MATTHEW W. GALLAGHER & TIMOTHY A. BROWN

14. INTRODUCTION TO CONFIRMATORY FACTOR ANALYSIS AND STRUCTURAL EQUATION MODELING

Confirmatory factor analysis (CFA) is a powerful and flexible statistical technique that has become an increasingly popular tool in all areas of psychology including educational research. CFA focuses on modeling the relationship between manifest (i.e., observed) indicators and underlying latent variables (factors). CFA is a special case of structural equation modeling (SEM) in which relationships among latent variables are modeled as covariances/correlations rather than as structural relationships (i.e., regressions). CFA can also be distinguished from exploratory factor analysis (EFA) in that CFA requires researchers to explicitly specify all characteristics of the hypothesized measurement model (e.g., the number of factors, pattern of indicator-factor relationships) to be examined whereas EFA is more data-driven. In this chapter we will provide a general introduction to how CFA and SEM can be used within educational research and other areas of psychology. We will begin with a nontechnical overview of the purpose of and methods underlying CFA and SEM before describing the various potential uses of CFA and SEM in educational research. We will then discuss the advantages of CFA and SEM over traditional methods of data analysis, provide an overview of the core steps in conducting CFA and SEM analyses, and discuss some practical issues in conducting these analyses such as software options. We then provide a brief summary of some of the more advanced methods in which CFA and SEM can be extended to conduct sophisticated analyses. We conclude with an illustrative series of example models in which the relationship between academic self-efficacy and academic performance is examined using CFA and SEM.

OVERVIEW AND GOALS OF CFA AND SEM

The goals of both CFA and SEM are to identify latent variables using a set of manifest indicators and to then evaluate hypotheses regarding the relationships among the latent variables. The conceptual background for conducting these analyses is the common factor model (Thurstone, 1947), which states that each manifest indicator is a linear function of one or more common factors and a unique factor. Factor analytic techniques therefore attempt to partition the variance of an indicator into (1) common variance, or the proportion of variance that is due to the latent variable, and (2) unique variance, which is a combination of random error variance

T. Teo (Ed.), Handbook of Quantitative Methods for Educational Research, 289–314.

(e.g., measurement error) and reliable variance that is specific to a particular item. Both EFA and CFA and SEM attempt to reproduce the observed intercorrelations/covariances between items with a more parsimonious set of latent variables. The primary difference, as mentioned above, is that CFA and SEM require researchers to explicitly specify every aspect of the models to be evaluated. CFA and SEM therefore require that researchers have a strong conceptual or empirical foundation to guide the specification and evaluation of models.

Common Uses of CFA and SEM

Some of the most common uses of CFA in educational and other areas of research include scale validation, construct validation, and evaluating measurement invariance. It is now considered standard practice to conduct a series of factor analyses when developing a new measure in psychological research. The standard progression is for researchers to begin by specifying an EFA model to evaluate an initial pool of items, and to then move to a CFA framework to provide a more rigorous evaluation of how a theoretical model represents the observed data. Through this process, researchers are able to determine the number of latent variables that best represents the constructs of interest and the pattern of relationships (i.e. factor loadings) between the observed items and latent variables. Thus, for instance, CFA can help researchers determine whether they should focus on the total score of a measure or subscales comprised of particular items from that scale. CFA also provides superior methods of evaluating other psychometric properties (e.g., reliability) of a scale than traditional methods such as Cronbach's alpha. For these reasons, educational researchers are strongly encouraged to use CFA when developing and validating new scales.

Another common application of CFA is to evaluate whether the measurement properties of an assessment are invariant. This is often an important second step in scale development. Measurement invariance can be tested cross-sectionally between groups or longitudinally between assessments of the same individuals. The use of CFA to evaluate measurement invariance across groups is discussed in detail in Chapter XX; in brief, these methods allow researchers to evaluate whether the relationship between indicators and latent variables is consistent between groups. For example, researchers could use CFA to evaluate measurement invariance between sexes on a test of mathematical proficiency. This analysis could help researchers determine whether any observed differences between sexes represent true differences between males and females or merely indicate that the items on a particular assessment function differently between sexes. The evaluation of measurement invariance is also a very important but underappreciated issue in longitudinal research, as the demonstration of measurement invariance across assessments provides the foundation for researchers to conclude that change in a latent variable across time truly represents growth or decline rather than inconsistent measurement. For additional information about how to test measurement invariance, readers are encouraged to consult Brown (2006), and Cheung and Rensvold (2002).

A third area in which CFA is commonly used is construct validation. CFA and SEM provide a useful framework for demonstrating both convergent and discriminant validity of theoretical constructs. Convergent validity is indicated by evidence that multiple indicators of theoretically linked constructs are strongly interrelated; for example, results on a series of tests that all purport to measure mathematical aptitude load on a single factor. Discriminant validity is indicated by evidence that indicators of theoretically distinct constructs do not correlate strongly with one another; for example, indicators of verbal and mathematical aspects of intelligence load on separate factors and correlate more so with indicators within the same domain of intelligence than with indicators within a different domain of intelligence. One of the most robust ways in which CFA can be used in construct validation is with the use of multitrait-multimethod techniques (Campbell & Fiske, 1959; Kenny & Kashy, 1992), a powerful yet infrequently used technique in which several constructs are measured using multiple methods and then modeled such that common variance due to method effects is separated from common variance due to latent traits.

Advantages of CFA and SEM

CFA and SEM have numerous advantages over traditional statistical techniques such as correlation and regression. One of the primary advantages of CFA and SEM is that they allow researchers to estimate the relationships between variables while accounting for measurement error. Traditional statistical techniques impose the generally unrealistic assumption that variables have been measured perfectly with no error. This assumption of error-free measurement is rarely appropriate in educational research or other areas of psychological research and results in parameter estimates that are biased to an unknown degree due to the failure to account for measurement error. By specifying latent variables that allow for the estimation of measurement error, researchers are able to obtain more accurate, reliable, and valid estimates of the relationships among latent constructs. This can also result in increased statistical power as the relationships between variables can be more precisely estimated after properly accounting for the role of measurement error. An important strength of CFA is the ability to model complex error structures among indicators to account for method effects (e.g., two self-report indicators of intelligence may correlate more strongly with one another than peer and teacher evaluations of intelligence would). Another important advantage of latent variable techniques such as CFA and SEM is that they permit the specification of complex longitudinal models that can help researchers to evaluate sophisticated theoretical models regarding change (e.g., latent growth curve models). A few of the more advanced methods are discussed later in this chapter. It is worth noting, however, that there are many circumstances in which CFA and SEM may not be the ideal method of data analysis. Most notably, if researchers are focusing on manifest variables that do not include measurement error (e.g., gender, grade point average), latent variable modeling techniques such as CFA and SEM may not be necessary.

CORE STEPS IN CFA AND SEM ANALYSES

CFA and SEM are complex statistical techniques that are performed in an iterative process and that present researchers with a number of important decisions during the process. The steps identified subsequently will provide readers with a general outline of the most common steps that researchers will follow when conducting CFA and SEM analyses. The subsequent steps presume that a researcher has already collected an appropriate dataset and has screened the data for outliers, univariate normality, and multivariate normality (Kline, 2011).

Specify Theoretical Model

The first step in conducting CFA and SEM analyses is for the researcher to clearly specify the theoretical model they are interested in testing. As mentioned previously, CFA and SEM differ from more data-driven procedures such as EFA and it is therefore crucial that researchers have a very clear idea of the specific models they want to test in advance. It is often helpful to diagram the planned models using common SEM notation and symbols. A fully notated example for a two-factor, six indicator CFA model can be seen in Figure 1. When diagramming SEM models, circles are used to denote latent variables, squares or rectangles are used to denote manifest or observed variables, correlations (standardized solutions) or covariances

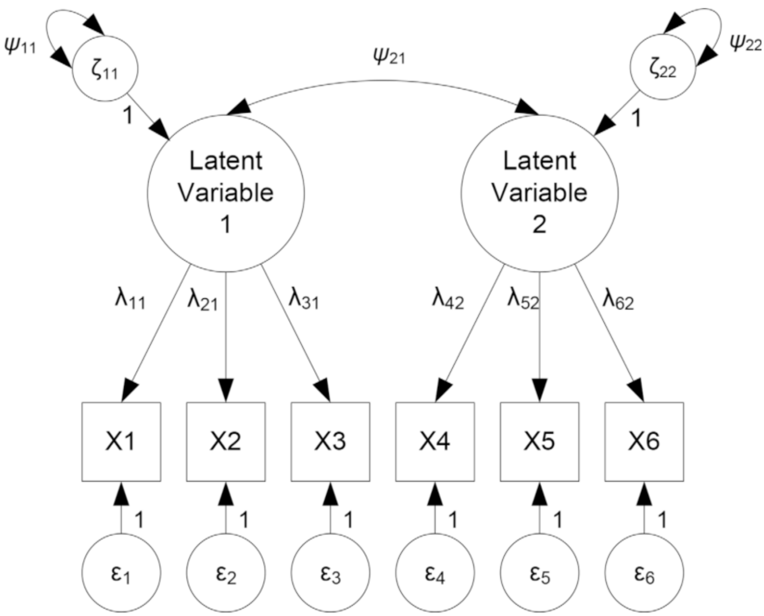

Figure 1. Example two factor CFA measurement model with six manifest indicators.

(unstandardized solutions) are denoted using double headed arrows, and single headed arrows are used to denote direct effects such as factor loadings or effects of one latent variable on another. Other common notations in SEM include the use of lambda (λ) to denote factor loadings, psi (ψ) to denote variances and covariances/correlations, and theta (θ) to indicate residuals and residual covariances.

Specify Measurement Model

After clearly specifying a theoretical model to be tested, researchers should next evaluate the measurement model for the latent variables of interest. The test of the measurement model should always be conducted prior to evaluating structural equation models. There are two important issues that researchers must consider when specifying the measurement model. The first issue is that researchers must ensure that CFA and SEM models are statistically identified. Adequate identification occurs when the number of parameters to be estimated in a model does not exceed the number of pieces of information in the variance-covariance input matrix. If a model is not adequately identified, then a solution cannot be solved as there are an infinite number of potential solutions. In CFA, the number of known pieces of information is determined by the size of the input variance/covariance matrix and can be calculated using the formula $b = [p * (p + 1)]/2$, where b is the number of elements in the input matrix and p is the number of variables included in the input matrix. For example, an input matrix of three variables would provide six pieces of information (three variances and three covariances) while an input matrix of two variables would only provide 3 pieces of information (two variances and 1 covariances). It is therefore only possible to freely estimate six parameters (i.e. three factor loadings and three residuals or two factor loadings, three residuals and the variance of the latent variable) in a model that uses an input matrix with three variables. When the number of freely estimated parameters equals the number of elements in the input matrix, then a model is just-identified and will fit the data perfectly. When the number of elements in the input matrix is greater than the number of freely estimated parameters, then a model is over-identified and the degrees of freedom (df) for the model can be determined by subtracting the number of freely estimated parameters from the number of known elements. When a model is overidentified, researchers are able to obtain goodness of fit statistics (discussed in more detail subsequently) that provide information about how well the specified CFA model reproduced the observed relationships in the sample data. It is also important to consider whether a model is locally identified in addition to being globally identified. Large models that include many variables will usually be over-identified but researchers should take care that each latent variable within a model is adequately identified. Situations in which the overall model is over-identified but certain components of the model are not locally identified are referred to as empirically under-identified solutions (e.g., selection of a marker variable that is unrelated to the other indicators that are specified to load on the same factor; see next paragraph). In these cases, either the model cannot be estimated or the model will converge but contain out of bounds parameter estimates (i.e., negative residual variances).

The second important issue in specifying CFA measurement models is setting the scale of latent variables. Latent variables do not have an inherent metric so the scale of these variables must always be set using one of the three methods. The most widely used method is the marker variable method, which involves fixing the factor loading of one indicator for each latent variable to be 1.0. This method results in setting the scale of the latent variable to the metric of the marker variable. Another common method is to standardize the factor variance, which involves fixing the variance of the latent variable to 1.0. The fixed factor method results in a standardized solution for the factor loadings and residuals. A third, but less common, method for setting the scale of latent variables is the effects coding approach (Little, Slegers, & Card, 2006). This method involves constraining the loadings of a latent variable to average 1.0. This is done by freely estimating all but one of the factor loadings and then fixing the remaining factor loading to equal the number of indicators minus each of the freely estimated factor loadings. The advantage of the effects coding method is that the parameters in a model (i.e., variances, means) reflect the observed scale of the indicator variables. The disadvantage of the effects coding method is that it requires slightly more complicated syntax. Each method is valid and will produce identical results in terms of model fit.

Estimate and Evaluate Measurement Model

The next step is to estimate the model using one of the many software packages designed for latent variable analysis (discussed later). The estimation process in CFA and SEM involves a fitting function (most commonly maximum likelihood; ML) that iteratively produces parameter estimates in an attempt to minimize the differences between the model-implied variance-covariance matrix and the sample variance-covariance matrix. For a more thorough description of the procedures involved in ML estimation and the circumstances in which other estimation methods are preferred, the reader is referred to Brown (2006), and Eliason (1993).

If a model converges successfully (i.e., a solution is obtained through ML estimation), researchers can then evaluate how acceptable the model fit the data. There are three primary components of the results that researchers should focus on when evaluating model fit. The first is overall goodness of fit, which reflects the degree to which the estimates of the CFA model reproduce the relationships between variables in the observed sample. A variety of fit statistics have been developed and it is generally recommended that researchers report multiple fit indices as they provide a more conservative and comprehensive evaluation of model fit. The classic goodness of fit index is model chi-square (χ^2). If the χ^2 value of a model exceeds the critical value from the χ^2 distribution (determined by the model's degrees of freedom), then the null hypothesis of adequate model fit is rejected. Although

χ^2 provides a very straightforward test of model fit, it has significant limitations including that it is overly sensitive to sample size and is therefore likely to reject very good models if the sample is large. For this reason, it is generally recommended that researchers report χ^2, but focus more on other fit indices when evaluating model fit. The most widely accepted global fit indices are the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992; Steiger & Lind, 1980), the comparative fit index (CFI; Bentler, 1990), the Tucker-Lewis index (TLI; Tucker & Lewis, 1973) which is sometimes also referred to as the non-normed fit index, and the standardized root mean square residual (SRMR; Bentler, 1995). For each of these fit statistics values generally range from 0 to 1. For the SRMR and RMSEA, values closer to 0 indicate better model fit, while values closer to 1 indicate better model fit for CFI and TLI. Recommendations vary in terms of what values of these fit statistics should be considered acceptable. Early guidelines for model fit suggested that CFI and TLI values greater than .9, and RMSEA values less than .1 should be considered acceptable (Bentler, 1990; MacCallum et al., 1996). More recently, the results of one of the most comprehensive simulation studies examining model fit (Hu & Bentler, 1999) suggested the following guidelines for considering a model to have good fit: (1) SRMR values close to or below .08, (2) RMSEA values close to or below .06, and (3) CFI and TLI values close to or above .95. It is important to recognize that these guidelines should used as general recommendations rather than rigid guidelines and that model fit should always be evaluated in terms of multiple fit indices rather than just a single fit statistic.

The second aspect of the results that researchers should examine when evaluating model fit is localized areas of poor fit. The global fit indices (e.g., RMSEA, CFI) provide a useful evaluation of the overall fit of a model but it is possible for a model to have good overall fit while poorly reproducing specific aspects of the model. The most common method for identifying localized misfit is by examining modification indices. Modification indices reflect the approximate change in the overall model χ^2 if a fixed or constrained parameter were to be freely estimated. Modification indices can be conceptualized as a χ^2 with 1 degree of freedom so modification indices of 3.84 or greater (i.e., the critical value of χ^2 with 1 df, $\alpha = .05$) suggest that the model fit could be significantly improved by freely estimating the parameter in question. Large modification indices may therefore provide researchers with information about how a particular model may be misspecified (e.g., the need for specifying a residual covariance between two indicators to account for a method artifact). However, it is important that researchers not make revisions to a model solely based on modification indices without a theoretical or empirical basis as this can lead to model overfitting and inappropriate capitalization on chance associations in the sample data (MacCallum, Roznowski, & Necowitz, 1992).

The third aspect of model evaluation is the interpretability, strength, and statistical significance of parameter estimates. It is important to confirm that model results do not contain any out of range values such as negative variances (often referred to as Heywood cases or offending estimates). This outcome can indicate significant problems in how the model was specified or problems with the sample data. Nonsignificant parameter estimates may indicate unnecessary parameters or items that are poor indicators of a latent construct. It is also useful to examine the completely standardized parameter estimates at this stage as these can be interpreted as correlations in the case of associations between latent variables, standardized regression coefficients in the case of factor loadings, and the proportion of variance unexplained in indicators in the case of residual variances. For example, a correlation approaching 1.0 between two latent variables may indicate that the two constructs are not truly distinct and that it may be more appropriate and parsimonious to collapse the variables into a single latent construct.

Consider Model Revisions

After estimating the measurement model and evaluating goodness of fit, the next step for researchers is to decide whether any revisions to the model are warranted. As mentioned previously, potential model revisions can be indicated based on modification indices or evaluation of the significance and strength of the parameter estimates. Any revisions should be made in an iterative fashion as modification indices are not independent of one another and minor changes in how a model is specified can produce large changes in both model fit and the parameter estimates. Researchers should err on the side of not making *post hoc* model revisions unless there is a strong theoretical or empirical foundation so as not to artificially inflate model fit by incorporating revisions that merely reflect sample-specific characteristics.

Specify Structural Models (If Applicable)

After establishing a satisfactory measurement model using CFA, researchers can then begin to specify structural equation models. Structural models allow researchers to explicitly model hypothesized relationships beyond the basic associations that are specified in CFA measurement models (i.e., factor covariances). More specifically, it is at this point that researchers can test hypothesizes regarding the presence or absence of regression effects among the latent variables, test models that involve the estimation of indirect effects to evaluate mediation hypotheses, test models that involve the estimation of interaction effects to evaluate moderation hypotheses, and test models that involve the specification of complex patterns of longitudinal growth such as latent growth curve models (Preacher et al., 2008). These are just a few of the many types of structural equation models that can be specified and researchers need to take care to use models that are appropriate for testing their specific theoretical hypotheses and models.

Reporting Results

The final step in conducting CFA and SEM analyses is to report the analyses in a clear and understandable manner so that it is possible for others to understand exactly how the models were specified and to replicate the models in independent samples (cf. McDonald & Ho, 2002). Given the complexity of many CFA and SEM models it is not always feasible to report every single parameter estimate from a model, but for transparency there are certain aspects of models that should always be presented. Researchers should clearly state how a model was specified, including information about the method of scale-setting used and justification of any *post hoc* model modifications. Multiple indices of model fit should be reported, preferably all five of the fit statistics described previously (i.e., χ^2, RMSEA, SRMR, CFI, and TLI). Researchers should also report information regarding the specific parameters of interest in a model, in both unstandardized and standardized form. Presentation of model parameters can often be accomplished most easily by presenting the results in a figure. It is also preferable to include the descriptive statistics (means and standard deviations) and the correlation/covariance matrix used to estimate the models so that readers can see the input matrix that was used to estimate the model (e.g., for data re-analysis).

PRACTICAL ISSUES IN USING CFA AND SEM

Software

There are now numerous software packages that are capable of estimating CFA and SEM models. Some of the most widely used programs include Mplus (Muthén & Muthén, 2008–2012), LISREL (Jöreskog & Sörbom, 1996), AMOS (Arbuckle, 2010), EQS (Bentler, 2006), CALIS (SAS Institute, 2005), Mx (Neale, Boker, Xie, & Maes, 2003), and multiple packages within the R statistical framework including SEM (Fox, 2006) and LAVAAN (Rosseel, 2011). All of these programs allow for the specification of CFA and SEM models either through the creation of syntax files or graphical interfaces. Each program is capable of estimating the most common CFA and SEM models but certain programs have unique characteristics or advantages. Mplus is in some ways the most flexible software program as it allows users to specify advanced models such as exploratory structural equation modeling (Asparaouhov & Muthén, 2009), multilevel structural equation modeling (Muthén & Asparouhov, 2008), and to use Bayesian estimation procedures that are not available or not easily specified in other programs. LISREL is a good program for didactic purposes as it allows researchers to specify models in terms of the matrices that comprise SEM models (e.g., lambda matrix for factor loadings). Mx has some advanced capabilities for estimating twin models and is the most common latent variable program in genetics research. All of the SEM packages within the R framework (e.g., LAVAAN) are open-source and free. AMOS and EQS both provide users with the option to specify models using a graphical interface. Although this capability may seem appealing, researchers should take caution as it is very easy to misspecify models when using graphical interfaces and ultimately, it is often easier to specify complex models using a syntax file.

Sample Size Requirements

As with any other area of research, the issue of statistical power is an important one when conducting CFA and SEM. There are multiple methods of determining power in CFA and SEM. One approach is based on statistical power for evaluating a model using RMSEA (MacCallum et al., 1996; Preacher & Coffman, 2006). This method requires researchers to specify the degrees of freedom for a model, alpha (typically .05), desired power (typically .80), and null and alternative values for RMSEA, and provides the sample size necessary to achieve the desired level of power in terms of the RMSEA model evaluation. An alternative and more flexible method of estimating power for latent variable models is the Monte Carlo method. Monte Carlo simulation studies allow researchers to evaluate the bias in specific parameter estimates and to determine power for detecting significant parameters based on population parameter estimates and varying sample sizes specified by the researcher. For a more detailed overview of methods for calculating power in CFA and SEM models and the use of Monte Carlo methods, readers are referred to Brown (2006), and Muthén and Muthén (2002).

Handling Missing Data

A common issue that applied researchers face when conducting CFA, SEM or any other form of statistical analysis is determining the most appropriate method for handling missing data. It is rare that researchers will collect a dataset in which no data are missing, and there are many reasons that data may be missing. Current typologies of missing data distinguish between three forms of missing data. In some situations data can be considered to be missing completely at random (MCAR) if, for example, a particular questionnaire was accidentally omitted in assessment packets for a few individuals. Data can also be considered missing at random (MAR) if, for example, attrition in a longitudinal study of academic outcomes is related to other variables in the data set such as academic engagement or motivation. Finally, data can be missing not at random or nonignorable if the pattern of missingness is related to some unobserved variable(s). A more complete description of the nature and implications of these patterns of missingness can be found in Enders (2010), but for the purposes of this chapter we will focus on what researchers can do to manage MAR and MCAR situations.

Many of the traditional methods of handling missing data (e.g., pairwise or listwise deletion) are inappropriate, as it has been repeatedly demonstrated that these approaches result in reduced statistical power and often produce biased parameter estimates (Allison, 2003; Enders, 2010; Schafer & Graham, 2002). The two methods that are the most appropriate strategies for handling missing data are full information maximum likelihood (FIML) estimation (also commonly called direct maximum likelihood), and multiple imputation. Both approaches are appropriate when data can be considered to be MAR or MCAR. We will focus our discussion on the FIML approach as this approach is easily implemented in many latent variable modeling software packages (e.g., Mplus, LISREL), and is generally regarded by methodologists as the most straightforward method of handling missing data (Allison, 2003). FIML methods use all of the available data to provide appropriate estimates of parameters and standard errors for a model while accounting for missing data. FIML is now the default estimator in Mplus and can be easily used in LISREL by including the MI keyword in the data line and indicating the missing data code in the dataset (e.g., MI = 9).

ADDITIONAL APPLICATIONS OF CFA AND SEM

Examining Mediation Using Structural Equation Modeling

Mediation can be defined as a process in which the effect of one variable (X) on another variable (Y) occurs through an intervening variable (M) (Baron & Kenny, 1986; MacKinnon, 2008). Mediation is an increasingly popular focus of research in educational research. SEM provides a very useful framework for evaluating mediational hypotheses. The use of latent variables allows researchers to obtain more accurate estimates of the overall indirect effect as well as the constituent parts of the indirect effect (i.e., M on X, Y on M). SEM also allows researchers to simultaneously evaluate multiple mediators and to extend mediation models to a longitudinal framework to evaluate how mediational processes unfold over time. Furthermore, it is possible to directly obtain bias-corrected and accelerated bootstrapped confidence intervals of the indirect effect, the current best-practice recommend method (Preacher & Hayes, 2008; MacKinnon, 2008) within SEM software packages such as Mplus. An example of how mediation can be examined within an SEM framework is presented later in this chapter.

Examining Moderation Using Structural Equation Modeling

Moderation (i.e., interactions) is also an increasingly popular area of research within education and other social sciences domains. Moderation can be tested in SEM for both categorical and continuous moderators. Categorical moderators can be evaluated using multiple groups models in which the parameters of interest are specified for each category of the moderator, with differences in the relationships between the groups considered evidence of moderation that can be tested for statistical significance using equality constraints. There are also multiple methods for evaluating continuous moderators within SEM. Little, Bovaird, and Widaman (2006) describe an approach in which a latent interaction term is specified by orthogonalizing the respective indicators of the independent variable and the moderator. Example syntax for how this approach be applied can be found in Schoemann (2010).

Longitudinal Extensions of Structural Equation Modeling

One of the most useful ways in which CFA and SEM can be extended is to examine longitudinal data. There are many ways in these methods can be extended to model complex patterns of change. Cross-lagged panel models allow researchers to evaluate how interindividual standing in latent constructs changes over time (Burkholder & Harlow, 2003). Latent growth curve models allow researchers to examine intraindividual trajectories of change and can be used to evaluate non-linear and other complex patterns of change (Bollen & Curran, 2006; Preacher et al., 2008). Latent difference score models are a third approach for modeling longitudinal change and allow researchers to examine intraindividual change in latent constructs between two assessments (McArdle, 2009). Each of these methods is well-suited to studying a variety of research topics and researchers interested in learning more about these topics are encouraged to consult Collins (2006), Selig and Preacher (2009), or Little, Bovaird, and Card (2007).

Multilevel Structural Equation Modeling

The final extension of CFA and SEM that we will mention is multilevel structural equation modeling (MSEM; Muthén & Asparouhov, 2008). MSEM is a relatively recent development and combines all of the advantages of hierarchical linear modeling (e.g., accounting for nested dependencies in data) and SEM (e.g., accounting for measurement error). MSEM is therefore an extremely robust statistical framework as it allows researchers to specify models that are not possible when using either hierarchical linear modeling or SEM. MSEM remains an infrequently used method given its complexity. However, descriptions of how these methods can be used in applied research are increasingly common (e.g., Preacher, Zyphur, & Zhang, 2010) and MSEM is likely to be a major area of growth in the next decade.

Illustrative Study

An example study will now be presented to demonstrate the sequence of steps that researchers will typically follow when conducting a study involving CFA and SEM. The data for these example models come from a study in which undergraduates completed a series of self-report questionnaires during their first semester of college to identify the psychological variables (e.g., self-efficacy, hope, engagement) that best predict academic performance during the first four years of college (Gallagher & Lopez, 2008). Participants were 229 students (129 males, 100 females) at a large Midwestern university who participated in exchange for psychology course credit. Prior to completing their first semester, participants completed the academic self-efficacy scale (Chemers, Hu, & Garcia, 2001), identified their goal for their GPA after four years of college, and provided consent to have their academic performance (semester GPA) tracked by the investigators through the University Registrar's office.

For illustration purposes, we will focus on just the relationships between academic self-efficacy, self-reported goals for GPA during the first semester of college, and cumulative GPA after four years of college. The descriptive statistics and correlation matrix used for these analyses are presented in Table 1. There were no missing data

Table 1. Sample correlations, standard deviations (SD) and means (M) for self-efficacy (SE) items, college grade point average goal (GPAGOAL), and four year college grade point average (GPA)

	SE1	*SE2*	*SE3*	*SE4*	*SE5*	*SE6*	*SE7*	*SE8*	*GPAGOAL*	*GPA*
SE1	1									
SE2	.408	1								
SE3	.365	.533	1							
SE4	.256	.247	.363	1						
SE5	.540	.432	.497	.325	1					
SE6	.432	.422	.507	.374	.756	1				
SE7	.246	.354	.441	.254	.476	.464	1			
SE8	.385	.385	.375	.316	.576	.579	.421	1		
GPAGOAL	.032	.051	.110	.148	.234	.309	.122	.142	1	
GPA	.263	.166	.278	.203	.302	.371	.102	.215	.350	1
N	229	229	229	229	229	229	229	229	228	147
M	5.14	5.45	4.92	4.83	4.95	5.07	4.67	5.92	3.39	2.96
SD	1.50	1.39	1.39	1.60	1.14	1.29	1.31	1.15	.34	.49

for the self-efficacy variables but data were missing for one person's GPA goals and 82 people were missing data on four year college GPA. These missing data were considered MAR and were accommodated using FIML. A series of four models will be presented to demonstrate the common steps researchers may take when using CFA and SEM. The first model uses CFA to evaluate the measurement model of the academic self-efficacy scale. The second model is an extension of the one-factor measurement model to include a correlated residual between two items. The third model examines the effect of academic self-efficacy on cumulative college GPA using SEM. The fourth model tests a mediation model in which participants' goals for GPA reported during their first semester of college partially mediates the effects of academic self-efficacy beliefs on cumulative GPA after four years of college. Mplus syntax for each of these examples will be presented, but each model could be conducted in the other latent variable software programs mentioned previously.

Evaluating the Measurement Model

The first step in evaluating the effects of academic self-efficacy on academic performance is to determine how well the latent construct of academic self-efficacy was measured. This can be accomplished using a basic one-factor CFA model. Annotated Mplus syntax and selected output from this model are presented in Table 2. As can be seen in Table 2, the syntax required for specifying a one-factor

Table 2. Mplus syntax and selected output of confirmatory factor analysis of the academic self-efficacy scale

SYNTAX:

```
TITLE: Academic Self-Efficacy Confirmatory Factor Analysis
DATA: FILE IS acaselfeff.dat;
VARIABLE:
    NAMES ARE id gpagoal gpa4year aselfe1-aselfe8;          !Identify all
    variables in data set
    USEVARIABLES ARE aselfe1-aselfe8;   !Specify variables to be used in model
    MISSING are all (-9);
ANALYSIS:
    TYPE IS GENERAL;
    ESTIMATOR IS ML;
MODEL:
    acaeffic by aselfe1-aselfe8;                 !Specify 8 items as indicators
                                  !Mplus defaults to marker variable identification
OUTPUT: MODINDICES(4) STANDARDIZED;          !Request completely
                                  standardized results and !modification indices
```

SELECTED OUTPUT:

TESTS OF MODEL FIT

Chi-Square Test of Model Fit	
Value	58.290
Degrees of Freedom	20
P-Value	0.0000
CFI/TLI	
CFI	0.944
TLI	0.921
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.091
90 Percent C.I.	0.065 0.119
SRMR (Standardized Root Mean Square Residual)	
Value	0.044

MODEL RESULTS

	Estimate	S.E	Est./S.E.	Two-Tailed P-Value
ACAEFFIC BY				
ASELFE1	1.000	0.000	999.000	999.000
ASELFE2	0.901	0.130	6.946	0.000
ASELFE3	1.008	0.135	7.491	0.000
ASELFE4	0.804	0.142	5.643	0.000
ASELFE5	1.132	0.123	9.189	0.000

ASELFE6	1.260	0.142	8.887	0.000
ASELFE7	0.864	0.125	6.934	0.000
ASELFE8	0.899	0.114	7.886	0.000
Intercepts				
ASELFE1	5.140	0.099	51.838	0.000
ASELFE2	5.445	0.092	59.276	0.000
ASELFE3	4.921	0.091	53.814	0.000
ASELFE4	4.825	0.106	45.686	0.000
ASELFE5	4.948	0.075	65.923	0.000
ASELFE6	5.074	0.085	59.473	0.000
ASELFE7	4.672	0.087	54.013	0.000
ASELFE8	5.917	0.076	78.046	0.000
Variances				
ACAEFFIC	0.745	0.165	4.518	0.000

COMPLETELY STANDARDIZED MODEL RESULTS

				Two-Tailed
	Estimate	S.E	Est./S.E.	P-Value
ACAEFFIC BY				
ASELFE1	0.575	0.049	11.835	0.000
ASELFE2	0.559	0.050	11.146	0.000
ASELFE3	0.628	0.045	14.009	0.000
ASELFE4	0.434	0.058	7.511	0.000
ASELFE5	0.861	0.024	36.073	0.000
ASELFE6	0.842	0.025	33.350	0.000
ASELFE7	0.570	0.049	11.644	0.000
ASELFE8	0.677	0.040	16.787	0.000
Variances				
ACAEFFIC	1.000	0.000	999.000	999.000

R-SQUARE

Observed				Two-Tailed
Variable	Estimate	S.E.	Est./S.E.	P-Value
ASELFE1	0.331	0.056	5.917	0.000
ASELFE2	0.313	0.056	5.573	0.000
ASELFE3	0.395	0.056	7.005	0.000
ASELFE4	0.188	0.050	3.755	0.000
ASELFE5	0.741	0.041	18.037	0.000
ASELFE6	0.709	0.043	16.675	0.000
ASELFE7	0.325	0.056	5.822	0.000
ASELFE8	0.458	0.055	8.393	0.000

MODEL MODIFICATION INDICES

Minimum M.I. value for printing the modification index 4.000

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
WITH Statements				
ASELFE2 WITH ASELFE1	4.201	0.204	0.204	0.145
ASELFE3 WITH ASELFE2	21.306	0.409	0.409	0.330
ASELFE4 WITH ASELFE3	4.267	0.224	0.224	0.145
ASELFE5 WITH ASELFE1	4.856	0.140	0.140	0.197
ASELFE5 WITH ASELFE2	5.477	–0.138	–0.138	–0.207
ASELFE5 WITH ASELFE3	5.311	–0.132	–0.132	–0.213
ASELFE5 WITH ASELFE4	4.298	–0.146	–0.146	–0.176
ASELFE6 WITH ASELFE1	5.362	–0.169	–0.169	–0.197
ASELFE6 WITH ASELFE2	4.483	–0.144	–0.144	–0.179
ASELFE6 WITH ASELFE5	13.271	0.203	0.203	0.505
ASELFE7 WITH ASELFE3	4.567	0.177	0.177	0.153

CFA model in Mplus is straightforward. The first few lines of syntax involve providing a title for the analysis, identifying the location of the data file (for Mplus the data file can be either tab-delimited, comma-delimited, or a fixed width ASCII file), providing variable names for all variables included in the dataset, selecting the specific variables that are included in the model to be analyzed, and identifying what numeric value used to indicate missing data (blanks can also be used as a missing data code if the data are in a fixed width ASCII file). The syntax for specifying the one-factor CFA model requires only two lines in Mplus. The first line signifies that the latent construct of Academic Self-Efficacy is identified by the eight items of the academic self-efficacy scale (Chemers et al., 2001). For this model, the latent construct of academic self-efficacy is identified using the marker variable method. This method of model identification is the default method in Mplus and simply requires that the corresponding indicators for the latent variable are specified (e.g., acaeffic by aselfe1-aselfe8;); by default, Mplus uses the first indicator after the "by" keyword (aselfe1) as the marker variable by fixing its unstandardized factor loading to 1.0. The final line of syntax instructs Mplus to provide additional output in the form of modification indices that equal 4.0 or above, and the standardized/ completely standardized estimates.

Because the CFA model of the academic self-efficacy scale converged successfully with no error messages, the first step is to examine the model fit statistics. The χ^2 test of model fit indicated significant model misfit ($p < .001$). However, as previously mentioned, the χ^2test is an overly conservative test and it is therefore more important to focus on the remaining model fit statistics. Although the SRMR is consistent with good model fit (.044), the CFI, TLI, and RMSEA indicate marginal fit (values of .944,.922, and .091, respectively). Taken together, these results suggest the specified model fit does not provide a good representation of the data, so the next step is to examine the modification indices to determine whether it may be possible to improve fit by respecifying the model. As noted earlier, this should only be done if substantively justified, as high modification indices do not necessarily indicate a relationship that is theoretically meaningful.

In the results presented in Table 2, the largest modification index is for the residual covariance between items 2 and 3 of the scale. The value for this modification index (21.31) is well above 3.84 and indicates there is a relationship between these two items that is not sufficiently accounted for by the latent variable of academic self-efficacy. An examination of the content of these two items reveals that this may be explained by a method effect arising from similar wording. Given the common stem of these items, it was deemed appropriate to specify a residual covariance between these two items to account for the method effect.

Revising the Measurement Model

The syntax and selected output from a second measurement model of the academic self-efficacy scale in which a residual covariance between items two and three is specified is presented in Table 3. As seen in Table 3, including the residual covariance requires an additional line of syntax (aselfe2 with aselfe3;). An examination of the model fit for this second model reveals that the inclusion of the residual covariance between the two items significantly improved model fit. The CFI and TLI values are both above .95, SRMR is below .08, and RMSEA equals .06. Together, these model fit statistics indicate good model fit for the one-factor measurement model of the academic self-efficacy scale that includes the residual covariance between items two and three. Inspection of modification indices indicates there are no remaining salient focal areas of ill fit.

An examination of the completely standardized factor loadings in this revised measurement model indicates that all eight of the items of the academic self-efficacy scale have moderate to large factor loadings (range = .43 to .87). The square of these loadings represents the proportion of the variance in the indicators explained by the latent constructs. Thus, the magnitude of these loadings indicates that a moderate proportion of the variance in the indicators could be explained by the latent variable of academic self-efficacy. It appears that the eight items are all adequate indicators of academic self-efficacy. Furthermore, an examination of the residual covariance parameter estimate indicates that this relationship was

Table 3. Mplus syntax and selected output of confirmatory factor analysis of the academic self-efficacy scale with residual covariance specified between items 2 and 3

SYNTAX:

```
TITLE: Academic Self-Efficacy Confirmatory Factor Analysis with Residual
   Covariance
DATA: acaselfeff.dat;
VARIABLE:
   NAMES ARE id gpagoal gpa4year aselfe1-aselfe8;
   USEVARIABLES ARE aselfe1-aselfe8;
   MISSING are all (-9);
ANALYSIS: ESTIMATOR IS ML;
MODEL:
   acaeffic by aselfe1-aselfe8;
   aselfe2 with aselfe3;                    ! specify residual covariance
OUTPUT: STANDARDIZED;
```

SELECTED OUTPUT:

TESTS OF MODEL FIT

Chi-Square Test of Model Fit	
Value	36.638
Degrees of Freedom	19
P-Value	0.0088
CFI/TLI	
CFI	0.974
TLI	0.962
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.064
90 Percent C.I.	0.031 0.094
SRMR (Standardized Root Mean Square Residual)	
Value	0.036

MODEL RESULTS

	Estimate	S.E	Est./S.E.	Two-Tailed P-Value
ACAEFFIC BY				
ASELFE1	1.000	0.000	999.000	999.000
ASELFE2	0.857	0.129	6.644	0.000
ASELFE3	0.973	0.134	7.256	0.000
ASELFE4	0.799	0.144	5.566	0.000
ASELFE5	1.153	0.126	9.179	0.000
ASELFE6	1.278	0.145	8.835	0.000
ASELFE7	0.861	0.126	6.852	0.000
ASELFE8	0.906	0.116	7.838	0.000

ASELFE2 WITH				
ASELFE3	0.413	0.098	4.222	0.000
COMPLETELY STANDARDIZED MODEL RESULTS				
				Two-Tailed
	Estimate	S.E	Est./S.E.	P-Value
ACAEFFIC BY				
ASELFE1	0.572	0.049	11.683	0.000
ASELFE2	0.529	0.052	10.140	0.000
ASELFE3	0.603	0.047	12.942	0.000
ASELFE4	0.429	0.058	7.377	0.000
ASELFE5	0.870	0.023	37.511	0.000
ASELFE6	0.849	0.025	34.235	0.000
ASELFE7	0.564	0.049	11.422	0.000
ASELFE8	0.677	0.040	16.778	0.000
ASELFE2 WITH				
ASELFE3	0.317	0.063	5.059	0.000

statistically significant. All subsequent models therefore include the residual covariance between items two and three.

Extending to Structural Equation Modeling with an Outcome

After establishing an appropriate measurement model, the next step would be to begin examining the relationship between academic self-efficacy and GPA. In this situation, cumulative GPA is a manifest variable outcome and we therefore do not include an intermediate model in which the measurement model for the outcome is also evaluated. The syntax and selected output from the structural equation model in which we examine the effect of academic self-efficacy on cumulative college GPA four years later is presented in Table 4. As seen in Table 4, the only modifications to the syntax required to specify this model is to add the GPA variable to the usevariables list and to add an additional line of syntax (gpa4year on acaeffic;). The inclusion of GPA as an outcome and the specification of the effect of academic self-efficacy on GPA did not worsen fit: as with the previous model, CFI and TLI values are both above .95, SRMR is below .08, and RMSEA equals .06. The results of this model indicate that academic self-efficacy is a significant predictor of cumulative college GPA four years later. The unstandardized effect was B = .239, SE = .055, $p < .001$. The completely standardized effect was $\beta = .415$ and academic self-efficacy predicted 17.2% of the variance in cumulative college GPA. The completely standardized results of this model are presented in Figure 2. These results support

Table 4. Mplus syntax and selected output of structural equation model examining the effect of academic self-efficacy scale on four-year college grade point average

SYNTAX:

```
TITLE: Academic Self-Efficacy SEM with 4year GPA as outcome
DATA: FILE IS acaselfeff.dat;
VARIABLE:
   NAMES ARE id gpagoal gpa4year aselfe1-aselfe8;
   USEVARIABLES ARE gpa4year aselfe1-aselfe8;
   MISSING are all (-9);
ANALYSIS: ESTIMATOR IS ML;
MODEL:
   acaeffic by aselfe1-aselfe8;
   aselfe2 with aselfe3;
   gpa4year on acaeffic;    !Estimate effect of Academic Self-Efficacy on GPA
OUTPUT: STANDARDIZED;
```

SELECTED OUTPUT:

TESTS OF MODEL FIT

Chi-Square Test of Model Fit	
Value	44.345
Degrees of Freedom	26
P-Value	0.0139
CFI/TLI	
CFI	0.974
TLI	0.964
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.056
90 Percent C.I.	0.025 0.083
SRMR (Standardized Root Mean Square Residual)	
Value	0.039

UNSTANDARDIZED MODEL RESULTS

	Estimate	S.E	Est./S.E.	Two-Tailed P-Value
GPA4YEAR ON				
ACAEFFIC	0.239	0.055	4.327	0.000

COMPLETELY STANDARDIZED MODEL RESULTS

	Estimate	S.E	Est./S.E.	Two-Tailed P-Value
GPA4YEAR ON				
ACAEFFIC	0.415	0.078	5.337	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GPA4YEAR	0.172	0.065	2.669	0.008

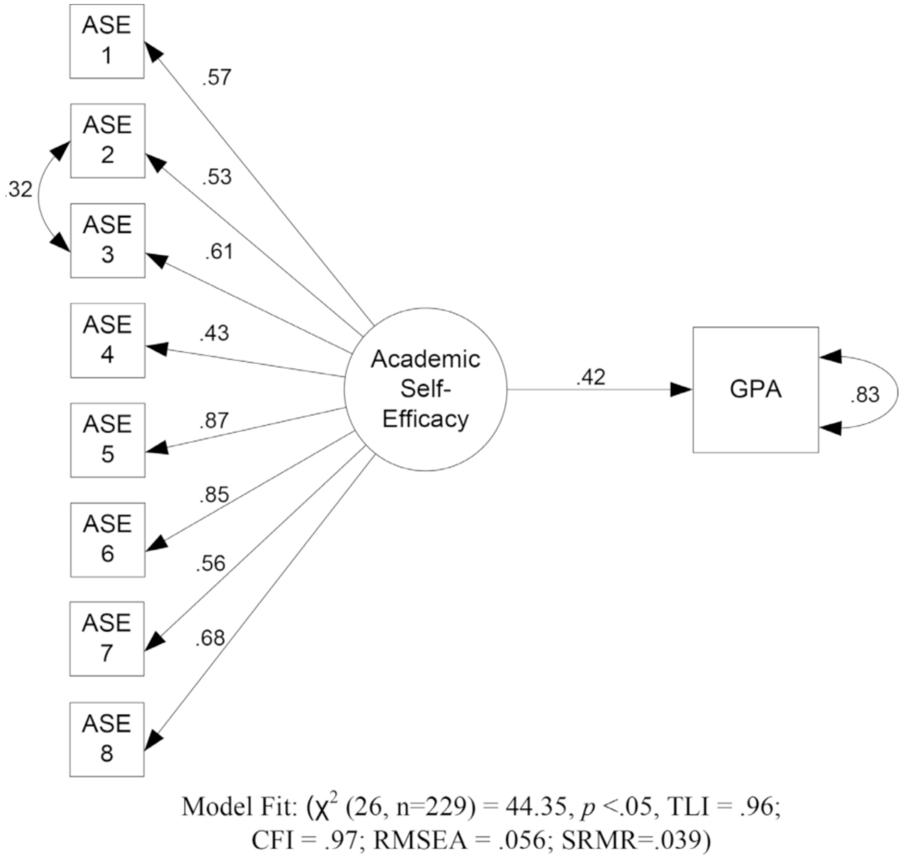

Figure 2. Example figure for presenting SEM results. Results correspond to the completely standardized results in Table 4.

the hypothesis that academic self-efficacy is a predictor of academic outcomes and provides the basis for examining potential mechanisms of the effects of academic self-efficacy on cumulative GPA.

Evaluating a Mediation Model

The final example model is a mediation model in which we examine whether the effects of academic self-efficacy on cumulative college GPA four years later are partially mediated by the GPA goals students set during their first semester of college. The syntax and selected output from the SEM in which we examine the indirect effect of academic self-efficacy on cumulative college GPA four years later via GPA goals are presented in Table 5. As seen in Table 5, the specification of this mediation model requires just a few minor additions to the syntax of the previous SEM model. The usevariables line is modified to include the additional variable of

Table 5. Mplus syntax and selected output of structural equation model examining the indirect effect of academic self-efficacy on four-year college grade point average via gpa goals in 1st semester

SYNTAX:

```
TITLE: Mediation Model: Aca Self-Efficacy → GPA Goal → 4yearGPA
DATA: FILE IS acaselfeff.dat;
VARIABLE:
   NAMES ARE id gpagoal gpa4year aselfe1-aselfe8;
   USEVARIABLES ARE gpagoal gpa4year aselfe1-aselfe8;
   MISSING are all (-9);
ANALYSIS: ESTIMATOR IS ML;
MODEL:
   acaeffic by aselfe1-aselfe8;
   gpa4year on acaeffic;
   gpa4year on gpagoal;
   gpagoal on acaeffic;
   aselfe2 with aselfe3;
Model Indirect:       !specify estimation of indirect effect
   gpa4year ind gpagoal acaeffic;
OUTPUT: CINTERVAL STANDARDIZED;
```

SELECTED OUTPUT:

```
TESTS OF MODEL FIT
Chi-Square Test of Model Fit
   Value                    58.633
   Degrees of Freedom       33
   P-Value                  0.0039
CFI/TLI
   CFI                      0.965
   TLI                      0.952
```

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.058	
90 Percent C.I.	0.033	0.082

SRMR (Standardized Root Mean Square Residual)

Value	0.043

UNSTANDARDIZED MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GPA4YEAR ON ACAEFFIC	0.172	0.043	3.996	0.000
GPAGOAL ON ACAEFFIC	0.097	0.024	3.995	0.000
GPA4YEAR ON GPAGOAL	0.413	0.115	3.595	0.000

COMPLETELY STANDARDIZED MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GPA4YEAR ON ACAEFFIC	0.345	0.079	4.349	0.000
GPAGOAL ON ACAEFFIC	0.284	0.067	4.232	0.000
GPA4YEAR ON GPAGOAL	0.282	0.075	3.759	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GPAGOAL	0.081	0.038	2.116	0.034
GPA4YEAR	0.254	0.068	3.737	0.000

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GPA4YEAR GPAGOAL ACAEFFIC	0.040	0.015	2.723	0.006

CONFIDENCE INTERVALS OF INDIRECT EFFECTS

	Lower .5%	Lower 2.5%	Estimate	Upper 2.5%	Upper. 5%
GPA4YEAR GPAGOAL ACAEFFIC	0.002	0.011	0.040	0.069	0.078

GPA goals, the effect of academic self-efficacy on GPA goals is specified (gpagoal on acaeffic;), the effect of GPA goals on cumulative GPA is specified (gpa4year on gpagoal;), the estimation of the indirect effect is requested by including “Model Indirect: gpa4year ind gpagoal acaeffic;”, and CINTERVAL is added to the output line so that confidence intervals of the indirect effect can be evaluated to determine whether there is evidence of mediation. The model fit for this mediation model was good: CFI and TLI values are above .95, SRMR is below .08, and RMSEA equals .06. The results indicated that there was a significant indirect effect of academic self-efficacy on cumulative college GPA four years later via GPA goals. The estimate of the

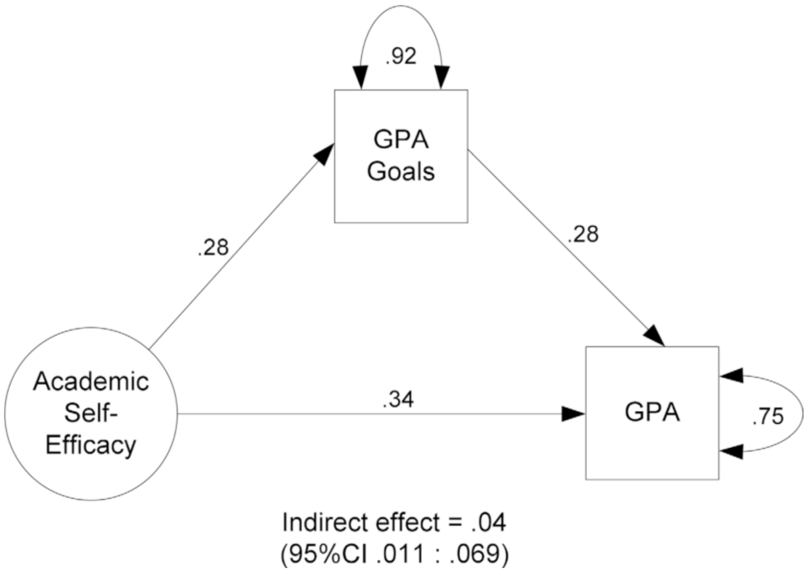

Figure 3. Example figure for presenting SEM mediation model results. results correspond to the completely standardized results in Table 5.

indirect effect was significant (B = .040, SE = .015, $p < .01$) and the 95% confidence interval of the indirect effect (.011 : .069) did not include 0. A path diagram with the completely standardized results of this mediation model can be seen in Figure 3. These results suggest that the academic self-efficacy beliefs may promote superior academic performance in college by causing students to set higher GPA goals for themselves. The theoretical implications of these results are not important for the purposes of this chapter, but the models described here and presented in Tables 2–5 provide an introduction to how CFA and SEM can be used in educational research.

SUMMARY

CFA and SEM are powerful statistical tools that have become increasingly popular in education research. The topics discussed within this chapter are just some of the many ways that these techniques can be used to evaluate measurement models and test complex theoretical models. The growth of these techniques has coincided with the development of more user-friendly statistical software for conducting these analyses and an increasing amount of publications providing didactic information about how these techniques can be applied to various research topics. Below we provide a few recommendations for resources that educational researchers may find helpful for additional information about how to apply these techniques in their own research programs.

SUGGESTIONS FOR FURTHER READING

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

Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge Press.

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

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201–226.

Teo, T., & Khine, M. S. (Eds.) (2009). *Structural equation modeling in educational research: Concepts and applications*. Rotterdam: Sense Publishers.

REFERENCES

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

Arbuckle, J. L. (2010). *IBM SPSS Amos 19 User's Guide.* Crawfordville, FL: Amos Development Corporation.

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

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

Bentler, P. M. (1990). Comparative fit indices in structural equation models. *Psychological Bulletin, 28*, 97–104.

Bentler, P. M. (2006). *EQS for Windows (Version 6.0)* [Computer software]. Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

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

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

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258.

Burkholder, G. J., & Harlow, L. L. (2003). An illustration of a longitudinal cross-lagged design for larger structural equation models. *Structural Equation Modeling*, *10*, 465–486.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Chemers, M. M., Hu, L., & Garcia, B. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*, 55–65.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Myths and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112,* 558–577.

Collins, L. M. (2006): Analysis of longitudinal data; The integration of theoretical model, temporal design, and statistical model, *Annual Review of Psychology 57,* 505–528.

Eliason, S. R. (1993), *Maximum likelihood estimation*. Newbury Park, CA: Sage.

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

Fox, J. (2006). Structural equation modeling with the sem package In R. *Structural Equation Modeling, 13*, 465–486.

Gallagher, M. W., & Lopez, S. J. (2008, August). *The unique effects of hope and self-efficacy on academic performance*. Poster presented at the American Psychological Association 116th Annual Convention, Boston, MA.

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

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

Kenny, D. A., & Kashy, D. A. (1992). The analysis of the multitrait-multimethod matrix using confirmatory factor analysis. *Psychological Bulletin, 112,* 165–172.

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

Little, T. D., Bovaird, J. A., & Card, N. A. (Eds.). (2007). *Modeling contextual effects in longitudinal studies*. Mahwah, NJ: Erlbaum.

Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and interaction terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling, 13*, 497–519.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling, 13,* 59–72.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201–226.

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

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

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Mahwah, NJ: Erlbaum.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577–605.

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64–82.

Muthén, B. O., & Asparouhov, T. (2008). 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. O., & Muthén, L. K. (2008–2012). *Mplus user's guide.* Muthén & Muthén: Los Angeles, CA.

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

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

Preacher, K. J., & Coffman, D. L. (2006, May). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://www.quantpsy.org/.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40, 879–891.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling.* Thousand Oaks, CA: Sage Publications.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15,* 209–233.

SAS Institute (2005). *SAS/STAT User's Guide,* Version 9. Cary, NC: SAS Institute Inc.

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

Schoemann, A. D. (2010). *Latent variable moderation.* Retrieved January 22nd, 2012, from http://www.quant.ku.edu/pdf/16.1%20latent%20variable%20moderation.pdf.

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development, 6,* 144–164

Steiger, J. H., & Lind, J. C. (1980). *Statistically-based tests for the number of common factors.* Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City. May 30, 1980.

Teo, T., & Khine, M. S. (Eds.) (2009). *Structural equation modeling in educational research: Concepts and applications*. Rotterdam: Sense Publishers.

Thurstone, L. L. (1947). *Multiple-Factor Analysis*. Chicago: University of Chicago Press.