CHAPTER 1

Structural Equation Modeling

An Overview

Rick H. Hoyle

Structural equation modeling (SEM) is a general statistical approach to modeling the mechanisms presumed to give rise to observed variability, covariation, and patterns in data. These mechanisms typically are of theoretical interest, though they may also include methodological and artifactual mechanisms. Although the data to which these mechanisms are presumed relevant are observed, models may include mechanisms that are unobserved, or latent. For this reason, SEM has been referred to as latent variable modeling. The primary data for many uses of SEM are covariances, which explains why SEM has also been referred to as covariance structure modeling. And the intent of many uses of SEM is to model putative causal effects between variables, explaining why SEM is sometimes referred to as causal modeling. Regardless of the label, the set of statistical methods referred to as SEM in this book offers a comprehensive and flexible approach to evaluating models of theoretical and methodological interest to researchers in the social and behavioral sciences.

As evidenced by the number of topics included in this second edition of the *Handbook* that were not in the first edition, SEM is an evolving and expanding statistical approach. Although the core capabilities of SEM have been well established since the early 1970s and generally accessible to researchers since the early 1980s, new capabilities are being developed and incorporated into computer programs for SEM analyses with regularity (see Matsueda, Chapter 2, for an informative history of SEM). These emerging capabilities coupled with powerful and intuitive computer programs for implementing them have spurred phenomenal growth in the amount and diversity of SEM usage. This thoroughly revised and updated *Handbook* is a response to that growth. The goal of this book is to provide detailed coverage of SEM, beginning with foundational concerns and moving through an impressive array of modeling possibilities.

In this opening chapter, I provide a brief introduction to SEM that also serves as an overview of the book. I begin by discussing the relation between SEM and statistical methods with which many readers new to SEM will be familiar. I then provide a brief description of the basic logic of SEM as it typically is used in the social and behavioral sciences. The heart of the chapter is the presentation of an implementation framework that serves as both context for the remainder of the chapter and an outline of the first part of the book. In the final section of the chapter, I offer a high-level view of data and models for which SEM can be profitably used and point the reader to chapters in the second and third parts of the book that offer detailed descriptions and demonstrations.

SEM IN RELATION TO OTHER STATISTICAL MODELS

As a linear model used primarily to model relations between variables, SEM is not unrelated to narrower and more familiar statistical models such as analysis of variance (ANOVA), multiple regression analysis, and principal factor analysis. Indeed, any of these analyses could be accomplished, and would yield identical results (assuming use of the same estimator, e.g., ordinary least squares), using SEM. As such, SEM can be described as, in part, a generalization, integration, and extension of these familiar models.

Consider, for example, tests involving means. In the most limited case, a single mean estimated from a sample is compared against a population value, often zero, and the difference tested for significance. This test can be usefully generalized to the situation in which both means are estimated from samples, which may be independent or dependent; alternatively, the means may come from two observations of the same sample. The same comparison could be made using ANOVA, which offers the additional benefit of allowing for both more than two means and means generated by more than one factor. The number of levels a factor might reasonably take on in ANOVA is relatively small, making it unsuitable for independent variables measured on a continuous or quasi-continuous scale such as survey items. Multiple regression analysis can accommodate both traditional ANOVA factors and quantitative measures that take on many values; thus, it has all the capabilities of ANOVA and more. Although both ANOVA and multiple regression analysis can accommodate multiple dependent variables, they are limited in how the relations between those variables can be specified. Furthermore, a variable can be either an independent or a dependent variable, but not both. SEM can accommodate both analytic situations. For instance, a set of variables might be used to predict a pair of outcomes that are correlated, uncorrelated, or related in such a way that one is regressed on the other. In the latter case, one of the dependent variables is also an independent variable because it is used to predict the other dependent variable. The use of SEM to compare means when one or more assumptions of ANOVA are not met (e.g., homogeneity of variance) is the topic of Chapter 21 (Thompson, Liu, & Green), which shows how ANOVA is a special case of SEM.

An alternative path to SEM that highlights additional capabilities begins with the zero-order correlation coefficient, which indexes the nondirectional association between two variables. The degree to which that association can be attributed to a common influence can be evaluated using partial correlation analysis, assuming the putative influence has been measured. In the case of three or more variables, this logic can be extended to consider common influences that are not measured using factor analysis. The traditional factor analysis model is referred to as exploratory factor analysis (EFA) because those influences, even in the presence of well-developed hypotheses, are not specified a priori. More an inconvenience than a limitation is the fact that an infinite number of factor scores can be derived from the parameters (factor loadings and uniquenesses) estimated by EFA (Steiger & Schönemann, 1978; see Devlieger & Rosseel, Chapter 17, for coverage of SEM analyses using factor scores). Finally, EFA requires that uniquenesses be uncorrelated. Factors in the context of SEM have traditionally been referred to as latent variables and are modeled in a more flexible, mathematically defensible manner that allows for a wide array of models that could not be evaluated using EFA. Applications of SEM that focus exclusively on the relations between latent variables and their indicators are referred to as restricted factor analysis or, more commonly, confirmatory factor analysis (CFA) (Brown, Chapter 14). Both labels are apt because it is the restrictions that CFA requires that make it confirmatory (i.e., subject to statistical testing). Conditional on appropriate restrictions (illustrated below), CFA permits specification and testing of a wide array of factor models including models with patterns of loadings nearly identical to those in rotated EFA solutions (Morin, Chapter 27).

Although each of these generalizations of basic statistical models is impressive in its own right, it is the integration of the two that constitutes the core strength of SEM. The traditional approach to integrating multiple regression analysis and factor analysis involves factoring a set of indicators of one or more predictors and outcomes, generating factor scores or creating unit-weighted composites of the highest-loading indicators, then using these variables as predictors or outcomes. SEM allows for these two components of the integrated analytic strategy to be achieved simultaneously; that is, the relations between indicators and latent variables and the relations between latent variables are examined in a single model.

This integration of regression analysis and factor analysis is illustrated in Figure 1.1. The model is one in which an outcome, Y, is regressed on a predictor, X. Y is operationally defined by three observed variables, y_1, y_2, and y_3, and X by four observed variables, x_1, x_2, x_3, and x_4. The observed variables, presumed to be fallible indicators of the latent variables, could be survey items, total scores on different instruments designed to measure X and Y, behavioral observations, or physical characteristics. Regardless of how the values on the indicators were obtained, it is assumed that x_1 to x_4 share in common their reflection of construct X but not Y and, conversely, y_1 to y_3 reflect construct Y but not X (i.e., there are no cross-loadings). In order to estimate the effect of X on Y using regression analysis, composite scores would need to be produced, perhaps by summing x_1 to x_4 and y_1 to y_3 or, if the indicators were on different scales, standardizing scores and taking a mean. As illustrated in Figure 1.1, the regression portion of the model involves only latent variables, designated by the larger ovals. These are unobserved forms of X and Y presumed to explain the associations between observed indicators of them, designated by squares. Variance in each indicator is attributable to two unobserved sources: one of the latent variables of interest, X or Y, and uniqueness, or specificity, designated by the small circles. The straight lines indicate directional effects, and the sharply curved lines indicate variances. The asterisks designate parameters to be estimated. These include factor loadings, uniquenesses, a regression coefficient, a disturbance (regression error of prediction), and the variance of X. This approach to depicting a model is called a "path diagram" (see Pek, Davisson, & Hoyle, Chapter 4). Importantly, although a model of this form is prototypical, it is but one of multiple ways latent variables and their interrelations can be modeled.

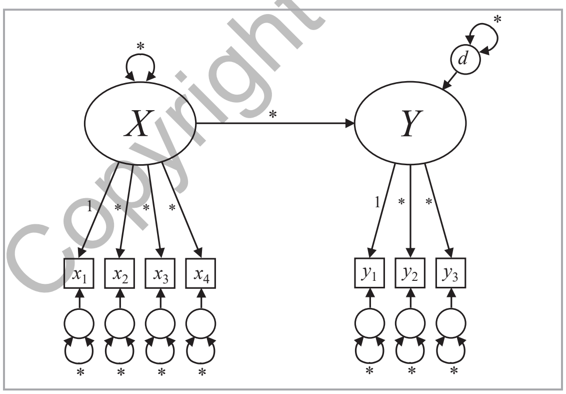

FIGURE 1.1. A prototypical model with measurement and structural components.

In many applications of SEM, the observed variables are assumed to be measured on a continuous scale, and any latent variables are assumed to be continuous as well. Yet variables often are measured coarsely (e.g., 5- or 7-point response scales) and sometimes categorically (e.g., *yes–no*), raising question as to the appropriateness of standard SEM approaches to estimation and testing. Fortunately, SEM accommodates data, models, estimators, and fit statistics for observed and latent categorical variables (Chen, Moustaki, & Zhang, Chapter 8; Koziol, Chapter 15; West, Wu, McNeish, & Savord, Chapter 10).

Although typical applications of SEM focus on relations between variables, in some cases, the hypothesis of interest requires modeling patterns of means or means of latent variables. These applications require moving beyond pure covariance structure modeling to a consideration of models that include a mean structure. This addition allows for the expansion of models such as the one shown in Figure 1.1 to include intercepts in the measurement and structural equations and means of the latent variables. In longitudinal data, it also permits modeling of individual patterns of means over time and their variability in latent growth curves (Grimm & McArdle, Chapter 30). When these variables are examined in relation to latent variables that explain associations among sets of indicators (e.g., X and Y in Figure 1.1), the model includes three components—measurement and structural equations, which, together, constitute the covariance structure, and the mean structure. The full generality and flexibility of SEM would be evident in a model that includes all three components and both continuous and categorical observed and latent variables.

BASIC LOGIC AND APPLICATION

The chapters in Part I of this book cover foundational topics relevant for understanding and effectively using SEM. I offer an overview in this introductory chapter as context for the material covered in those chapters and the basic applications covered in the first few chapters in Part II of the book.

A fundamental difference between SEM and more familiar statistical models such as ANOVA and multiple regression analysis is the target of parameter estimation. In typical applications of multiple regression analysis, for example, the regression coefficients are estimated using ordinary least squares (OLS). The coefficients define a regression line that minimizes the average squared distance between the case-level data points (the target) and the line. Residuals index the degree to which the estimated line misses each data point, that is, the degree of error in predicting the observed data points from those estimated by the model. The goal of estimation in SEM is the same—to find values of the parameters that best account for the observed data given a substantively interesting model. A major difference, however, is what constitutes the observed data, or target. In the prototypical application of SEM—for example, the model shown in Figure 1.1—the data are the variances of and covariances between the observed variables. The goal of estimation, typically by the maximum likelihood method, is to find values for the parameters that, given the model, maximize the likelihood of the observed data. Stated differently, as with OLS regression, the goal is to minimize the difference between the observed and estimated data, but the observed and estimated data in prototypic applications of SEM are variances and covariances. Thus, the residuals are the differences between the observed variances and covariances and those estimated by the model given the data (see Chen et al., Chapter 8, for detailed coverage of estimation in SEM).

Returning to the model depicted in Figure 1.1, the data are the seven variances of the observed variables plus the 21 covariances between them (easily calculated as $p(p + 1)/2$, where p is the number of observed variables). As with the case-level observed data in OLS regression, the degrees of freedom available for model testing are derived from the number of data points—28 in this case. This number is the same regardless of sample size. As with tests involving case-level data, the number of degrees of freedom for a given test is the number of available degrees of freedom, 28 in this instance, minus the number of parameters to be estimated. Referring again to Figure 1.1 and counting asterisks, there are 15 parameters to be estimated, leaving 13 degrees of freedom for tests of model fit. A model that fits the data well implies covariances that are close in magnitude to the observed covariances (the implied and observed variances will be the same, as all variance in observed variables is fully accounted for in the model).

Models such as the one shown in Figure 1.1 are specified by researchers; that is, there is no default model for covariance matrices based on seven observed variables. A given specification offers a putative explanation for the pattern of observed covariances and reflects the researcher's hypotheses about those relations; it also reflects certain technical constraints necessary to ensure the model can be estimated. When the parameters in a model are estimated from data, they can be used in combination with the data to produce an estimated, or implied, covariance matrix equivalent to fitted values on the outcome variable in OLS regression. The difference between the implied and observed matrices is the residual matrix, which is implicated directly or indirectly in various tests and indices of fit. Generally speaking, a model fits the data when the elements of the residual matrix are uniformly near zero. Models initially specified by researchers often result in one or more residual covariances that differ from zero, meaning they are not adequately explained by the model given the data. In such cases, models often are respecified, estimated, and tested, the equivalent of post hoc comparisons in ANOVA. When support is obtained for either an a priori or respecified model, it is compared against plausible alternative models, interpreted, and presented. Each of these steps in the basic application of SEM are discussed and illustrated in Part I of the book; considerations specific to particular models are presented in Parts II and III. In the next section of the chapter, I present a framework that integrates the general steps involved in the implementation of SEM.

SEM IMPLEMENTATION FRAMEWORK

Despite its flexibility and generality, in practice, SEM is nearly always implemented following a series of discrete steps. In this section, I present an SEM implementation framework that positions these steps in relation to each other, providing context for the foundational topics and applications presented in the remainder of the book. For each step I provide an overview and refer to the relevant chapters. The framework, shown in diagram form in Figure 1.2, comprises four steps—specification, estimation, evaluation of fit, and interpretation and reporting—that are always followed. Because they are important considerations for how the steps are implemented, I also include the related concerns of data acquisition/preparation, identification, respecification, and model selection; these are shown in Figure 1.2 as boxes connected by dashed lines to one or more of the primary steps in implementation.

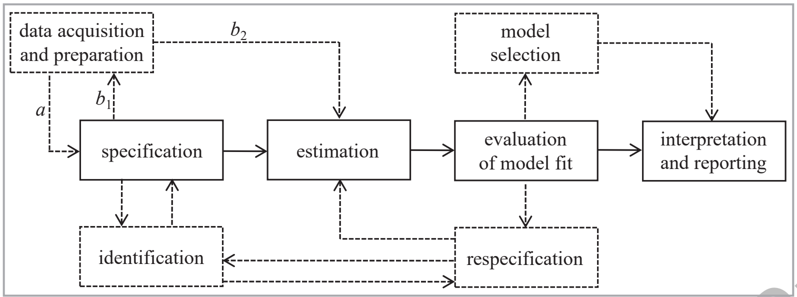

FIGURE 1.2. Steps in the implementation of SEM.

SEM can be used with different intents, and it is useful to review them here as context for the presentation of the implementation framework. Specifically, Jöreskog (1993) described three common intents when using SEM. Although somewhat rare in practice, SEM can be used with strictly confirmatory intent. In such cases, a single a priori model is specified and evaluated. The model either provides an acceptable account of the data or it does not. No attempts are made at modifying the model or comparing it with alternative models. Alternatively, the researcher's intent may include both evaluating the fit of a model in an absolute sense and comparing it with alternative models that reflect competing theoretical accounts or offer a more parsimonious account of the data. When SEM is used with the intent of comparing alternative models, these models should be specified a priori and, when possible, specified in such a way that direct statistical comparisons can be made. Finally, the intent of an SEM analysis might be the generation of a model for subsequent evaluation in a strictly confirmatory or alternative models analysis. Although an initial model must be specified, that model might originate from results of prior analyses (e.g., multiple regression analysis, EFAs) or from SEM analyses of an a priori model that offers a sufficiently poor account of the data that it must either be modified or abandoned. Many uses of SEM begin with strictly confirmatory or alternative model comparison intent, but they become exercises in model generation when a priori models do not meet fit criteria. At the other extreme, it is possible to begin with a commitment to no particular model and use data mining strategies to generate models (see Brandmaier & Jacobucci, Chapter 39). With these distinctions in mind, I now turn to an overview of the implementation framework displayed in Figure 1.2.

Specification

The typical use of SEM always begins with the specification of a model. A "model" is a formal statement of the mechanisms assumed to have given rise to the observed data. Those mechanisms reflect the substantive hypotheses that motivated the analysis, as well as characteristics of the sample and research design. As discussed later in this section, the model also includes features that ensure that unique values can be obtained for the parameters to be estimated (see Pek et al., Chapter 4, for detailed coverage of specification).

As shown in Figure 1.2, specification can take place either before or after data are acquired and prepared for analysis. The dashed line labeled a corresponds to the situation in which specification follows data collection, whereas the line labeled b_1 corresponds to the situation in which data collection follows specification then, as indicated by line b_2, directly precedes estimation. Again, using the model depicted in Figure 1.1 as an example, a researcher might have access to a set of data that includes x_1 to x_4 and y_1 to y_3. These may be data the researcher collected but did not collect with this specific model in mind, or data acquired from a secondary source (e.g., U.S. Census data). Note that in this situation the options for specification are constrained by the contents of a set of data that were not collected with the researcher's model of interest in mind. In such cases, multiple indicators might not be available, precluding the specification of latent variables, the spacing of longitudinal data might not be ideal for the mechanisms being modeled, or in any number of other ways the data might limit the researcher's ability to specify the model that ideally would be tested. For this reason, the preferred approach is the acquisition of data that allows for the estimation and testing of a model that comprises all that the researcher believes relevant to the process or structure of interest. Referring again to Figure 1.1, a model, when specified before the data are acquired, serves as a guide to data collection or the selection of a secondary data source. Only a data set that includes x_1 to x_4 and y_1 to y_3 would be suitable for the analysis.

The specific actions and concerns in specification are the same whether a model is specified before or after the acquisition and preparation of data. In terms of actions, specification involves designating the variables, relations among the variables, and the status of the parameters in a model. With regard to designating variables, the decisions are which observed variables to include and which latent variables, if any, to model (see Bollen & Hoyle, Chapter 5, for a detailed treatment of latent variables). Having decided which observed and latent variables to include in the model, the researcher must then decide which variables are related and, for those that are related, whether the relation is nondirectional or directional. Finally, the status of parameters in a model must be specified. In general, a parameter can be specified as either fixed or free. Fixed parameters are those whose values are set by the researcher and, therefore, not estimated. For instance, in the model shown in Figure 1.1, the loading of x_1 on X is fixed to 1. Less apparent is the fact that the loadings of x_1 to x_4 on Y and the loadings of y_1 to y_3 on X are fixed to 0; otherwise, the loadings are free parameters and will be estimated from the data (see Pek et al., Chapter 4, for additional detail on fixed and free parameters).

A specified model is expressed formally using a system of notation in either a set of equations or a diagram. Historically, each computer program for conducting SEM analyses accepted only one means of depicting a model. For example, early versions of the LISREL program required specification using matrix notation (see Pek et al., Chapter 4). Early versions of the EQS program required equations and double-label notation. The first program designed specifically for use on desktop computers, Amos, accepted either line by line code or path diagrams constructed using the program's built-in drawing capability. These and other programs such as Mplus and the `lavaan` package in R now allow for model specification using multiple means, as well as program-specific shorthand coding schemes. Model specification in Mplus and `lavaan`, the programs used for nearly all of the examples presented in this volume, is described and illustrated in Chapter 13 (Geiser).

A key concern in specification is identification (see Kenny & Milan, 2012, for a detailed treatment). Each parameter in a specified model must be identified and, if all parameters are identified, the model is said to be an identified model. A parameter is identified when it takes on a single value given the model and observed data. Parameters can be identified in two ways. The most straightforward and direct means of identifying a parameter is to fix its value. Because a fixed parameter can, by definition, assume no other value, it is identified. Free parameters are identified if there is but one estimated value for them that satisfies the estimation criterion and are obtained when the data are used to solve relevant structural and measurement equations. In some models, there is more than one way to obtain the estimate for a free parameter from these equations. As long as all such computations produce the same estimate, the parameter is overidentified. If a single value for a given parameter cannot be obtained through estimation, the parameter is unidentified and, as a result, the model is unidentified. Although a few straightforward rules of thumb offer some assurance that a model is identified, the only way to ensure identification is to show mathematically that a single value can be obtained for each parameter in all ways it might be expressed as a function of other parameters in the model. As illustrated in Figure 1.2, identification is linked to (re)specification by dotted lines. This designation is not to suggest that identification is optional. Rather, it indicates that it is possible to specify and estimate a model without attending to identification. Ideally, prior to estimation, researchers would verify that all parameters are identified; however, some SEM software includes certain parameter specifications by default that ensure basic identification (e.g., a single loading on each latent variable to establish its metric). All SEM software produces error messages that signal identification issues, though those messages often do not point to the specific unidentified parameter or set of parameters. In such cases, the researcher is forced to attend to identification. It bears noting that not all identification problems are related to specification. Parameter estimates near zero and highly correlated parameters can result in empirical underidentification, which can only be detected by attempting estimation.

An additional concern related to specification is the statistical power of tests of model fit. The model that best reflects the researcher's hypotheses about the mechanisms that gave rise to the data may be perfectly captured in the specification with all parameters identified, but the likelihood of finding support for the model or specific parameters in the model given the specification and data is too low to justify the analysis. The statistical power of SEM analyses is affected by multiple factors (e.g., degrees of freedom, sample size, correlations between parameters) that may vary from one fit index or statistical test to the next. The role of degrees of freedom—which derive, in part, from model specification—in the statistical power of SEM analyses argues for the consideration of statistical power as part of model specification. Detailed treatment of statistical power in the SEM context is provided by Feng and Hancock in Chapter 9.

Estimation

Once a model has been specified, its parameters identified, and the data prepared for analysis, the implementation moves to estimation. The goal of estimation is to find values for the free parameters that minimize the discrepancy between the observed covariance matrix and the implied covariance matrix given the model and the data. The means by which parameter estimates are derived depend on which of a number of possible estimation methods is used. Examples are maximum likelihood, unweighted least squares, generalized least squares, weighted least squares, and asymptotically distribution free estimators (see Chapter 8 by Chen et al., for detailed coverage of estimation and estimation methods). By far the most commonly used method of estimation is maximum likelihood, the default in most SEM computer programs. Because the validity of model evaluation rests most fundamentally on the integrity of estimates, a critical concern for researchers is whether maximum likelihood estimation is appropriate given their data and model. If it is not, then a decision must be made as to which alternative estimator overcomes the limitations of maximum likelihood without introducing additional concerns about the integrity of estimates. The key assumptions and how they are evaluated are discussed in Chapter 7 (Kline). The robustness of different estimators to violations of assumptions often is determined by simulation studies, the logic and interpretation of which are covered in Chapter 6 (Leite, Bandalos, & Shen).

Most estimation methods, including maximum likelihood, are iterative. They begin with a set of start values for the free parameters. These values are, in effect, used along with the fixed parameter values to solve the equations that define the model and produce an implied covariance matrix. The degree of discrepancy between the observed and implied covariance matrices is reflected in the value of the fitting function, the computation of which varies from one estimator to the next. The goal of estimation is, through iterative updating of parameter estimates (beginning with the start values), to minimize the value of the fitting function, which takes on a value of zero when the observed and implied covariance matrices are identical. Because the start values are not based on a consideration of the data given the model, the initial estimates typically result in substantial discrepancy between the observed and implied covariance matrices reflected in a relatively large value of the fitting function. The first few iterations typically result in substantial reductions in the discrepancy between the two matrices and corresponding declines in the value of the fitting function. When the value of the fitting function cannot be minimized further through updates to the parameter estimates, the process is said to have converged on a solution. Often convergence is achieved in 10 or fewer iterations, though complex models or estimation situations in which start values are highly discrepant from the final estimates may require more. Unidentified models and models estimated from ill-conditioned data typically do not converge, requiring the researcher to revisit the model specification or data evaluation and preparation. Although convergence is necessary for evaluation of fit, the number of iterations required for convergence has no relevance for that evaluation.

Evaluation of Fit

Although a set of parameter estimates obtained from suitable data for an identified model are those estimates that minimize the discrepancy between the observed and implied covariance matrices, that discrepancy may be relatively large or small; that is, the fixed and estimated parameters may imply a covariance matrix that is sufficiently similar to the observed covariance matrix to support an inference that the model fits the data, or it may imply a covariance matrix in which one or more values are sufficiently discrepant from the observed data that an inference of fit is not warranted. In an SEM analysis, the evaluation of fit concerns whether the specified model offers an acceptable account of the data or should be rejected (if the intent is strictly confirmatory) or respecified (if the original or reconsidered intent is model generation). How this evaluation is done and a decision reached remains a topic of research and debate among methodologists (for a review and recommendations, see West et al., Chapter 10).

A useful starting point for considering how decisions about fit are made is a value based on the value of the fitting function and sample size, which is assumed to follow a central χ^2 distribution. In reality, the value typically labeled χ^2 is an approximation that, under conditions typical of SEM analyses, is a poor approximation. Moreover, the statistical test, when it is legitimate, is of a hypothesis that few researchers would venture: that the specified model fully accounts for the observed data (i.e., there is no discrepancy between the observed and implied covariance matrices; see Preacher & Yaremych, Chapter 11, for discussion of the limited value of this hypothesis). Nonetheless, it is prototypical of goodness-of-fit tests, the goal of which is to inform inferences about the correspondence between the observed data and the data implied by a model.

Relatively early in the history of SEM, the χ^2 goodness-of-fit test fell into disfavor as a test of the absolute fit of a specified model. The earliest alternatives were indices that reflected the improvement of a specified model over a model that assumed no relations between the variables (i.e., the independence, or null, model), with some taking into account model complexity. In some cases, these values were standardized so that their values ranged from 0 to 1, with higher values indicating greater improvement of the specified model over a model that offered no account of the relations between variables. A drawback to these comparative fit indices is that because they do not follow a known probability distribution, they cannot be used to construct formal statistical tests. As such, their use is governed by rules of thumb, typically involving the designation of a criterion value that must be exceeded for a model to be considered acceptable (see West et al., Chapter 10, for a discussion of the challenges associated with setting criterion values for these indices).

Because of the critical importance of the decision to reject or accept and interpret a specified model and the absence of a number that can be used for unambiguous inferences for all data and modeling circumstances, the development of new fit statistics and indices continues. The most promising of these follow a known probability distribution, focus on absolute rather than comparative fit, evaluate the hypothesis of approximate rather than perfect fit, and account for the complexity of the model. In Chapter 10, West and colleagues review a wide range of fit statistics and indices and offer recommendations for using them to judge the adequacy of a specified model. For alternative model applications, Preacher and Yaremych (Chapter 11) discuss the use of fit information to select from among a set of alternative models.

Beyond the evaluation of overall model fit, and typically only when overall fit is deemed acceptable, are tests of the magnitude of the estimated parameters. These typically are tested for difference from zero using a test that is comparable to the test of coefficients in multiple regression analysis (i.e., estimate/standard error). Additional tests focused on parameters might consider whether two or more estimates are equivalent, as in evaluations of measurement invariance (see Gonzalez, Valente, Cheong, & MacKinnon, Chapter 22) or follow a pattern of theoretical interest as in latent curve analyses (see Grimm & McArdle, Chapter 30).

Respecification

As shown in Figure 1.2, the evaluation of fit may be followed by one of three next steps in the SEM implementation process. If the intent of the analysis is, to use Jöreskog's (1993) descriptor, strictly confirmatory, then the next step is interpretation and reporting. If the evaluation of fit indicates that an a priori model does not offer an acceptable account of the data, the research may engage in model generation by respecifying the model to improve fit based on an examination of the residual matrix or software-supplied modification indices. If, rather than considering a single model, the researcher wishes to consider several alternative models, then he or she must compare models in order to select the one to be interpreted and reported. The larger and more complex a specified model, the greater the likelihood of misspecification and, therefore, the greater the likelihood that respecification will be necessary to attain the values of fit indices generally required for interpretation and reporting.

Decisions about how a model might be respecified to improve its fit are based on specification searches, the goal of which is to find sources of misspecification among the fixed and free parameters in the initially specified model. Specification searches can be manual, which involves a visual inspection of the residual matrix in search of subjectively large residuals, or automated, which involves the use of a statistical algorithm that evaluates the incremental improvement in fit if each fixed parameter is freed (e.g., Lagrange multiplier test) or free parameter is fixed (e.g., Wald test). Note that respecification requires a reconsideration of identification then a return to estimation and evaluation of fit. Furthermore, in addition to concerns about whether, given sample sizes typical of research that uses SEM analyses, specification searches find modifications that would replicate in another sample from the same population (MacCallum, Roznowski, & Necowitz, 1992) are concerns about the validity of critical values of indices and test statistics, which are not adjusted for the researcher degrees of freedom associated with respecification (Wicherts et al., 2016). Transparent and complete reporting of all analyses coupled with careful interpretation of results is critical when the selected model was not among the models posited before analyses began.

Model Selection

Only in the strictly confirmatory application of SEM is a single model evaluated, moving the implementation directly to interpretation and reporting. In typical applications, more than one model is evaluated either by design or out of necessity when a single model specified a priori is not consistent with the data. Multiple models put forward prior to analyzing the data may represent alternative theoretical accounts of the structure or mechanism under investigation or alternative models that differ primarily in terms of complexity. The models may be nested such that one is specified by fixing or freeing parameters in the other, or they may be nonnested. In either case, the models can be compared formally using various indices of fit and decision criteria (see Preacher & Yaremych, Chapter 11, for detailed coverage of model selection). Some alternative models of interest are equivalent; their estimation results in identical fit information (for a review, see Williams, 2012). Because such models cannot be differentiated on statistical grounds, the choice of one of the alternatives requires conceptual justification based on deep understanding of the focal structure or mechanism.

The goal of model selection is to move to the final step of implementation with a single model that will be interpreted, then disseminated in a research report (see Figure 1.2). The need to move beyond evaluations of fit for several models to the selection of one model requires more than a superficial understanding of fit criteria and features of models that contribute to misspecification and unacceptable fit. Such considerations are particularly important when more than one candidate model meets fit criteria. In these cases, simple decision rules based on statistical criteria may not lead to the selection of the model that offers the best balance of parsimony, generalizability, and informativeness with respect to the structure or mechanism under investigation. The selection is particularly challenging for competing but equivalent models, for which the use of statistical criteria is not an option. In all instances of model comparison and selection, considerations beyond those related to the concepts and relations between them such as research design and sample size are relevant. In short, model selection may require a consideration of statistical, design, and conceptual information in order to select from among a set of alternative models given a set of data.

Interpretation and Reporting

When a model has been selected, attention turns to the final step in the implementation framework. Given the technical challenges associated with specification, estimation, and evaluation of fit (including model comparisons), it is perhaps surprising that many of the criticisms leveled at SEM have focused on the interpretation and reporting of results. For that reason, the researcher who uses SEM must take special care at this final stage of the SEM implementation process.

With respect to interpretation, the primary concerns are the basis for the model, the meaning of particular parameters in the model, and the degree to which the model is unique in accounting for the observed data. Generally speaking, the basis for the model can either be a priori, as in models that reflect theoretical accounts or form a set of interrelated hypotheses that perhaps derive from multiple theories, or post hoc, as in models that include modifications to the initially specified model or have their basis in exploratory analyses of the same data to which they were fit. The former affords more confident inferences and allows for more straightforward interpretation based primarily on the concepts and their interrelations. The latter requires qualifying with reference to the means by which the model was derived or modified.

A second interpretational issue concerns the meaning of certain parameters in the model. Specifically, I refer to parameters associated with directional paths and the degree to which they can be interpreted as reflecting causal effects. In this regard, the prevailing wisdom among methodologists has moved from a willingness to view tests of parameters as tests of causal effects in the 1960s and 1970s to an increasing reluctance to interpret parameters in this way beginning in the

1980s and continuing into the early 2000s. As detailed in Chapter 3 (Pearl), there is evidence of a move away from such conservative interpretation of directional effects to a view that, when properly justified, parameters can be interpreted as tests of causal effects even when the design is cross-sectional and the data are correlational.

Finally, an issue that has received too little attention from researchers who use SEM, despite repeated expressions of concern by methodologists (e.g., Breckler, 1990; MacCallum, Wegener, Uchino, & Fabrigar, 1993), is the degree to which the model accepted by the researcher is the only model that offers an acceptable account of the data. As discussed in the previous section, this may include nested or non-nested models that meet fit criteria or equivalent models, for which values of fit indices are identical. With respect to equivalent models, a particular concern is models that include paths that directly contradict those in the preferred model (see Pek & Hoyle, 2016, for discussion of the problem of equivalent models for tests of mediation in cross-sectional data). The degree to which the researcher can successfully manage these interpretational concerns influences the credibility, impact, and reproducibility of their application of SEM.

Beyond these interpretational concerns is a more mundane set of concerns that focus on what is to be included in research reports describing SEM analyses and results. Given the flexibility of SEM and the multiple approaches to estimation and evaluation of fit, the research report must include information that generally is not expected in reports of ANOVA, multiple regression, or factor analysis. At the most basic level, the reader needs full information regarding the model specification, including the full array of fixed and free parameters and an accounting for degrees of freedom. Additional information includes the estimation method used and the outcome of evaluating its assumptions, the information to be consulted in order to evaluate fit, and the specific criteria that will distinguish a model that offers an acceptable account of the data from one that does not. Information about missing data, if any, and how it was managed in the analysis is important, particularly given the fact that some approaches to managing missing data affect model specification (e.g., inclusion of auxiliary variables; see Enders, Chapter 12, for information about methods for addressing missing data in SEM analyses). Once this background information has been provided, the researcher must decide what statistical information from an SEM analysis to report and how to report it. Best practices in reporting SEM results are outlined and illustrated in a number of published papers (e.g., Hoyle & Isherwood, 2013; McDonald & Ho, 2002; Raykov, Tomer, & Nesselroade, 1991; Schreiber, Stage, King, Nora, & Barlow, 2006).

This general framework captures the primary steps in any implementation of SEM, regardless of the type of model or data under study. In the final major section of the chapter, I describe the various types of models and the types of data for which they would be appropriate. Instances of each type are discussed in detail and illustrated in Parts II and III of this book.

TYPES OF MODELS

A covariance matrix to be modeled using SEM, especially a large matrix, affords a wide array of modeling possibilities, constrained only by features of the sampling strategy, the research design, and the hypotheses or patterns the researcher is willing to entertain. In fact, an infinite number of models is possible with even a few observed variables (e.g., Raykov & Marcoulides, 2001). Of course, not all models that might be specified and estimated are plausible or interesting. The point is that SEM allows for the specification and testing of a wide array of models using a single comprehensive and integrative statistical approach. In the remainder of this section, I describe a sample of the models for which SEM is well suited; references are provided to relevant chapters in the book. Although these models do not sort cleanly into a small number of categories, for efficiency, I present them in relatively homogeneous groups based on the type of data and hypotheses for which they are appropriate.

Models Primarily Focused on Latent Structure

The variables implicated in many research questions cannot be directly observed in pure form, if at all. Rather, they must be inferred from fallible indicators, such as administrative records, observer ratings, self-reports, or the status of some biological characteristic, such as heart rate or changes in blood volume in selected regions of the brain. A means of separating variance in these indicators attributable to the variable of interest from variance attributable to other factors is to gather data on multiple indicators that share in common only their reflection of the unobserved variable of interest. This latent variable is assumed to be a relatively pure reflection of the variable of interest, free of the error and idiosyncrasies of the individual indicators (though not free of other sources of variance common to all indicators; see Bollen & Hoyle, Chapter 5, for further details and discussion of other types of latent variables). This notion of "common variance as latent variable" is familiar to many researchers as the basic premise of EFA. In the SEM context, it is the basic logic and building block for a large number of models.

The most straightforward model concerned primarily with the latent structure of a set of indicators in the first-order factor model with reflective indicators. The two factors in the model depicted in Figure 1.1 are first-order factors assumed to account for the covariances among the seven indicators. Unlike in EFA, indicators typically are assigned a priori to factors and, in many cases, each indicator is assumed to reflect only one factor (but see Morin, Chapter 27, on measurement models in exploratory SEM). This prototypical model can be used to test a wide array of hypotheses, such as whether the factors are correlated and, if so, whether they are distinguishable; whether each item is, in fact, a reflection of only one factor; whether the loadings are equal; and whether subsets of the uniquenesses are correlated. The basic first-order model and extension of it are discussed in Chapter 14 (Brown). Considerations with respect to the number of indicators per factor and methods for reducing that number when it is large (e.g., a questionnaire with many items) are covered in Chapter 28 (Marcoulides, Yuan, & Deng) and Chapter 16 (Sterba & Rights), respectively.

If the model includes enough first-order factors, the researcher might choose to explore the latent structure of the first-order factors. In the same way that the common variance among indicators can be attributed to a smaller number of latent variables, it is possible that the common variance among first-order factors can be attributed to a smaller number of second-order factors. The classic example is Thurstone's use of EFA to argue for the presence of seven primary (i.e., first-order) mental abilities but later to concede that a single (i.e., second-order) unobserved thread, presumably general intelligence, ran through them (Ruzgis, 1994). With enough first-order factors, it is possible to have multiple second-order factors.

Another class of models concerned primarily with the latent structure of a set of indicators comprises models with subfactors, which are additional first-order factors that explain commonality in subsets of indicators that may span one or more broader first-order factors of interest (e.g., Hoyle & Lennox, 1991; Reise, Mansolf, & Haviland, Chapter 18). In such models, some or all indicators are directly influenced by two first-order factors. For example, returning to Figure 1.1, imagine that x_2, x_4, and y_2 were negatively worded and for that reason shared a source of variance not captured by *X* and *Y*. In order to account for this common variance, a subfactor, *Z*, could be specified that influences x_2, x_4, and y_2 despite the fact that they span *X* and *Y*. The inclusion of subfactors can be used strategically to tease apart trait and method variance, as in multitrait–multimethod models (Eid, Koch, & Geiser, Chapter 19), or trait and state variance, as in trait–state models (Cole & Liu, Chapter 33). These models, as well as first- and higher-order models, can be estimated for indicators that are continuous or categorical. The specific concerns of measurement models that include categorical indicators are discussed in Chapter 15 (Koziol).

Regardless of the specific model of latent structure, the question of whether a single model applies to all members of a given population may be of interest. (The same question may be asked of any model, regardless of type.) There are two approaches to studying model equivalence. When the subpopulations for which the model is to be compared can be distinguished by an observed variable (e.g., gender, ethnicity), then multigroup modeling may be used (Sörbom, 1974). In multigroup modeling, a model is estimated separately for different groups subject to constraints placed on individual parameters or groups of parameters. For instance, the loadings in a factor model might be constrained to be equal across groups and compared to a model in which they are free to vary as a means of evaluating the equivalence of the loadings. This approach is described and illustrated by Widaman and Olivera-Aguilar (Chapter 20). It is also possible that a given model does not describe the data for all members of the population but the variable that defines homogeneous subgroups in terms of parameter values is not observed. In such cases, factor mixture modeling can be used to estimate a categorical latent variable that indexes subgroup membership (Lubke & Muthén, 2005; for general coverage of mixture models, see Steinley, Chapter 29).

Models Primarily Focused on Directional Effects

A second type of model is concerned primarily with the estimation of the directional relations between variables, which may be latent or observed. The most basic model of this type is equivalent to the multiple regression model, in which the relations between a set of potentially correlated predictor variables and a single outcome are estimated. In this simplest structural model, all variables are observed and there are no directional relations between the predictor variables. SEM extends this basic model in three primary ways: (1) Any of the variables may be observed or latent (with the use of factor scores in SEM as presented in Chapter 17, by Devlieger & Rosseel, offering a hybrid option), (2) there may be multiple outcomes among which there are directional relations, and (3) there may be directional relations between predictors. The first extension is illustrated in our example model, in which latent variable *X* predicts latent variable *Y*. The second and third extensions are somewhat redundant as instances of models in which variables are both predictor and outcome. In fact, it is possible to have a model in which one of many variables is only a predictor and all other variables serve as predictors with reference to some variables in the model and outcomes with reference to others. Additional coverage of the distinction between predictor-only and predictor-and-outcome or outcome-only variables—exogenous and endogenous variables, respectively—is provided in Chapter 4 (Pek et al.).

This distinction is evident in a relatively straightforward but highly useful model: the model that includes an indirect, or mediated, effect. Imagine that we add a variable, *Z*, to the model depicted in Figure 1.1. This variable is presumed to mediate the effect of *X* on *Y*. To evaluate this hypothesis, *Z* is positioned between *X* and *Y* with a directional path running from *X* to it and from it to *Y*. Thus, *Z* is both an outcome and a predictor. This particular model, the topic of Chapter 22 (Gonzalez et al.), has received considerable attention from methodologists and is widely used in some research literatures.

Discussions of statistical mediation often compare and contrast it with statistical moderation—the qualification of a direct effect by another variable. Moderation is tested by interaction or product terms, which are routinely included in ANOVAs, less frequently considered in multiple regression analyses, and rarely included in models analyzed using SEM. In part, the relative neglect of interaction terms in SEM analyses may be attributed to the complexity of specifying interactions involving latent variables. Recent developments in modeling latent interactions have resulted in approaches that significantly reduce the complexity of specification and estimation while expanding the forms of interaction effects that can be modeled. These strategies are reviewed and demonstrated in Chapter 23 (Kelava & Brandt). The inclusion of dynamic moderation effects in longitudinal models is covered in Chapter 24 (Zyphur & Ozkok). Interaction effects receive additional coverage in Chapter 37 (Harring & Zou) as an instance of nonlinear effects.

A particularly useful class of models focused on directional relations is for data on the same sample at multiple points in time. These models can be distinguished in terms of the intensity of assessment or observation. Traditional longitudinal models involve the collection of data at relatively few points in time (typically two to four) at relatively long time intervals (typically 1–6 months). Intensive longitudinal models involve the collection of data at many time points at short time intervals (occasionally even in a continuous stream). The prototypical model for traditional longitudinal data is the autoregressive model, in which each variable is included in the model at each point in time. This permits estimation of the effect of one variable on another from one wave to the next while controlling for stability of the variables from wave to wave (basic coverage is provided in Zyphur & Ozkok, Chapter 24). When the data collection is more intensive, as in the case of many observations over a short period of time, SEM can be used to model dynamic and patterned change as it is observed taking place. Dynamic SEM and continuous-time modeling are covered in Chapter 31 (Hamaker, Asparouhov, & Muthén) and Chapter 32 (Chow, Losardo, Park, & Molenaar), respectively. In Chapter 34, Chen, Song, and Ferrer show how models of dynamic change are extended to the dyadic case.

These longitudinally intensive data, as well as data appropriate for a subset of models described in the next section, are clustered; that is, the individual observations of each individual are almost certainly more related to each other than they are to the individual observations of other individuals in the data set. The same concern applies when each individual observation applies to a different individual, but subsets of individuals share an experience (e.g., treatment by one of several health care professionals) or place in an organization (e.g., one of several classrooms or schools) that is not shared by all individuals in the sample. SEM permits modeling of such clustering while retaining all of the flexibility in modeling described in this section of the chapter. Chapter 26 (Heck & Reid) describes and illustrates the specification, estimation, and testing of these multilevel models using SEM methods.

Models that Include Means

The goal of most models estimated using SEM, including all those described to this point, is to account for covariances between variables. An additional model type, which may be integrated with the models reviewed thus far, focuses on estimating the pattern of observed means or estimating latent means. These models require as input an augmented matrix either derived from raw data or produced by adding a line for means to an observed variance–covariance matrix. Models fit to such matrices add intercepts to the measurement and structural equations, which allows for the modeling and comparison of means of latent variables, as well as attempts to account for, and perhaps predict, the pattern of means between groups or over time. The additional considerations raised by the inclusion of means and hypotheses involving means that can be evaluated using SEM are covered by Thompson and colleagues (Chapter 21).

Particularly useful is a set of models that are longitudinal, multilevel, and focused on modeling means—latent growth models. These models express as latent variables the variability between individuals in the pattern of means over time. For instance, bonding to school might be assessed annually on four occasions beginning with the first year of middle school. These assessments are clustered within individual; thus, the model is multilevel. With four time points, both linear and quadratic patterns could be modeled, yielding three latent growth factors reflecting variances in intercepts and linear and quadratic slopes. In multilevel terms, these factors are Level 2 variables that can be related to other Level 2 (i.e., individual level) latent and observed variables as described in the earlier sections. The basics of this modeling approach and variations on it are described by Grimm and McArdle (Chapter 30).

To further extend a model that already leverages many of the capabilities SEM affords, a researcher might ask whether there is evidence in the data of distinct subsets of individuals who show evidence of a similar pattern of bonding to school scores across the four time points. Although it is possible that the researcher has anticipated and measured the characteristic that defines these subsets, more often the heterogeneity in growth either is unexpected or, if expected, its source unknown. In such cases, growth mixture modeling can be used to model a categorical latent variable that defines subsets of individuals with similar patterns of bonding to school scores. This latent variable is not unlike the latent variables discussed thus far, except that its interpretation is not as simple as inferring the source of commonality among its indicators. Rather, it can be correlated with or predicted by other variables, latent or observed, to examine potential explanations for membership in these emergent groups defined by different patterns of bonding to school. Growth mixture modeling combines features of mixture modeling (Steinley et al., Chapter 29), latent growth modeling (Grimm & McArdle, Chapter 30), and latent class analysis (Lanza & Rhoades, 2013) to model common patterns of change attributable to unobserved sources of between-group differences. These models can be expanded to include predictors of membership in the emergent classes or compare classes on other observed or latent variables.

These different model types can be adapted to a wide array of data and analytic situations. For instance, SEM is increasingly used to model genetic data (Bruins, Franić, Dolan, Borsboom, & Boomsma, Chapter 35). A relatively new application is for modeling meta-analytic data (Cheung, Chapter 36). And, across an array of data types, SEM has proven useful as an integrative approach to measurement scale development and validation (Raykov, Chapter 25). Across all these data and model types, parameters can be estimated and models selected using Bayesian methods, which are now available in widely used SEM computer programs. An introduction and demonstration of the Bayesian approach to SEM analyses is provided by Depaoli, Kaplan, and Winter (Chapter 38).

CONCLUSION

SEM is a comprehensive and flexible approach to modeling patterns and mechanisms of theoretical interest in a wide array of data types. Historically used primarily to model covariances between variables measured on continuous scales, the capabilities of SEM have expanded dramatically to allow for modeling of many data types using an array of estimation methods and to accommodate means, patterns of means, latent interaction terms, nonlinear relations, categorical latent variables, clustered data, and models tailored to the needs of researchers working with complex data historically not analyzed using sophisticated multivariate methods. Though SEM is not necessary, or even desirable, for every hypothesis test or modeling need, it is unrivaled in its capacity to fulfill many varied multivariate hypotheses and model types. How this capacity is harnessed and used to full advantage is the topic of the 38 chapters that follow.

REFERENCES

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

Hoyle, R. H., & Isherwood, J. C. (2013). Reporting results from structural equation modeling analyses in *Archives of Scientific Psychology. Archives of Scientific Psychology, 1*, 14–22.

Hoyle, R. H., & Lennox, R. D. (1991). Latent structure of self-monitoring. *Multivariate Behavioral Research, 26*, 511–540.

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

Kenny, D. A., & Milan, S. (2012). Identification: A nontechnical discussion of a technical issue. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 145–162). New York: Guilford Press.

Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*, 157–168.

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

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

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

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

Pek, J., & Hoyle, R. H. (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass, 10*, 150–163.

Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling, 8*, 142–149.

Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in *Psychology and Aging*: Some proposed guidelines. *Psychology and Aging, 6*, 499–503.

Ruzgis, P. (1994). Thurstone, L. L. (1887–1955). In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 1081–1084). New York: Macmillan.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research, 99*, 323–338.

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.

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis* (pp. 136–178). Chicago: University of Chicago Press.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., Robbie, C. M., van Aert, R. C. M., et al. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid *p*-hacking. *Frontiers in Psychology, 7*, Article 1832.

Williams, L. J. (2012). Equivalent models: Concepts, problems, alternatives. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 247–260). New York: Guilford Press.