides, G. A. (2006). Estimation of generalizability coefficients via a structural equation modeling approach to scale reliability evaluation. *International Journal of Testing, 6*(1), 81–95.

Raykov, T., & Penev, S. (2010). Evaluation of reliability coefficients for two-level models via latent variable analysis. *Structural Equation Modeling, 17*(4), 629–641.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.

Reinartz, W. J., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Market Research, 26,* 332–344.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor model and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145–154.

Rosellini, A. J., & Brown, T. A. (2011). The NEO five-factor inventory: Latent structure and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment, 18*(1), 27–38.

Rudolph, K. D., Troop-Gordon, W., Hessel, E. T., & Schmidt, J. D. (2011). A latent growth curve analysis of early and increasing peer victimization as predictors of mental health across elementary school. *Journal of Clinical Child and Adolescent Psychology, 40*(1), 111–122.

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*(1), 73–103.

Saris, W. E., Satorra, A., & van der Veld, W. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16*(4), 1–24.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology, 22,* 249–278.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *Proceedings of the American Statistical Association* (pp. 308–313). Alexandria, VA: American Statistical Association.

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., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75,* 243–248.

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

Savalei, V. (2010). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling, 17*(2), 241–264.

Savalei, V., & Bentler, P. M. (2005). A statistically justified pairwise ML method for incomplete nonnormal data: A comparison with direct ML and pairwise ADF. *Structural Equation Modeling, 12,* 183–214.

Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling, 16,* 477–497.

Savalei, V., & Yuan, K.-H. (2009). On the model-based bootstrap with missing data: Obtaining a p-value for a test of exact fit. *Multivariate Behavioral Research, 44,* 741–763.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement, 71*(1), 95–113.

Shapiro, A. (2009). Asymptotic normality of test statistics under alternative hypotheses. *Journal of Multivariate Analysis, 100,* 936–945.

Shapiro, A., & Browne, M. W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association, 82,* 1092–1097.

Shapiro, A., & ten Berge, J. M. F. (2000). The asymptotic bias of minimum trace factor analysis, with applications to the greatest lower bound to reliability. *Psychometrika, 65*(3), 413–425.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*(1), 935–943.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120.

Skrondal, A, & Laake, P. (2001) Regression among factor scores. *Psychometrika, 66*(4), 563–575. doi: 10.1007/BF02296196

Skrondal, A., & Rabe-Hesketh, S. (2011). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. New York, NY: Chapman & Hall/CRC.

Snijders, T., & Bosker, R. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229–239.

Song, X.-Y., Lee, S.-Y. (2007). Bayesian analysis of latent variable models with nonignorable missing outcomes from exponential family. *Statistics in Medicine, 26,* 681–693.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7,* 149–162.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893–898.

Stein, J. A., Nyamathi, A., Ullman, J. B., & Bentler, P. M. (2007). Impact of marriage on HIV/AIDS risk behaviors among impoverished, at-risk couples: A multilevel latent variable approach. *AIDS and Behavior, 11*(1), 87–98. doi:10.1007/s10461-005-9058-2

ten Berge, J. M. F., & Soèan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69*(4), 613–625.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209–220.

Treiblmaier, H., Bentler, P. M., & Mair, P. (2011). Formative constructs implemented via common factors. *Structural Equation Modeling, 18,* 1–17.

Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling, 18*(1), 110–131.

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling, 17*(2), 165–192.

Ullman, J. B. (2007). *Structural equation modeling.* In B. G. Tabachnick & L. S. Fidell, *Using multivariate statistics* (5th ed., pp. 676–780). Boston, MA: Allyn & Bacon.

Ullman, J. B., Stein, J. A., & Dukes, R. L. (2000). Evaluation of D.A.R.E. (drug abuse resistance education) with latent variables in the context of a Solomon four group design (pp. 203–232). In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions*. Mahwah, NJ: Erlbaum.

Velicer, W. F., & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3,* 321–327.

von Oertzen, T. (2010). Power equivalence in structural equation modelling. *British Journal of Mathematical and Statistical Psychology, 63*(2), 257–272.

Wall, M. M., & Amemiya, Y. (2003). A method of moment technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology, 56,* 47–63.

Wang, L., & Zhang, Z. (2011). Estimating and testing mediation effects with censored data. *Structural Equation Modeling, 18*(1), 18–34.

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice, 29*(3), 39–47.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8,* 16–37.

Williams, L. J., & O'Boyle, E. H. Jr. (2011). The myth of global fit indices and alternatives for assessing latent variable relations. *Organizational Research Methods, 14*(2), 350–369.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58–79.

Wu, E. J. C., & Bentler, P. M. (2011). *EQSIRT—A user-friendly IRT program*. Encino, CA: Multivariate.

Wu, J., Witkiewitz, K., McMahon, R. J., & Dodge, K. A. (2010). A parallel process growth mixture model of conduct problems and substance use with risky sexual behavior. *Drug and Alcohol Dependence, 111*(3), 207–214.

Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research, 45,* 420–452.

Ximénez, C. (2009). Recovery of weak factor loadings in confirmatory factor analysis under conditions of misspecification. *Behavior Research Methods, 41*(4), 1038–52.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling, 17*(3), 392–426.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research, 40*(1), 115–148.

Yuan, K.-H. (2008). Noncentral chi-square versus normal distributions in describing the likelihood ratio statistic: The univariate case and its multivariate implication. *Multivariate Behavioral Research, 43,* 109–136.

Yuan, K.-H. (2009b). Normal distribution based pseudo ML for missing data: With applications to mean and covariance structure analysis. *Journal of Multivariate Analysis, 100,* 1900–1918.

Yuan, K.-H., & Bentler, P. M. (1998a). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 51,* 63–88.

Yuan, K.-H., & Bentler, P. M. (1998b). Structural equation modeling with robust covariances. In A. Raftery (Ed.), *Sociological methodology* (pp. 363–396). Malden, MA: Blackwell.

Yuan, K.-H., & Bentler, P. M. (1999). F tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics, 24,* 225–243.

Yuan, K.-H., & Bentler, P. M. (2000a). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology 2000* (pp. 165–200). Washington, DC: American Sociological Association.

Yuan, K.-H., & Bentler, P. M. (2000b). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika, 65,* 43–58.

Yuan, K.-H., & Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 54,* 161–175.

Yuan, K.-H., & Bentler, P. M. (2002a). On normal theory based inference for multilevel models with distributional violations. *Psychometrika, 67,* 539–562.

Yuan, K.-H., & Bentler, P. M. (2002b). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67,* 251–259.

Yuan, K.-H., & Bentler, P. M. (2003). Eight test statistics for multilevel structural equation models. *Computational Statistics & Data Analysis, 44,* 89–107.

Yuan, K.-H., & Bentler, P. M. (2004a). On chi-square difference and *z* tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement, 64,* 737–757.

Yuan, K.-H., & Bentler, P. M. (2004b). On the asymptotic distributions of two statistics for two-level covariance structure models within the class of elliptical distributions. *Psychometrika, 69,* 437–457.

Yuan, K.-H., & Bentler, P. M. (2005). Asymptotic robustness of the normal theory likelihood ratio statistic for two-level covariance structure models. *Journal of Multivariate Analysis, 94,* 328–343.

Yuan, K.-H., & Bentler, P. M. (2006a). Mean comparison: Manifest variable versus latent variable. *Psychometrika, 71,* 139–159.

Yuan, K.-H., & Bentler, P. M. (2006b). Asymptotic robustness of standard errors in multilevel structural equation models. *Journal of Multivariate Analysis, 97,* 1121–1141.

Yuan, K. H., & Bentler, P. M. (2007b). Robust procedures in structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 367–397). Amsterdam, The Netherlands: North-Holland.

Yuan, K.-H., & Bentler, P. M. (2007c). Multilevel covariance structure analysis by fitting multiple single-level models. In Y. Xie (Ed.), *Sociological methodology 2007* (Vol. 37, pp. 53–82). New York, NY: Blackwell.

Yuan, K.-H., & Bentler, P. M. (2010a). Consistency of normal-distribution-based pseudo-maximum likelihood estimates when data are missing at random. *American Statistician, 64*(3), 263–267.

Yuan, K.-H., & Bentler, P. M. (2010b). Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical and Statistical Psychology, 63,* 273–291.

Yuan, K.-H., & Bentler, P. M. (2010c). Finite normal mixture SEM analysis by fitting multiple conventional SEM models. In T. F. Liao (Ed.), *Sociological Methodology 2010* (pp. 191–245). Hoboken, NJ: Wiley.

Yuan, K.-H., Bentler, P. M., & Chan, W. (2004). Structural equation modeling with heavy tailed distributions. *Psychometrika, 69,* 421–436.

Yuan, K.-H., Bentler, P. M., & Kano, Y. (1997). On averaging variables in a confirmatory factor analysis model. *Behaviormetrika, 24,* 71–83.

Yuan, K.-H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis: The univariate case and its multivariate implication. *Sociological Methods & Research, 34,* 249–258.

Yuan, K.-H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics & Data Analysis, 52,* 4842–4858.

Yuan, K.-H., Chan, W., & Bentler, P. M. (2000). Robust transformation with applications to structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 53,* 31–50.

Yuan, K.-H., Cheng, Y., & Zhang, W. (2010). Determinants of standard errors of MLE's in confirmatory factor analysis. *Psychometrika, 75*(4), 633–648.

Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology, 59,* 397–417.

Yuan, K.-H., & Hayashi, K. (2011). Fitting data to models: Structural equation modeling diagnosis using two scatter plots. *Psychological Methods, 15*(4), 335–351.

Yuan, K.-H., Hayashi, K., & Bentler, P. M. (2007). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypotheses. *Journal of Multivariate Analysis, 98,* 1262–1282.

Yuan, K.-H., Hayashi, K., & Yanagihara, H. (2007). A class of population covariance matrices in the bootstrap approach to covariance structure analysis. *Multivariate Behavioral Research, 42,* 261–281.

Yuan, K.-H., Kouros, C. D., & Kelley, K. (2008). Diagnosis for covariance structure models by analyzing the path. *Structural Equation Modeling, 15,* 564–602.

Yuan, K. H., & Lu, L. (2008). SEM with missing data and unknown population distributions using two-stage ML: Theory and its application. *Multivariate Behavioral Research, 43,* 621–652.

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika, 67,* 95–122.

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology, 33,* 241–265.

Yuan, K.-H., Wallentin, F., & Bentler, P. M. (2011). *ML versus MI for missing data with violation of distribution conditions*. Under editorial review.

Yuan, K.-H., Wu, R., & Bentler, P. M. (2011). Ridge structural equation modeling with correlation matrices for ordinal and continuous data. *British Journal of Mathematical and Statistical Psychology, 64,* 107–133.

Yuan, K.-H., & Zhong, X. (2008). Outliers, leverage observations and influential cases in factor analysis: Minimizing their effect using robust procedures. *Sociological Methodology, 38,* 329–368.

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*(3), 297–330.

Zu, J., & Yuan, K.-H. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research, 45,* 1–44.