CHAPTER 19

PATH ANALYSIS AND STRUCTURAL EQUATION MODELING WITH LATENT VARIABLES

Rick H. Hoyle

The focus of this chapter is a family of statistical methods and strategies collectively referred to as *structural equation modeling* (SEM), of which path analysis is a special case. These range from the relatively straightforward and familiar to the complex and new. My goal in the current treatment is threefold and reflected in the structure of the presentation: In the first section of the chapter, I develop a context for understanding the origins of SEM by tracing its emergence and positioning it within the constellation of statistical methods that are familiar to many psychological scientists. In the second section, which constitutes the core of the chapter, I describe and illustrate the steps involved in using SEM. In the third and final section, I present an array of prototypic models that illustrate the range of structures and processes that could be modeled using SEM. A firm grasp of material presented in the chapter will prepare readers to understand most published reports of SEM analysis in psychological science and position them to make an informed decision about whether their own research agenda could benefit from using SEM.

BACKGROUND AND CONTEXT

SEM is a growing family of multivariate statistical methods for modeling data. SEM is substantially more flexible than statistical methods that have dominated data analysis in psychological science since the early 20th century. It allows for multiple independent and dependent variables, which may be observed or implied by the pattern of associations among observed variables. Directional relations, as in analysis of variance (ANOVA) and multiple regression analysis, can be modeled *between* independent variables and *between* dependent variables. Complex models of the latent structure underlying a set of observed variables (i.e., unmeasured sources of influence and their interrelations) can be evaluated. These models can be estimated from continuous or ordered categorical data and include correlations between variables, direct and indirect effects, and focus on both relations between variables and patterns of means across group or over time. The flexibility of SEM permits modeling of data in ways that are not possible with other, more commonly used, statistical methods. For this reason, SEM both facilitates tests of hypotheses not adequately tested by other methods and suggests hypotheses that might not otherwise be considered for lack of a framework within which to venture them.

Historical Context

The roots of contemporary SEM can be traced to the earliest form of path analysis, Wright's (1934) method of path coefficients. Wright, a geneticist, developed path analysis for the purpose of modeling the relative influence of heredity and environment on the color of guinea pigs (Wright, 1920). These influences were assumed to be causal, and Wright

While writing this chapter, the author was supported by National Institute on Drug Abuse (NIDA) Grant P30 DA023026. Its contents are solely the responsibility of the author and do not necessarily represent the official views of NIDA.

DOI: 10.1037/13620-019
APA Handbook of Research Methods in Psychology: Vol. 2. Quantitative, Qualitative, Neuropsychological, and Biological, H. Cooper (Editor-in-Chief)

referred to his use of path analysis as a means of testing causal effects when "more direct attempts" (i.e., randomized experiments) were not feasible. The unfortunate result of this co-occurrence of non-experimental design and causal inference is that from the outset, path analysis and SEM have been associated with causal inference (Denis & Legerski, 2006). Although such inferences could be defended for many of Wright's models, given their focus on genetic influences, such is not the case for the lion's share of models that are tested using path analysis and SEM in the social and behavioral sciences.

Wright also is credited with the invention of the path diagram, the widely used graphic means of representing models. His (and therefore *the*) first path diagram was a stylized depiction of the concurrent influences of the genetic contribution of a sire and a dam, environmental factors, and chance on the color of guinea pig offspring (Wright, 1920). From this stylized depiction, Wright developed a more formal and general diagram that could be used to depict the full array of relations between variables in a "system of causes and effects" (p. 330). Wright's *system* corresponds to the contemporary notion of a model. An important feature of Wright's diagram was the inclusion of path coefficients (a term coined by Wright) on paths indicating the magnitude and direction of statistical relations in the system. In a description that captures well the activity of SEM as elaborated later in the chapter, Wright (1920) characterized path analysis as the activity of "expressing the known correlations in terms of unknown path coefficients" (p. 330).

Although there is evidence that social and behavioral scientists were aware of Wright's innovation relatively soon after his seminal publications (e.g., Burks, 1928), it was not until the 1960s that a wave of interest began to build. The stimulus was an important book by Blalock (1964), *Causal Inferences in Nonexperimental Research*, which seemed to underscore the inference that researchers drew from Wright's work—that path analysis/SEM could be used to test causal hypotheses using data from correlational research (a misinterpretation of both Wright's and Blalock's writing). Sociologists extended the use of path analysis to longitudinal models and, though the methodology for including latent variables (i.e., unmeasured sources of influence) had not yet been fully developed, sociologists indicated an awareness of the importance of accounting for unmeasured variables (Tomer, 2003). It was a sociologist who published the first article in a psychology journal highlighting the potential of path analysis/SEM for psychological research (Duncan, 1969). By 1970, there was evidence that path analysis had begun to find traction in psychological science (Werts & Linn, 1970). Preceding and overlapping developments by methodologists in sociology were important developments by econometricians. Perhaps the most important of these concerned the method by which parameters (e.g., regression coefficients) were estimated. As early as the 1940s, it became evident that ordinary least squares was inadequate for estimating parameters in multiequations systems and that maximum likelihood could be used effectively (e.g., Mann & Wald, 1943). Goldberger and Duncan (1973) integrated the sociological approach to path analysis with the simultaneous equations approach in economics and the factor analytic approach in psychology (e.g., Duncan, 1975; Goldberger, 1971), yielding the generalization of path analysis now known as SEM (described by Bentler, 1986b, as "the literal grafting of a factor analytic model upon a simultaneous equation model," p. 41). This general model was formalized and extended in the 1970s by Jöreskog (1973), Keesling (1972), and Wiley (1973), producing what became known as the LISREL (*LI*near *S*tructural *REL*ations) model.

Bentler (1986b) credited the "spread from the methodology laboratory to the research laboratory with unusual rapidity" to the fact that SEM allows researchers "to effectively study substantive problems that could not easily be investigated using alternative approaches" (p. 35). The effectiveness of SEM was introduced to psychologists primarily by Bentler and colleagues in an early set of publications that used SEM to evaluate complex multivariate hypotheses in a unified and efficient matter (e.g., Bentler & Speckart, 1979; Huba & Bentler, 1982). These and other compelling demonstrations (e.g., Maruyama & McGarvey, 1980) coupled with the growing accessibility of LISREL, the primary computer program for implementing SEM, and the introduction of Bentler's (1985) EQS fueled the early growth in the use of SEM in psychological science.

By the late 1980s developments in SEM methods and reports of substantive research using SEM were appearing with increasing frequency. During the period from 1987 to 1994, the overall number of such publications increased from 80 to 185. During the period, the number of SEM methods-focused articles remained steady, whereas the number of substantive articles increased nearly threefold (Tremblay & Gardner, 1996). This pattern of growth continued during the period from 1994 to 2001, with the number of substantive publications almost doubling and the number of different psychology journals in which a report of research using SEM appeared increasing as well (Hershberger, 2003). Importantly, across the period of these reviews, substantive publications reporting results from analyses using other multivariate methods (e.g., multivariate analysis of variance, factor analysis, cluster analysis) remained steady or declined (Hershberger, 2003; Tremblay & Gardner, 1996). Within a relatively short period of time, SEM has moved from relative obscurity and use by a small number of methodologically minded researchers to its current status as well-known multivariate method used by researchers across the spectrum of psychological science.

Statistical Context

An effective way to begin developing an understanding of SEM is to compare and contrast it with more familiar statistical methods, each of which can be viewed as a special case of SEM. An overly simplistic, but useful, view of SEM is as a hybrid of multiple regression analysis and factor analysis. Some readers will be familiar with the two-step strategy of using factor analysis to determine the latent influences in a set of variables, and then using factor scores or unit-weighted composites to focus the data analysis on the latent influences. In SEM, these two steps are accomplished simultaneously in such a way that actual scores for the latent influences are not needed. Instead, those influences in the form of latent variables (i.e., factors) are estimated from the data when they are included as predictors or outcomes in a set of regression-like equations. Importantly, however, unlike multiple regression analysis, for which outcomes are addressed one at a time, multiple, possibly latent, outcomes can be included in a single model. Moreover, predictive relations between the outcomes can be modeled if there is reason to do so.

With this general idea in mind, we can see how SEM is a generalization of a host of narrower, more familiar, statistical methods. For example, the *t* test is a special case of ANOVA, which is, in turn, a special case of multiple regression analysis. As noted, multiple regression analysis is a special case of SEM. Focusing on the latent variable component of SEM, covariances (unstandardized zero-order correlations) also are a special case of SEM, though they also are the building blocks for factor analysis. SEM can be used to model categorical latent variables (i.e., latent classes), which are an extension of methods based on contingency tables as well as latent class and latent transition analysis (e.g., Kaplan, 2008; Marsh, Lüdtke, Trautwein, & Morin, 2009). To this set of capabilities can be added the modeling of patterns of means as in trend analysis in ANOVA. And, when the means are from repeated assessments of a sample of individuals, these patterns of means can be treated as predictors or outcomes in multilevel models (e.g., Curran, 2000). The end result for SEM is a very general model that includes components of many narrower statistical models with which psychological scientists are familiar but, bringing them together in a single framework, allows for models and hypothesis tests not possible using those narrower models.

BASIC CONCEPTS

An initial understanding of a number of features of SEM will set the stage for a somewhat more detailed description of the steps involved in using SEM. In each of the short sections that follow, I juxtapose a feature of SEM against a comparable feature typical of statistical methods commonly used by psychological scientists. The goal is not to show that one is superior to the other, but rather to highlight the features of SEM that are likely to be unfamiliar to most readers.

Modeling Versus Analyzing Data

Psychological scientists are accustomed to analyzing data. By *analyze*, I mean test specific differences (e.g., *t*-test, ANOVA) and coefficients (e.g., correlation, regression), typically against zero, using tailored methods that involve relatively little concern for the full array of influences apparent in the data. Modeling data, on the other hand, involves accounting for features of the research design and substantive influences that explain the pattern of relations across a set of variables (for a fuller treatment of this distinction, see Rodgers, 2010). Whereas the outcome of analysis typically is evidence for or against a posited difference or coefficient, the outcome of modeling is a statement about the correspondence between a system of relations between variables specified by the researcher and a set of observed data on those variables. Psychological scientists occasionally engage in modeling, as in factor analysis, in which the goal is to discover a plausible model to explain the relations between a set of variables, or hierarchical multiple regression analysis, in which the goal is to incrementally build a model of the relations that produce variance in single outcome; however, psychological scientists, in the main, are data analysts. As is evident from its name, SEM is a strategy and set of methods for modeling. Thus, its use requires a shift in thinking about hypothesis testing.

Covariances Versus Raw Scores as Data

Whether analyzing or modeling data, the goal of most statistical methods is to account for variability in observed data. The degree to which the difference or relation tested by the model does not account for the data typically is termed error; thus, the goal of data analysis or modeling might be conceptualized as an exercise in minimizing error. SEM differs from methods to which psychological scientists are accustomed in its definition of error. In least squares methods such as multiple regression analysis, the goal is to find the set of coefficients that minimize the difference between each individual's observed value on the outcome variable and the value predicted for them by the regression line on the basis of their scores on the predictor variables. The focus, then, is on the correspondence between observed and predicted case-level data. In SEM, the data of interest typically are the observed variances and covariances of the variables (although means are sometimes of interest). The adequacy of a model is judged by the correspondence between the observed variances and covariances and those predicted by the model. For this reason, SEM is sometimes referred to as covariance structure analysis. The interpretation and use of SEM requires a shift in focus from accounting for individual scores to accounting for covariances.

Specification of a Model Versus Running an Analysis

Most of the data-analytic strategies with which psychological scientists are familiar are, to use a computer analogy, plug and play. All that is required to run an analysis is choosing the variables, in some cases indicating which are independent and which are dependent variables, and identifying them as such in the computer program of choice. Because of the narrow and tailored nature of the methods, relatively few, if any, decisions need to be made beyond which variables to include and where to include them. The execution of an SEM analysis is substantially more involved. This is, in part, a by-product of the high degree of flexibility afforded by the method. In SEM, there is no default model. The selection and designation of variables is just the first in a series of steps involved in specifying a model. As will become apparent as you work your way through the chapter, any number of models might be specified for a set of variables. Indeed, for a large number of variables, the number of models that might be specified is extremely large. The important point is that computer programs for running SEM analyses are not "plug and play." Rather, the researcher must make a potentially large number of decisions about which variables in a set are related to each other and how. Collectively, these decisions are referred to as model specification. Psychological scientists occasionally engage in specification (without using the label), as in choosing how many factors to extract and how to rotate them in a factor analysis, or deciding how to group variables and the order in which groups are entered in hierarchical multiple regression analysis. In SEM, all analyses require specification.

All Parameters Versus a Select Few

In commonly used methods such as ANOVA and multiple regression analysis, the researcher typically is provided with, or elects to attend to, only a subset of the parameters that are estimated. This is because many of the parameters involved in the analysis are neither under the control of, nor typically of interest to, the researcher. Thus, for example, the variances of predictors and the covariances between them in multiple regression analysis are rarely seen when that analysis is run, except during diagnostics for colinearity. The uniquenesses, variance in variables not accounted for by the factors in a factor analysis are not routinely provided in computer output. In SEM, every parameter in a model is "in play." In fact, a decision must be made about how every parameter will be handled in the estimation of the model. Although this requirement adds to the work required before a model can be estimated using SEM, it suggests a potentially large number of hypotheses that might be tested but routinely are not formally considered.

Goodness of Fit Versus Difference From Zero

As noted, the typical focus of data analysis in psychological science is the question of whether a particular difference or coefficient differs from zero. Although these differences and coefficients may be thought of as continuous variables that vary in magnitude, it remains the case that, in many quarters of the discipline, they are categorical variables that can take on two values—significant or nonsignificant. Furthermore, although some statistical methods allow for testing of a set of relations, as in tests of R^2 in multiple regression analysis, most focus on individual relations (e.g., main effects and interactions in ANOVA). In SEM, the primary focus is a system of relations as specified in a model. Tests of specific coefficients are consulted only after the question of whether the model provides an acceptable account of the data has been addressed. By "acceptable account," I mean the degree to which the covariances predicted by the researcher's specified model mirror the observed covariances. When the two *sets* of covariances are statistically equivalent, the model *fits* the data. Thus, a key difference between methods to which psychological scientists are accustomed and SEM, is that SEM focuses on the collective adequacy of the system of relations in a model. The magnitude of specific coefficients within the model, although related to the adequacy of the model as a whole, typically is of secondary interest.

Latent Versus Observed Variables

A hallmark of SEM is the ability to specify relations between latent variables, or factors. Latent variables are unobserved sources of influence that typically are inferred from the pattern of relations between a set of observed variables, referred to as indicators. Although, as noted, relations between latent variables sometimes are approximated by a piecemeal strategy using factor or principal components analysis followed by ANOVA or multiple regression analysis using factor scores, this combination is implemented seamlessly in SEM. As illustrated in the final section of the chapter, the ability to model and test hypotheses about latent variables is the basis for a number of rigorous and sophisticated strategies for decomposing variance in observed variables. The most straightforward benefit, however, is the ability to estimate relations between variables from which unreliability has been removed. This approach not only results in larger coefficients indexing relations between variable but also, because of the dependency between coefficients in a model, it sometimes results in smaller coefficients (e.g., as in mediation models). In both cases, the coefficients are assumed to better approximate the true relation between constructs than coefficients between fallible observed variables.

STEPS IN THE USE OF STRUCTURAL EQUATION MODELING

Although many types of models can be evaluated using SEM, the steps involved in applying SEM are virtually always the same. These steps are used to implement SEM in the service of one of three goals (Jöreskog, 1993). In a *strictly confirmatory* use of SEM, the goal is to evaluate the degree to which a single, a priori model accounts for a set of observed relations. Alternatively, instead of focusing on a single model, SEM might be used to compare two or more competing models in an *alternative models* strategy. Finally, a use of SEM might have a *model-generating* focus. If, for example, an a priori model does not adequately account for the observed data, rather than abandoning the data, the researcher might use it to generate an explanatory model (McArdle, in press). Of course, using the data to generate a model of the data is inferentially risky (MacCallum, Roznowski, & Necowitz, 1992); however, careful modification of a poor-fitting a priori model with the goal of finding a model that accounts for the data can lead to discoveries that, if replicated, increase understanding of a psychological structure or process. Regardless of the goal, an application of SEM follows an ordered set of steps that begins with specification and concludes with interpretation.

I illustrate those steps with an empirical example. The data are from two waves of a longitudinal study of problem behavior among middle school students (Harrington, Giles, Hoyle, Feeney, & Yungbluth, 2001). A total of 1,655 students from 14 schools participated in the study. Students first completed the self-report survey by early during the academic year when they were from 11 to 13 years old. They completed the survey a second time near the end of the academic year, about 8 months later. (An intervention and a third wave of data detract from the illustrative benefits of the data set and are therefore ignored for this example.) The analysis data set includes 10 observed variables measured at two occasions and assumed to reflect three latent variables. The focal outcome is *problem behavior*, for which indicators are drug use (composite of the number of days out of the past 30 that each of a set of illicit substance were used), sexual activity (ranging from affectionate physical contact to sexual intercourse), and interpersonal aggression (ranging from teasing to physical fighting). One predictor is *risk*, indicated by scores on three individual differences: impulsive decision making, sensation seeking, and (low) self-esteem. The other predictor, *protection*, is reflected by four composite scores on scales designed to tap values and lifestyle variables assumed to be incompatible with problem behavior. These variables include bonding to school, a personal and public commitment to avoid problem behavior, an assumption that prevailing norms are not to engage in problem behavior, and the view that problem behavior interferes with a productive and otherwise desirable lifestyle. Data from all students on these 10 indicators were available at both time points. Additional information about the data set, including the management of missing data, can be found in the published report (Harrington et al., 2001).

MODEL SPECIFICATION

All applications of SEM begin with the specification of a model. Model specification involves three sets of decisions that stem from questions about a model:

1. Which variables will be included and in what form (observed or latent)?
2. Which variables will be interrelated and how?
3. Which parameters (i.e., coefficients) will be estimated and which will be set to specific values?

Before detailing how these questions are addressed in SEM, let us consider how they are addressed in familiar methods such as ANOVA, multiple regression, and factor analysis. Although each of these statistical methods, when applied to research questions, involve specification, most aspects of specification in these are established by convention and imposed by computer programs used to analyze data. The exception would be the first question, which typically is addressed directly by the researcher without consulting the data, although some data-reduction and data-mining methods involve consulting the data to decide which variables are to be included in a model. A decision about variables that is key in SEM is whether latent variables will be included and, if so, how many. In factor analysis, the decision to extract factors is a decision to include latent variables, and the decision about how many to extract is a decision about how many to include. As with data-reduction and data-mining methods, this decision about including latent variables typically is made on the basis of patterns in the data. In SEM, the decision about whether to include latent variables and how many can be made before data have been collected and certainly should be made before the data have been consulted.

The second question involved in model specification is which variables will be interrelated and how they will be interrelated. Again, restricting our discussion to methods familiar to psychological scientists, these decisions often are part and parcel of using a particular statistical method. For instance, a 2 × 2 factorial design yields 3 degrees of freedom for specifying a model that will capture the pattern of means. In the default specification, one degree of freedom is used for each main effect, and one is used for the interaction. The remaining $(n - 1) - 3$ degrees of freedom is the divisor in the error term. Even with this simple design, other specifications are possible. These alternative specifications would require defining sets of contrasts that differ from those implied by the standard main-effects-and-interaction specification. With more complex designs, the number of sets of such contrasts is large. In factor analysis, the decision to use an orthogonal rotation method is a decision to not allow factors to correlate with each other; the use of an oblique rotation allows those correlations. A final component, one that is particularly salient in SEM, is the direction of influence between variables that are directionally related. In analyses of data from randomized experiments, this decision is straightforward: Influence runs from the manipulated variables to variables assessed in the experimental context after the manipulation. In data from quasi- or nonexperimental studies, the decision is not straightforward. In its simplest form, it involves deciding, for instance, whether a particular variable will be positioned on the predictor or outcome side of a multiple regression equation. Such decisions constitute model specification within the rather restrictive data-analytic frameworks of methods such as ANOVA and multiple regression analysis.

The final question is which parameters in the model will be estimated and which will be set to a specific value and, therefore, not estimated. This aspect of specification is less apparent than identifying variables and their interrelations, but it is no less important in defining models. Examples of parameters that are estimated include regression coefficients and factor loadings. Parameters that are set are less obvious. For example, in exploratory factor analysis, the uniqueness components of indicators are independent; that is, the covariances between them are set to zero. As noted, an orthogonal rotation specifies uncorrelated factors, which is equivalent to setting the covariances between the factors to zero. There is relatively little room to fix parameters in the standard statistical methods used by psychological scientists, but every parameter in a model to be estimated and tested using SEM can be set to a particular value, or fixed, given certain mathematical constraints.

Before discussing and illustrating model specification in SEM, it is worth considering what constitutes a *model* in SEM. This consideration is particularly important when drawing inferences from the results from estimating and testing a model in SEM. One interpretation of a model is as a literal reflection of reality as captured by the data. Of course, "reality" is remarkably complex, impossible to fully grasp, and, on some counts, simply uninteresting, rendering this idea of a model less than appealing. An alternative interpretation of a model is as an account anchored in reality, but one that includes only those aspects relevant to a research question (although accounting for salient features of research design). This view of a model assumes that the models of interest to psychological scientists will never fully account for a set of data. Rather, the models will provide a parsimonious account of relations evident in a set of data with reference to an interesting and potentially useful set of constructs and theoretical relations. Consistent with this idea, Box (1979) famously quipped, "All models are wrong, but some are useful" (p. 201). In other words, the goal of model specification is the identification of a model that is testable and useful, even if it fails to account for all aspects of the reality that produced the data. Pearl (2000) offered a definition that will guide our thinking in the remainder of the chapter. A model is "an idealized representation of a reality that highlights some aspects and ignores others" (p. 202).

Now, let us focus specifically and concretely on model specification in SEM by returning to the example introduced at the beginning of this section. Recall that a large sample of middle-school students provided information on 10 observed variables at each of two occasions. All of the variables are relevant to the research question, which concerns the role of certain risk and protective factors in adolescent problem behavior. Although the research question might focus on the contribution of each of the three risk and four protective factors to each of the three problem behaviors, the focus is broader: To what extent do risk and protection evident in certain dispositional, perceptual, and attitudinal constructs, contribute to problem behavior, broadly defined? To address this question, I need three additional variables that reflect the general constructs of risk, protection, and problem behavior. In fact, because the variables are assessed at two points in time, I need to include these constructs twice. Thus, in terms of identifying variables, I will include 26 variables—10 observed variables and three latent variables observed on two occasions.

Having designated the variables to be included in the model, I now need to specify the relations between variables. I will provide additional detail later in this section, but for now I can describe the set of relations in general terms. I intend to explain the commonality among impulsive decision making, sensation seeking, and low self-esteem with a latent variable I label risk. A second latent variable, protection, will account for the commonality among bonding to school, commitment to avoid problem behavior, belief that problem behavior is not the norm, and belief that problem behavior interferes with success. And I will model the commonality among drug use, sexual activity, and interpersonal aggression as a third latent variable, problem behavior. These relations and any correlations between latent variables constitute a subcomponent of the model referred to as the *measurement model*. As outlined in Table 19.1, the measurement component of the model links measured indicators to latent variables and includes three types of parameters: loadings, variances of latent variables, and unexplained variance in indicators, or uniquenesses. Each of these parameters must be designated as fixed (i.e., set at a specific value) or free (i.e., to be estimated from the data). I outline considerations involved in these designations in the next section.

In the model of interest, the measurement model is not an end in itself. Although I need to establish empirically that the observed variables reflect the latent variables in the pattern I have proposed, my ultimate goal is estimating the directional relations between the latent variables. Because this is a longitudinal model, I am particularly interested in the directional relation between each variable at Time 1 and the other variables at Time 2. These paths and the correlations between the latent variables at Time 1 and unexplained variance in the latent variables at Time 2 constitute the *structural model*. The parameters included in the structural model are listed and described in the lower portion of Table 19.1. These include all directional paths, variances of predictor variables, unexplained variance in outcomes, and

TABLE 19.1

Parameters to Be Considered When Specifying a Model

Parameter	Description
Measurement component	
Loading	Coefficient resulting from the regression of an indicator on a latent variable
Variance, latent variable	Variance of a latent variable
Variance, uniqueness	Variance in the indicator not attributable to the latent variables on which it loads
Structural component	
Variance, independent variable	Variance of a measured or latent variables that is not predicted by other variables in the model
Variance, disturbance	Unaccounted for variance in a measured or latent variable that is predicted by other variables in the model
Nondirectional path	Covariance between two variables
Directional path	The amount of change in a dependent variable attributable to an independent variable controlling for other relations in the model

covariances between predictors and unexplained variance in outcomes. As with the measurement model, each of these parameters must be designated as fixed or free.

The specification of a model for estimation requires considerably more detail than offered in the written description that I have provided. For that reason, different strategies and conventions have emerged to formalize the aspects of model specification that I have presented. I describe three strategies, using components of the example model to illustrate.

Path Diagrams

One means of formalizing the specification of a model to be analyzed using SEM is the path diagram. A path diagram of a portion of the example model is shown in Figure 19.1. This example includes all of the components necessary to fully specify all but a few highly specialized models. First, note the squares, labeled using Bentler–Weeks notation (Bentler & Weeks, 1980) as v_1 to v_{10}. These represent observed variables, in this case, the 10 variables described earlier as assessed at one occasion. Next, notice the three large ovals, labeled *F1* to *F3*, which include the names of the constructs on which the model is focused. These ovals designate latent variables, or factors. Smaller circles indicate variances of two types. Attached to each indicator, and labeled e_1 to e_{10}, are uniquenesses. These, too, are latent influences on the indicators. The uniquenesses reflect both variance attributable to random processes and variance attributed to processes specific to that indicator. A lone small circle, labeled d_3, is attached to the problem behavior latent variable. Typically referred to as a *disturbance*, this latent influence reflects all processes that contribute to variance in *F3* that is not attributable to *F1* and *F2*. The straight lines indicate directional influence, either among latent variables and indicators or between independent and dependent latent variables. The curved arrow between *F1* and *F2* indicates a nondirectional path, or covariance. The sharply curved, double-headed arrows pointing to the uniquenesses, the disturbance, and the independent latent variables indicate variances. Finally, notice that associated with each path is either an asterisk or a number. The asterisks indicate free parameters, whose values will be estimated from the data. The numbers indicate fixed parameters, whose value does not change as a function of the data.

Some readers might be surprised to see paths in the measurement component of the model running from the latent variables to their indicators. This specification corresponds to the common factor model, which assumes that the commonality among a set of variables is attributable to one or unmeasured influences. These influences are assumed to account for a portion of the variance in the indicators. In SEM terms, indicators related to latent variables in this way are referred to as reflective indicators; they are fallible reflections of the underlying variable of interest. Virtually all latent variables in psychological science are related to their

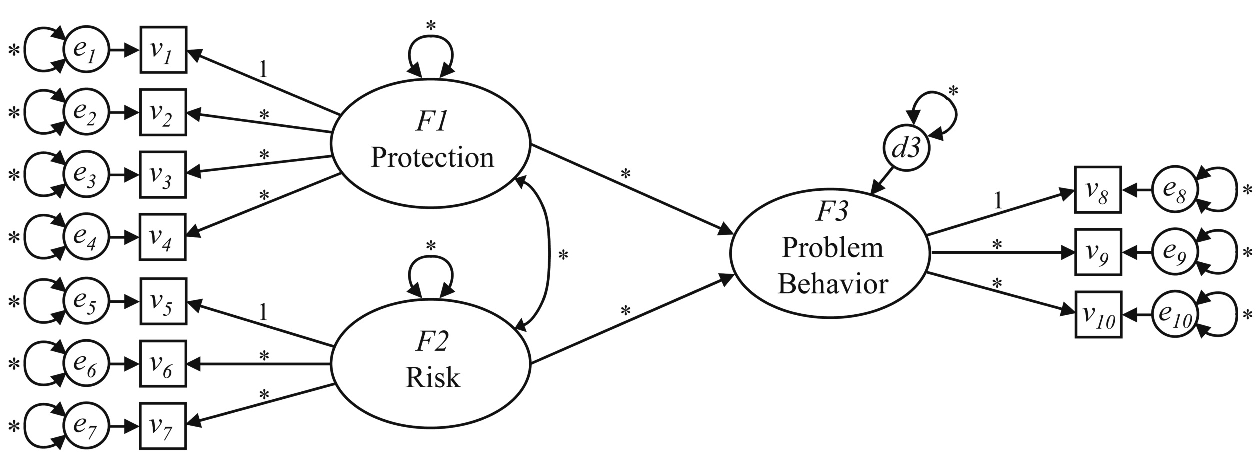

FIGURE 19.1. Path diagram illustrating the components of path diagrams using a portion of the empirical example.

indicators in this way. Nonetheless, it bears mention that an alternative specification reverses the direction of the arrows so that the latent variables are assumed to be caused by its indicators. In this case, the observed variables are referred to as formative indicators. In that alternative specification each measured indicator is assumed to represent a unique component that, when added to the other components, yields the latent variable. Such models pose significant estimation challenges and this, coupled with their rarity in psychological research, justifies my focus solely on reflective indicators in the remainder of the chapter (for an informative discussion of this distinction, see Edwards & Bagozzi, 2000).

Although it is evident from the path diagram that a loading on each latent variable has, for reasons detailed in the section on estimation, been fixed to one, it is not evident that many other parameters have been fixed. There are no paths between uniquenesses, implying that the covariances between them have been fixed to zero. Each indicator is specified to load on only one latent variable, meaning that its loadings on the other two latent variables have been fixed to zero. Thus, although the path diagram is an appealing means of presenting a model, a significant weakness in terms of model specification is that parameters fixed at zero are generally not shown. Of course, the paths I have identified could be added to the model with coefficients of zero associated with them; however, such a diagram would be too cluttered (e.g., imagine a path running from *F3* to v_1) to effectively communicate the basic features of the model.

Equations

An alternative means of specifying a model is through equations. A straightforward means of doing so is based on the work of Bentler and Weeks (1980) and implemented in the EQS computer program for SEM analyses (Bentler, 1989). In addition to using regression-like equations to specify directional paths in a model, the approach uses an intuitive labeling scheme, a portion of which is used to label variables in the path diagram presented earlier. As shown in the path diagram, observed variables are labeled *v*, and all latent variables are labeled *F*. Uniquenesses are labeled *e*, and disturbances *d*. In addition to identifying variables in a model, these labels can be used to identify nondirectional paths and specific parameters using double-label notation. For instance, the loading of v_1 on *F1* would be labeled, *F1*,v_1. The effect of *F1* on *F3* would be labeled, *F1,F3*. Variances are denoted by double reference to the variable in question. Thus, the variance of *F1* is labeled *F1,F1*, the variance of d_3 as d_3,d_3, and so on. Finally, as in Figure 19.1, free parameters are denoted by asterisks.

Using these labels and a series of simple equations, I can specify the model shown in Figure 19.1. The relations between the indicators and their latent variables are characterized in a series of measurement equations:

$$\begin{aligned} v_1 &= 1F1 + e_1 \\ v_2 &= {*}F1 + e_2 \\ v_3 &= {*}F1 + e_3 \\ v_4 &= {*}F1 + e_4 \\ v_5 &= 1F2 + e_5 \\ v_6 &= {*}F2 + e_6 \\ v_7 &= {*}F2 + e_7 \\ v_8 &= 1F3 + e_8 \\ v_9 &= {*}F3 + e_9 \\ v_{10} &= {*}F3 + e_{10}. \end{aligned} \tag{1}$$

With the exception of d_3, which is part of the structural component of the model, all of the variables in the path diagram are represented in these equations. Missing from these equations are the variances for the 10 uniquenesses and the two factors that are predicting the third latent variable.

$$\begin{aligned} &e_1, e_1 = * \\ &. \\ &. \\ &. \\ &e_{10}, e_{10} = * \\ &F1, F1 = * \\ &F2, F2 = *. \end{aligned} \tag{2}$$

The measurement component of the model is completed by specifying the covariance between *F1* and *F2*:

$$F1, F2 = *. \tag{3}$$

APA-COOPER_V2-11-0601-019.indd 342 12/13/11 7:11:04 PM

The remaining paths and parameters in the model constitute the structural model. The paths between latent variables are specified in structural equations:

$$F3 = *F1 + *F2 + d_3. \tag{4}$$

The model is completed by specifying the variance of the disturbance:

$$d_3, d_3 = *. \tag{5}$$

As with the path diagram, free parameters are easily detected in the specification. Some fixed parameters, such as the fixed loading on each latent variable are evident as well. As with the path diagram, however, parameters fixed at zero are not explicitly noted. The covariances between uniquenesses that are fixed at zero would be denoted, for example,

$$e_1, e_2 = 0. \tag{6}$$

Cross loadings fixed at zero (by implication because of their omission from the model) could be included in the measurement equations. For example,

$$v_2 = 1F1 + 0F2 + 0F3 + e_1.$$

Matrix Notation

Another approach to model specification is matrix notation, which is most closely associated with the LISREL computer program (though no longer required for using it). Although it is more tedious and intimidating for those who are new to SEM, model specification using matrix notation has two significant advantages over path diagrams and equations. First, except for certain shortcuts, every parameter, fixed and free, is explicitly shown. Second, matrix notation is the means by which new developments in SEM typically are communicated. As such, an understanding of how models are communicated using matrix notation is essential for keeping abreast of the new developments. Of importance to individuals working on research questions that cross disciplinary boundaries is the fact that matrix notation is used to communicate specifications and findings from substantive research in some disciplines.

Using matrix notation, model specification requires addressing elements in one or more of three matrix equations. A distinction is made within the measurement component between latent variables that are not explained within the model (*F1* and *F2* in our example) and those explained within the model (*F3*). A third matrix equation specifies the relations between latent variables.

Using matrix notation, indicators of latent variables not explained within the model are labeled as *x* and the latent variables as ξ. Loadings are denoted as λ. Thus, the matrix equation for those latent variables is,

$$x = \lambda_x \xi + \delta. \tag{8}$$

Unpacking the equation yields an equation in which specific parameters are made explicit.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & 1 \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \end{bmatrix}. \tag{9}$$

When presented in this way, it is apparent that all of the equations for *F1* and *F2* presented in the previous section are included. (Note that *x*s are the *v*s, the ξs are *F*s, the δs are *e*s, and the λs are *s.) Importantly, it is apparent that, as before, for each indicator, one loading is fixed to zero.

The measurement model for the independent variables is completed by specifying the variances for the latent variables and uniquenesses and any covariances between them. The variances and covariances for the latent variables are specified in the matrix, Φ, which for our model takes the form,

$$\begin{bmatrix} \phi_{11} & \\ \phi_{21} & \phi_{22} \end{bmatrix}. \tag{10}$$

The diagonal elements are variances and the off-diagonal element is the covariances. The variances and covariances for the uniquenesses are specified in Θ_δ.

$$\begin{bmatrix} \delta_{11} & & & & & & \\ 0 & \delta_{22} & & & & & \\ 0 & 0 & \delta_{33} & & & & \\ 0 & 0 & 0 & \delta_{44} & & & \\ 0 & 0 & 0 & 0 & \delta_{55} & & \\ 0 & 0 & 0 & 0 & 0 & \delta_{66} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \delta_{77} \end{bmatrix}. \tag{11}$$

Because uniquenesses typically are not allowed to covary, this matrix often is written as a vector including only the diagonal. For our purposes, the advantage of providing the matrix is that it makes clear that a host of model parameters are fixed at zero in the specification.

For the latent variable explained within the model, *F3*, the matrix equation is,

$$y = \lambda_y \eta + \varepsilon. \quad (12)$$

In expanded form, the equation is,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix} \begin{bmatrix} \eta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}. \quad (13)$$

Note in the path diagram that no variances are associated with *F3* (η_1 in the matrix equation). This is because variance in such latent variables—that is, latent variables that reflect outcomes in a model—is accounted for by paths directed at them and their disturbance terms. Because their variances are not parameters in the model, when there is more than one latent outcome variable in the model their covariances are not parameters in the model either.

Relations between latent variables specified in these matrices are specified in the structural equation,

$$\eta = B\eta + \Gamma\xi + \zeta. \quad (14)$$

For our example, this equation is expanded to reveal individual parameters and their status in the model. (The 0 reflects the fact that variance of η_1 is not estimated. It is partitioned between the paths from ξ_1 and ξ_2 and the disturbance, ζ_1.)

$$\begin{bmatrix} \eta_1 \end{bmatrix} = [0] \begin{bmatrix} \eta_1 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \zeta_1. \quad (15)$$

The only remaining parameter to be specified is the variance of the disturbance (ζ_1 here, but d_3 in the equation and path diagram approaches). Variances of disturbances and any covariances between them are specified in the matrix Ψ, which in the example model, is simply

$$[\Psi_{11}]. \quad (16)$$

Either of these strategies, path diagrams, equations, or matrix notation, allow for specification of a model. Each indicates the observed and latent variables to be included in the model, their relative positions in the model, the relations among them, and the various fixed and free parameters that characterize those relations. Regardless of the method used, once a model has been formally specified, assuming data are available for the observed variables, it can be estimated.

ESTIMATION OF PARAMETERS

The goal of estimation is finding the optimal set of estimates for free parameters in a model given the observed data and the relations specified in the model. By "optimal," I mean a statistical criterion targeted by an estimator. As an example, consider the ordinary least squares estimator, which is typically used to find the optimal, or best-fitting, regression line in multiple regression analysis. The criterion targeted by that estimator is the minimization of the average squared distance between the observed value on the dependent variable for each case and the value predicted for them given the regression line. Although the least squares estimator could be used for SEM, the nature of parameter estimation in SEM is such that least squares is not well suited to finding the parameter estimates best suited to testing the model against the data. A number of alternative estimators are available. These generally fall into one of two categories. Normal theory estimators assume multivariate normality and, typically, continuous measurement. Alternative estimators relax one or both of these assumptions.

Because of the complexity of models in SEM, the degree to which the data satisfy the assumptions of the estimator to be used is of even greater concern than is typical. Of course, data rarely are ideal. Consequently, the relevant question is how robust a given estimator is to violations of its assumptions. For instance, the assumption of continuous measurement is virtually never met in psychological research, in which ordered-categorical response options (e.g., Likert-type scales) are the norm. Fortunately, most normal-theory estimators are reasonably robust to violations of the assumption of continuous measurement, producing acceptable results with measures that include five or more categories (Johnson & Creech, 1983; Wirth & Edwards, 2007). The assumption to which most attention has been directed is multivariate normality. Note that the assumption concerns *multivariate*, not univariate, normality. In other words, the joint distribution of all observed variables in the model must be normal (see DeCarlo, 1997, for a discussion of multivariate kurtosis). Because evaluating multivariate normality, especially when the number of observed variables is large, is not straightforward, evaluation typically focuses on univariate distributions. The practical payoff of focusing on univariate distributions is evident; however, it is important to keep in mind that normal univariate distributions do not guarantee multivariate normality. In either case, normal theory estimators seem to be robust to modest departures from multivariate normality, and relatively straightforward corrections are now available to account for consequential departures from multivariate normality.

The most widely used estimator in SEM is maximum likelihood (ML). In the same way that ordinary least squares is assumed in most uses of multiple regression analysis, ML is assumed in most applications of SEM. It is the default estimator in most SEM computer programs. The goal of ML estimation is to find a set of estimates for the free parameters that, when the data and fixed parameters are taken into account, maximize the likelihood of the data given the specified model (Myung, 2003). ML estimation is an iterative procedure that begins with a somewhat arbitrary set of start values for the free parameters and updates these values until the difference between the observed data and the data implied by the model is minimized. At this point, the estimation procedure is said to have converged, and the adequacy of the resultant model is evaluated.

The use of estimators other than ML usually is motivated by one or more characteristics of the data that do not meet the assumptions of ML estimation. For categorical or coarsely categorized data, robust weighted least squares (Flora & Curran, 2004) and the categorical variable methodology (B. O. Muthén, 1984) are appropriate. For data that are, practically speaking, on continuous scales but nonnormally distributed, asymptotic distribution-free (ADF) methods (Browne, 1984) or postestimation corrections to normal-theory estimates (Satorra & Bentler, 1994) are available. ADF methods, though appealing in theory, typically are impractical given the large sample sizes required for them to evidence theoretical properties. The latter method, typically described as a scaling correction, is particularly promising and is illustrated with our example later in this section of the chapter. Because the validity of model evaluation rests most fundamentally on the integrity of estimates, determined in part by an appropriate estimator choice, a critical concern for researchers is whether ML estimation is appropriate and, if it is not, which alternative estimator overcomes the limitations of ML without introducing additional concerns about the integrity of estimates.

Before returning to the example, I briefly treat an important consideration in estimation: identification. Although identification is evaluated and addressed as part of model specification, I address it here because its consequences are most evident during estimation. In general, identification concerns the degree to which a unique estimate can be obtained for each parameter in a model. Although identification typically is discussed at the global level (i.e., for the model as a whole), if any parameter in a model is not identified (i.e., local identification status), then the model is not identified. Importantly, if it is not identified, attempts at estimation will not meet the criterion inherent in most estimators—minimizing the difference between the observed data and the data implied by the model.

With regard to identification, the status of a model is referred to in one of three ways. An identified model is one in which a single, unique value can be obtained through at least one set of manipulations of other parameters in the model given the data. If the value of one or more parameters can be obtained in more than one way, then the model is *overidentified* and subject to testing for model fit. If the value of each parameter can be obtained in exactly one way, then the model is *just identified*. Although the parameter estimates in just identified models are valid, such models cannot be tested. Finally, a model in which a single, unique value for each parameter cannot be obtained is underidentified, or simply, *unidentified*.

Determining the identification status of a specified model can be a challenge, sometimes requiring the determination of whether the specification meets one or more technical criteria. Those criteria are detailed in comprehensive treatments of SEM (e.g., Bollen, 1989). In some instances, those criteria manifest as relatively straightforward rules of thumb. For instance, for the variance of a latent variable to be defined, it must either be fixed either to a specific value (e.g., 1.0) or to the value of the common portion of one of the indicators, which typically is accomplished by fixing the loading for that indicator to a value of 1.0 (Steiger, 2002). Another rule of thumb concerns the number of indicators per latent variable. If a latent variable is modeled as uncorrelated with other variables in a model (e.g., a one-factor model or a model with orthogonal factors), it must have at least three indicators to be identified. This is because of the general identification rule that a model cannot have more free parameters than the number of nonredundant elements in the observed covariance matrix. For present purposes, it is important only to understand that, although SEM is flexible in terms of the kinds of relations between variables that can be specified, that flexibility can be limited somewhat by technical considerations having to do with how parameters are estimated.

Returning now to the example model and ML estimation, I illustrate the estimation process and describe, in context, additional considerations. When illustrating model specification, to simplify the presentation, I made use of only a portion of the full model to be estimated from the data described earlier. Recall that a relatively large sample of middle school students provided data on 10 variables at two points in time. I hypothesized that the 10 observed variables were indicators of three latent variables of interest at two points in time. My interest is whether variability in a latent variable for problem behavior is prospectively predicted by latent variables for personal characteristics that reflect risk for and protection from problem behavior. Displayed in Figure 19.2 is a path diagram depicting the set of relations in the model. Notice that the observed variables are omitted from the diagram. The reason for their omission is twofold: (a) Their inclusion would yield a figure too cluttered to be of use, and (b) although the measure-

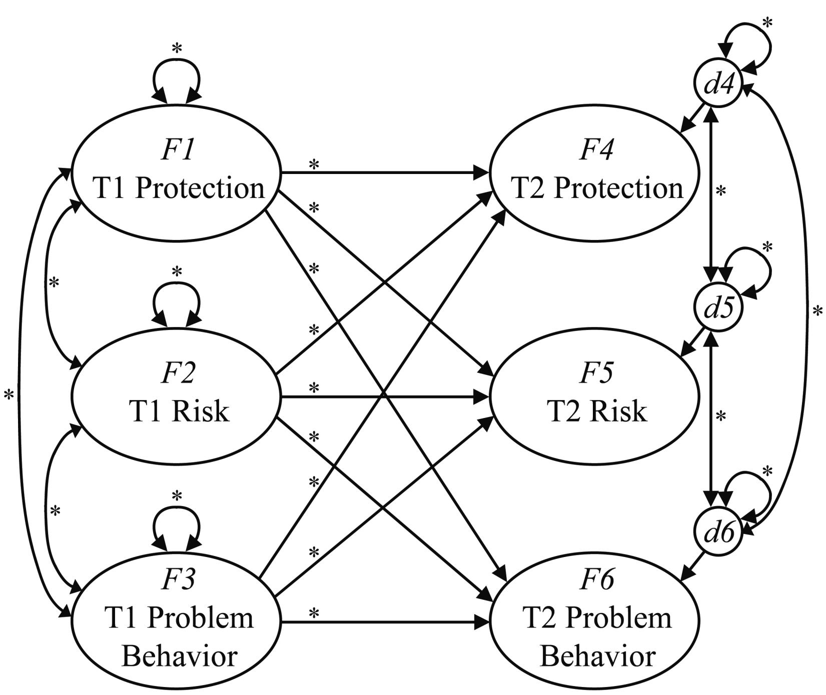

FIGURE 19.2. Structural component of the full example model. T1 and T2 refer to the first and second waves of data collection, respectively. Indicators of the latent variables are not shown.

ment model is a component of the model, for this specific model, it is not the component of greatest interest. The structural component, which specifies the direction relations between the latent variables is of primary substantive interest, and it is this component shown in the figure.

Because the figure shows only the structural component of the model to be estimated, it does not provide full detail regarding model specification. The part of the model not shown concerns the relations and parameters in the measurement component. In our model, these fall into three categories: (a) the relations between indicators and latent variables; (b) uniquenesses associated with indicators; and (c) covariances between uniquenesses. Referring to Figure 19.1, it is apparent that the 10 observed variables at each time point are arrayed so that four reflect protection, three reflect risk, and three reflect problem behavior. It also is evident that a uniqueness is associated with each indicator. What cannot be seen in Figure 19.1, because it does not include repeated assessments of the variables, are nonzero covariances between selected uniquenesses in the model. These are illustrated for indicators of one of the latent variables, protection, in Figure 19.3. Notice the curved arrow and accompanying asterisk between each indicator measured at the two waves of data collection. These autocorrelations are neither substantively interesting nor require justification. They reflect the fact that any unique, nonrandom

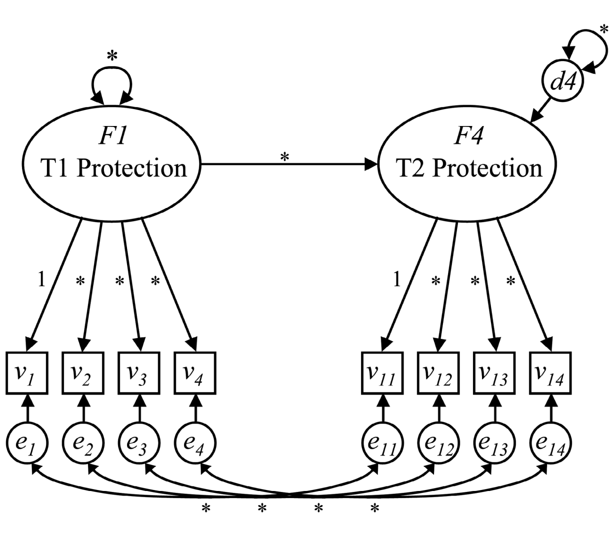

FIGURE 19.3. Portion of the example model illustrated autocorrelated uniquenesses.

variance in an observed variable at one point in time should be evident at other points in time. Imposing Figure 19.3 over the top portion of Figure 19.2 gives a sense of what the full model would look were the path diagram complete.

As noted, the goal of estimation is to obtain values for the free parameters that, when coupled with the fixed parameters, minimize the discrepancy between the observed data and the data implied by the model. Moreover, these values are obtained iteratively, beginning with a guess based on convention and values in the data and culminating in the optimal estimates. The processes of estimation are best understood by dissecting an example.

First, it is useful to pinpoint all of the free parameters to be estimated in a model. In our model there are 65. Their locations in the model are as follows:

- 7 loadings at Time 1 (recall that one is fixed on each latent variable)
- 7 loadings at Time 2
- 10 uniquenesses at Time 1
- 10 uniquenesses at Time 2
- 10 covariances to reflect autocorrelated errors (as in Figure 19.3)
- 3 variances, one for each latent variable, at Time 1
- 3 covariances between the latent variables at Time 1
- 9 paths from Time 1 to Time 2 latent variables
- 3 disturbances at Time 2
- 3 covariances between disturbances at Time 2

In Table 19.2, I provide details from the iteration history only for *F1*, the protection latent variable at Time 1. That part of the model includes nine parameters: four loadings, four uniquenesses, and one latent variable variance. These are shown under the headings “Loadings” and “Variances” in Table 19.2. In the leftmost column is iteration number. The first line is labeled “Start,” indicating that values in this line were those with which iteration began. Although these can be specified by the researcher, they typically are inserted automatically by the computer program used for estimation. Note that, for the loadings and the variance of the latent variable, the default start value is 1.0. For the EQS program, used here, the default for variances of the uniqueness terms is 90% of the variance of the observed variables

TABLE 19.2

Iteration History From Start Values to Convergence (Parameter Estimates for Protection Latent Variable in Example Model)

	Loadings				Variances					
Iteration	v_1	v_2	v_3	v_4	e_1	e_2	e_3	e_4	*F1*	F_{ML}
Start	1.000	1.000	1.000	1.000	.300	.248	.345	.294	1.000	
0	1.000	.674	.734	.706	.235	.185	.282	.231	.329	11.996
1	1.000	.805	.967	.779	.231	.148	.193	.219	.217	6.070
2	1.000	.913	1.191	.831	.216	.129	.130	.216	.177	3.231
3	1.000	.973	1.317	.844	.199	.119	.096	.216	.165	1.783
4	1.000	1.039	1.443	.838	.176	.105	.062	.218	.156	1.227
5	1.000	1.040	1.405	.812	.173	.100	.065	.220	.161	.645
6	1.000	1.055	1.438	.833	.177	.101	.061	.219	.156	.630
7	1.000	1.055	1.433	.829	.176	.100	.062	.220	.157	.630
8	1.000	1.056	1.436	.830	.177	.100	.061	.220	.156	.630

Notes. Tabled values are unstandardized maximum likelihood estimates. F_{ML} = maximum likelihood fitting function. Loading for v_1 fixed to identify variance of *F1*.

(e.g., the .300 for e_1 is 90% of the observed variance of .331 for v_1). Importantly, the value in the Start line for v_1, although it is the same as for v_2 to v_4, is not a start value. As shown in the specifications, this is a fixed value, which is evident because its value does not change from one iteration to the next.

Before discussing the pattern of parameter estimates as the estimator moves from start values to convergence, I draw your attention to the final column in Table 19.2. The values in this column are values of the maximum likelihood fitting function. They basically reflect the difference between the observed data in the form of the covariance matrix and the covariance matrix implied by the model given the data at each iteration. The goal of estimation is to minimize that difference and, as can be seen working from top to bottom in the column, the values decrease with each iteration until they stabilize at a minimum value. Although I have rounded those values to three decimal places, minimization is evaluated at the fifth decimal place (unless the computer program is instructed to do otherwise). When the value of F_{ML} does not change at the first five decimal places from one iteration to the next, iteration stops. At this point, the estimation process has converged on a solution; that is, it has produced the set of parameter estimates that maximize the likelihood of the data given the model.

Let us turn now to the columns for loadings and variances. Notice that there is no value for F_{ML} on the Start line. This is because no value has been generated at that point. The next row contains values for the parameters returned when the fixed and start values are substituted into the equations and values for free parameters produced when taking the data into account. It is at this point that the iterative attempt to minimize F_{ML} begins. Notice that, except for the fixed value of the loading for v_1, all of the parameters have changed from their start values. Relative to other values in the F_{ML} column, the value for the model at this point is very large, indicating a relatively poor fit of the model to the data. After a single iteration, the values are updated to produce a substantial drop in the value of F_{ML}. Although subsequent updates result in smaller values of F_{ML}, the decrease is smaller after each update to the parameter estimates. This pattern is reflected in the parameter estimates as well, which change relatively little after the fourth iteration. Notice the pattern of change in the different parameters. For instance, although the start value for the v_2 loading was relatively close to the optimal estimate, the first iteration moved the estimate farther away rather than closer to that estimate. The value for the loading of v_2 had moved close to its optimal value after only two iterations.

Estimation in this case was smooth. Eight is a relatively small number of iterations, and there is nothing in the iteration history to cause hesitation in moving to evaluation of fit. Such is not always the case. For instance, sometimes estimation fails to converge. A failure to converge is a relative term, because an upper bound ordinarily is imposed on the number of iterations to be attempted. The value for that number is 30, and although it is possible that an attempt at estimation would converge if allowed additional iterations (which can be done by overriding the computer program's default), this typically is not the case. A failure to converge can result from the presence of one or more unidentified parameters in a model or ill-conditioned data. In some cases, a model may be identified in terms of specification, but because a parameter estimate is in effect zero (equivalent to having fixed the path to zero in the specification), it is not. Such models are empirically unidentified. Otherwise, failures to converge often occur in complex models or with data that are not normally distributed. A failure to converge is a signal to reevaluate the model specification and, if no problems are identified there, to reexamine the scaling and distribution of variables in the data set.

Computer Programs

Models typically are estimated by computer programs written specifically for SEM. In the early history of SEM, the only computer program for this purpose was LISREL (Jöreskog, & Sörbom, 2006). Early versions of LISREL (first released in 1970) were challenging to use, requiring at least some knowledge of matrix notation to specify models. Despite these challenges, LISREL became (and, in some quarters, remains) synonymous with SEM. The release of EQS (Bentler, 1985) offered an appealing alternative to LISREL, exchanging matrix notation for a simpler specification approach on the basis of equations similar in form to multiple regression equations. By the early 1990s, statistical computing had begun the transition from mainframe to desktop, and SEM programs followed suit. Early desktop versions of LISREL and EQS were, for the most part, direct translations of the mainframe programs. AMOS (Arbuckle, 1995) was the first fully featured SEM program written to run in the desktop environment, offering the option of model specification by drawing path diagrams. LISREL and EQS have evolved to take full advantage of the desktop environment and, with AMOS, have been joined by other programs for estimating and testing models. A subset of these programs are stand-alone; that is, they do not require another resident program to run. Others are a module within a general purpose statistical analysis program such as IBM SPSS, SAS, or Statistica. In the remainder of this section, I list the alternatives currently available for running SEM analyses and reference relevant websites and publications.

The computer programs one might use for running SEM analyses can be divided into three categories. Perhaps best known are the commercial stand-alone programs, of which LISREL is the prototype. I have mentioned three of these: AMOS, EQS, and LISREL. An increasingly influential addition to this group is M*plus* (L. K. Muthén & Muthén, 1998). Early versions of these programs did not interface well with popular data manipulation programs (e.g., Microsoft Excel) and did not, themselves, offer much in the way of data manipulation, thus requiring the preparation and exporting of data from other programs prior to use. Now, to varying degrees, the programs allow data manipulation and viewing within a spreadsheet format. A critical advantage of these programs is that behind each is a methodologist who contributes regularly to the technical literature on SEM. For that reason, new capabilities are added frequently and, as a result, new versions are released significantly more often than would be typical of statistical software. AMOS (Blunch, 2008; Byrne, 2010) is distributed by SPSS (now IBM SPSS); a basic description is provided at http://www.spss.com/amos/ (see also http://www.spss.com/media/whitepapers/SEMEB.pdf). Although the program runs, and can be purchased, as a stand-alone program (Windows operating system only), it typically is bundled with the SPSS software and often is included in enterprise licenses at education institutions. EQS (Byrne, 2006; Mueller, 1999) is available for multiple computing platforms (detailed information is available at http://www.mvsoft.com/). LISREL (Byrne, 1998; Kelloway, 1998; Mueller, 1999), now in Version 8, includes SIMPLIS, which allows for simple, intuitive specification of models, and PRELIS, which can be used for data manipulation before analysis. Information, including examples, references, and announcements of workshops can be found at http://www.ssicentral.com/lisrel/. M*plus* (Byrne, 2011) is particularly well suited for categorical data and offers the most wide ranging array of modeling capabilities, including multilevel and mixture models. The website for M*plus* (http://www.statmodel.com/), includes extensive information about the program as well as an active discussion list, copies of technical reports, and access to published articles about or using M*plus*. Although not a stand-alone program, I include STREAMS (Structural Equation Modeling Made Simple) in this section, because it features an appealing user interface for manipulating data, specifying models, and presenting results (information at http://www.mwstreams.com/). It does not include the capacity to estimate models, instead using an instance of EQS, LISREL, or M*plus* installed on the same computer.

A second category of SEM computer programs includes those that are integrated into comprehensive commercial programs for quantitative data analysis. A significant strength of these programs is access to the full range of data manipulation and description capabilities offered by the programs. A significant drawback is that the programs typically are slow to incorporate new technical developments compared with the stand-alone commercial programs. Also, documentation, particularly third-party books of the sort available for the stand-alone programs, is sparse. The oldest of the programs in this category are RAMONA, offered by Systat (http://www.systat.com/) and SEPATH, available in Statistica (http://www.statsoft.com/products/statistica-advanced-linear-non-linear-models/itemid/5/#structural). TCALIS is the successor to the CALIS procedure in the powerful SAS System (http://support.sas.com/documentation/cdl/en/statugtcalis/61840/PDF/default/statugtcalis.pdf). Finally, SEM analyses can be done in Stata with GLLAMM (http://www.gllamm.org/; Rabe-Hesketh, Skrondal, & Pickles, 2004).

A final category of computer programs for SEM analyses are no-cost, often open-access, stand-alone programs. As might be expected, these programs do not offer an impressive user interface and provide little or no data manipulation capability. Aside from the fact that they can be obtained at no cost, a subset of these programs allows for modification by users. Perhaps the most widely used from this category is OpenMx (http://openmx.psyc.virginia.edu/), which is particularly well suited to SEM analyses of heredity. The R sem package (http://cran.r-project.org/web/packages/sem/index.html; Fox, 2006) is a good choice for researchers already familiar with data analysis using R. The AFNI *1dSEM* package (http://afni.nimh.nih.gov/sscc/gangc/PathAna.html) is offered as C source code by the National Institute of Mental Health. Finally, SmartPLS (http://www.smartpls.de/forum/), which allows for estimation by the method of partial least squares only, runs on the JAVA platform. These programs make good use of online discussion forums to announce new developments and provide for community-based user support.

Beyond availability, key considerations in choosing a computer program for SEM analyses are as follows:

- What options are available for specifying models (e.g., matrix notation, equations, path diagrams)?
- Does the program have the capabilities needed for the models I will evaluate (e.g., multilevel and multigroup designs)?
- Does the program offer an estimator appropriate for my data (e.g., noncontinuous measurement scales, nonnormal distributions, and missingness)?
- Does the program provide the data manipulation and output control I desire? If not, does it offer straightforward transfer of manipulated data and output to the programs I intend to use for these purposes?

Trial versions (i.e., limited number of variables or cases) of a number of the programs can be obtained, allowing for a firsthand evaluation of their interface and capabilities.

EVALUATION OF FIT

Once a set of optimal parameter estimates have been obtained, the fit of a model can be evaluated.

In theory, this evaluation can be done in a standard null hypothesis statistical testing framework. That is, a null hypothesis can be posed and either be rejected or not using some test statistic. In the case of structural equation modeling, the null hypothesis is,

$$H_0\colon \Sigma = \Sigma(\hat{\Theta}), \quad (17)$$

where Σ is the population covariance matrix, estimated by the observed covariance matrix, and $\Sigma(\hat{\Theta})$ is that matrix written as a function of the model parameter values, or the implied covariance matrix. The goal is to fail to reject this null hypothesis or, in statistical test terms, to find a probability *greater* than .05 associated with the test statistic. The minimized value of the ML fitting function can be used to produce a value that, under certain circumstances, is distributed as a χ^2 statistic with degrees of freedom equal to the difference between the number of nonredundant elements in the covariance matrix and the number of free parameters in the model. In theory, using this value and the appropriate reference distribution, one can conclude that a specified model either fits or does not fit the data.

For two principal reasons, this hypothesis test is, in practice, not useful. Most fundamentally, the conditions necessary for the value on the basis of the minimized value of the fit function to approximate a χ^2 statistic are never met in substantive applications of SEM. These conditions include multivariate normal data, a sample large enough to ensure that the asymptotic properties of the estimator are realized, and a model that is correct in the population. Even if the conditions were met in typical applications, the test suffers from the shortcoming of all null hypothesis statistical tests: Unless the implied and observed covariance matrixes are identical, the null hypothesis can, with enough power, be rejected. Conversely, the implied and observed covariance matrices can be nontrivially different but, if power is sufficiently low, the null hypothesis cannot be rejected. Thus, the statistical test, even if valid on theoretical grounds, is as much a reflection of power as it is of meaningful difference between the implied and observed covariance matrices.

Because of these limitations, alternative means of evaluating model fit have been developed. A prevalent strategy, pioneered by Bentler and Bonett (1980), involves comparing the fit of a specified model against the fit of a model that specifies no relations among the variables (referred to as the null, baseline, or independence model). These comparisons typically yield a value between 0 and 1 that is interpreted as a proportion. The resultant values are indexes, not statistics, and therefore are descriptive rather than inferential. A widely used index is the comparative fit index (CFI; Bentler, 1990). An alternative strategy is to index the absolute fit of a model by summarizing the differences between corresponding elements in the implied and observed covariance matrixes (i.e., the residuals). The root mean square error of approximation (RMSEA) is a commonly used index of this sort (Steiger & Lind, 1980). The RMSEA has the additional benefits of including a correction for model complexity and an interpretation that allows for tests of close, rather than exact, fit (Browne & Cudeck, 1993).

Returning now to the running example, I illustrate the general activity of evaluating fit and the use of these specific indexes of fit. Although I have shown the iteration history and parameter estimates only for the Time 1 protection latent variable in Table 19.2, the value of the fitting function is for the full model, which includes six latent variables and the relations between them. As noted, the value of F_{ML} can be used to produce a value that often is interpreted with reference to the χ^2 distribution. That value is computed as the product of F_{ML}, .630 in the current case, and the sample size minus one (1,654 in this case), which yields a value of 1,042.02. This value is referenced against the χ^2 distribution for the number of degrees of freedom in the model. That value, as noted, corresponds to the difference between the number of elements in the covariance matrix (computed as $p(p + 1)/2$, where p is the number of observed variables) and the number of free parameters the model. For our example, there are 210 observed variances and covariances and 65 free parameters in the model, yielding 145 degrees of freedom. The critical value for $p < .05$ for 145 degrees of freedom is 174.1, well below our observed value of χ^2. Assuming the unrealistic assumptions described earlier are met, this outcome indicates that the data implied by our model do not match the observed data within the sampling error.

Before moving to alternative means of evaluating fit, it will be useful to look more closely at the values reflected in this summary statistic. Again, for illustrative purposes, focusing only on that portion of the model relevant to the Time 1 protection latent variable, I have included two covariance matrixes in Table 19.3. On the left is the portion of the observed covariance matrix indicating the variances of, and covariances between, the indicators of the latent variable in question. Our model accounts for those covariances through the specification of the latent variable *F1*. If this specification is sufficient to explain the pattern of covariances, the parameter estimates should yield an implied, or reproduced, covariance matrix very close to the observed matrix. The implied covariance matrix for these variables is shown on the right in Table 19.3. Looking first to the diagonal, the values are, within rounding, identical. This reflects the fact that we fully account for variance in the variables in the model, either by the influence of the latent variable or the influence of a uniqueness component. The values of most interest are the off-diagonal elements, the observed and implied covariances between the variables. Note that some values in the implied matrix are higher and some are lower than their counterparts in the observed matrix; however, on the whole, the values are quite similar. The differences between corresponding elements in the two matrixes yield a third matrix, the residual matrix. For models that do not provide an acceptable account of the data as reflected in the values of fit indixes, examination of the residual matrix can be useful in pointing to specific relations that are over- or, more typically, underestimated in the specified model.

Returning now to the evaluation of fit, given the well-documented concerns about the validity of the χ^2 variate, I could consult one of a number of alternative indexes that have been proposed and evaluated and that generally have found support for most modeling circumstances. The typical recommendation is to choose at least two, each from a different class of index. As noted earlier, I favor CFI, which indexes incremental, or comparative fit, and *RMSEA* (with confidence intervals), which reflects absolute fit. Values of CFI can range from 0 to 1, with higher values indicating better fit. The value for a specified model is interpreted as the proportionate improvement in fit of the specified model over the null, or independence, model. Although values of .90 and higher are commonly interpreted as indicative of acceptable fit, simulation studies indicate that .95 is a more appropriate lower bound, with values between .90 and .95 interpreted as marginal fit (Hu & Bentler, 1999). The value of CFI for the example model is .952, indicating that the specified model provides an acceptable account of the data.

RMSEA indexes the degree of misspecification in a model per degree of freedom. To explain, a model can only imply data different from the observed data if it is more parsimonious than the unstructured data. To wit, a just identified model, which has the same number of free parameters as elements in the observed covariance matrix will perfectly reproduce the observed data. Such a model has zero degrees of freedom. Fixing a parameter yields a degree of freedom, but it also requires introducing an assumption in the model that is not born out in the data. To the extent the value of the fixed parameter is incorrect, it will result in a poorer fit of the model to the data

TABLE 19.3

Values in Observed and Implied Covariance Matrices for Indicators of One Latent Variable in the Example Model

	Observed matrix				Implied matrix			
Variable	v_1	v_2	v_3	v_4	v_1	v_2	v_3	v_4
v_1	.331				.333			
v_2	.149	.274			.165	.275		
v_3	.219	.244	.384		.224	.237	.383	
v_4	.141	.132	.183	.327	.130	.137	.186	.327

APA-COOPER_V2-11-0601-019.indd 352 12/13/11 7:11:08 PM

because of misspecification. RMSEA captures this property across the model. Thus, as noted, our example model has 65 degrees of freedom, which can be interpreted as 65 ways in which an assumption was imposed on the model that could be incorrect. RMSEA addresses the question of whether, on average, these assumptions are tenable given the data. If the assumptions imposed by fixed parameters are perfectly consistent with the observed data, they will result in no misspecification and yield a value for RMSEA of zero, indicating perfect model fit. Unlike χ^2, RMSEA typically is interpreted with reference to a criterion of close fit. Thus, values of .08 or less are viewed as indicative of acceptable model fit. Values from .08 to .10 indicate marginal fit, and values greater than .10 indicate poor fit. The standard error of RMSEA is known, and therefore a confidence interval can be put on the point estimate. It is customary to report the upper and lower limits of the 90% confidence interval. In such cases, the focus is on the degree to which the upper limit falls beneath the .08 and .10 criterion values. For the example model, the point estimate of RMSEA is .061, and the 90% confidence limits are .058 and .065. Because the upper confidence limit falls below .08, RMSEA indicates a close fit of the model to the data.

My use of ML implies that the data are multivariate normal; however, as might be assumed given the behaviors under study and the age of the participants, distributions for several of the measures are highly skewed. Indeed, the normalized value of 289.50 for Mardia's coefficient, a test of the correspondence of the distribution in the observed data to the multivariate normal distribution is very high. Examination of the observed variables indicates significant kurtosis in all three of the problem behaviors, particularly polydrug use. As noted in the section on estimation, normal theory estimates such as those produced by ML can be adjusted by a scaling correction that accounts for departure of the data from normality (Satorra & Bentler, 2001). This adjustment yields new values of fit indexes and adjusted standard errors for parameter estimates that, in simulation studies, closely approximate their values when estimated using distribution-free estimators. When this scaling correction is imposed on the ML solution for our example, it results in a substantially smaller value of $\chi^2 = 679.35$. The ratio of the normal theory χ^2 (1,042.02 in this case) to the scaled χ^2 is an informal index of the degree of nonnormality in the data. The value of about 1.5 for the ratio in our example indicates that the non-normality results in a normal theory χ^2 that is about 50% larger than it should be. Because the scaling adjustment affects both the specified and independence model fits, it is not necessarily the case that the value of CFI will increase with the adjustment. In fact, the CFI value drops slightly to .942, just below the cutoff, following the scaling adjustment. RMSEA declines so that the point estimate is .047 and the 90% confidence limits are .044 and .051. In this instance, the statistics to report and interpret are the scaled statistics, which suggest that the model offers an acceptable account of the data.

If, as in the present case, support for the overall fit of a model is obtained, then attention turns to the values of the parameter estimates. It is possible for a model to provide acceptable fit to the observed data but not yield parameter estimates that are consistent with expectation. The most basic question about model parameter estimates is whether they differ from zero, a hypothesis tested, as in other contexts, by the ratio of the parameter estimate to its standard error. The resulting statistic, the critical ratio, is interpreted as a *z* statistic. Other questions that can be addressed about parameter estimates are their difference from other assumed values (e.g., .30) and their difference from each other.

Because of the number of free parameters in the example model, I present parameters in the measurement and structural components separately. Parameter estimates for the measurement component are presented in Table 19.4. I have chosen to identify parameters using double-label notation; however, I might have used matrix notation or verbal descriptions depending on the audience. For each parameter, I include the estimated (or fixed) value, the critical ratio, and the standardized estimate. The estimate is an unstandardized value, the equivalent of an unstandardized regression coefficient or a covariance (an unstandardized correlation coefficient). The critical ratio is formed by dividing the estimate by its standard error—in this case, robust standard errors that have been scaled to

TABLE 19.4

Parameter Estimates and Test Statistics for Measurement Component of Example Model

	Time 1				Time 2		
Parameter	Estimate	Critical ratio	Standardized estimate		Estimate	Critical ratio	Standardized estimate
Loadings							
$F1,v_1$[a]	1.00		.69	$F4,v_{11}$[a]	1.00		.66
$F1,v_2$	1.06	19.92	.80	$F4,v_{12}$	1.35	27.79	.88
$F1,v_3$	1.44	26.41	.92	$F4,v_{13}$	1.58	27.86	.92
$F1,v_4$	0.83	18.10	.57	$F4,v_{14}$	0.82	18.40	.51
$F2,v_5$[a]	1.00		.40	$F5,v_{15}$[a]	1.00		.40
$F2,v_6$	1.96	12.01	.54	$F5,v_{16}$	2.00	11.83	.55
$F2,v_7$	1.55	12.96	.78	$F5,v_{17}$	1.51	11.88	.77
$F3,v_8$[a]	1.00		.64	$F6,v_{18}$[a]	1.00		.63
$F3,v_9$	4.61	16.29	.77	$F6,v_{19}$	4.42	17.64	.70
$F3,v_{10}$	0.60	10.41	.67	$F6,v_{20}$	0.70	10.42	.65
Unique variances							
e_1,e_1	0.18	24.09	.73	e_{11},e_{11}	0.18	22.77	.75
e_2,e_2	0.10	6.96	.60	e_{12},e_{12}	0.08	17.67	.48
e_3,e_3	0.06	10.38	.40	e_{13},e_{13}	0.06	13.15	.38
e_4,e_4	0.22	22.56	.82	e_{14},e_{14}	0.26	24.24	.86
e_5,e_5	0.24	26.70	.92	e_{15},e_{15}	0.25	27.23	.92
e_6,e_6	0.42	20.94	.84	e_{16},e_{16}	0.42	21.60	.83
e_7,e_7	0.07	13.10	.63	e_{17},e_{17}	0.07	11.76	.63
e_8,e_8	0.07	13.52	.77	e_{18},e_{18}	0.08	11.40	.77
e_9,e_9	0.73	13.36	.64	e_{19},e_{19}	1.08	16.15	.72
e_{10},e_{10}	0.02	5.76	.74	e_{20},e_{20}	0.04	4.79	.76
Latent variable variances[b]							
$F1,F1$	0.16	14.28	1.00				
$F2,F2$	0.05	7.04	1.00				
$F3,F3$	0.05	6.97	1.00				

Notes. Parameter estimates obtained using maximum likelihood. The standard errors for critical ratios were robust standard errors, rescaled to account for non-normality in the data. All estimates are significantly different from zero at $p < .001$. v_1 and v_{11} = protective normative beliefs, v_2 and v_{12} = realization of conflict between problem behavior and ideals, v_3 and v_{13} = commitment to avoid problem behavior, v_4 and v_{14} = bonding to school, v_5 and v_{15} = low self-esteem, v_6 and v_{16} = sensation seeking, v_7 and v_{17} = impulsive sensation seeking, v_8 and v_{18} = interpersonal aggression, v_9 and v_{19} = sexual activity, v_{10} and v_{20} = polydrug use. *F1* and *F4* = Protection. *F2* and *F5* = Risk. *F3* and *F6* = Problem Behavior.
[a]Value in unstandardized estimate column is fixed value of parameter, for which there is no standard error and, therefore, no critical ratio. [b]Time 2 latent variable variances are not estimated directly.

account for nonnormality in the data. This value is distributed as a *z* statistic; thus, critical values are 1.96, 2.58, and 3.33 for $p < .05$, $p < .01$, and $p < .001$, respectively. The standardized estimate is the value psychological scientists are inclined to present and interpret from multiple regression analysis and factor analysis. The standardized estimates for loadings are the equivalent of factor loadings in exploratory factor analysis. As is apparent from the large values for the critical ratio, all factor loadings at both time points are significantly different from zero. The standardized values range from .40 for the loading of low self-esteem (v_5 and v_{15}) on risk (*F2* and *F5*) to .92 for the loading of commitment to avoid problem behavior (v_3 and v_{13}) on protection (*F1* and *F4*). The moderate-to-large loadings indicate that the observed variables for each construct share considerable commonality, which is presumed to reflect the construct of interest. Moreover, the equivalent values for corresponding loadings at the two time points suggest that the latent variables are similarly defined across this period of time, a point to which I return later in this section.

The estimates of greatest interest are those that index the relations between latent variables. Those estimates, and other estimates within the structural component of the model, are provided in Table 19.5. Although the autocorrelations between uniqueness might have been presented with estimates from the measurement component of the model, they span the two waves of data collection and therefore are included as part of the structural component. The estimates of these covariances are highly significant, suggesting that, had these paths been omitted from the specification, the fit to the data would have been poor. The standardized estimates are correlation coefficients, and their magnitudes make clear that there is considerable and stable nonrandom variance in the indicators not explained by the latent variables. Moving toward the bottom of Table 19.5 and continuing the focus on covariances, the expected high correlations between the latent variables at Time 1 are apparent. Similarly, the covariances between unexplained variances in the latent variables at Time 2 are significant and strong.

These ancillary estimates appear in order, so our attention can now move to the directional relations between the latent variables at Time 1 and Time 2. First, I consider the autoregressive paths, which reflect the stability of the latent variables over the period of time covered by the study, about 8 months. It is no surprise that there is a high degree of stability in risk-related personality. The standardized estimate suggests that, once stability is taken into account, there is relatively little variance in Time 2 risk to be predicted by the other Time 1 latent variables. The stability of protection as reflected in attitudes, values, and perceptions also is quite high. Somewhat surprising is the nonsignificant autoregressive path for problem behavior, suggesting that the behaviors included do not necessarily reflect stable characteristics of individuals or situations at this age. Moreover, the weak stability coefficient indicates that the majority of variance in problem behavior at Time 2 could be predicted from variables at Time 1.

TABLE 19.5

Parameter Estimates and Test Statistics for Structural Component of Example Model

Parameter	Estimate	Critical ratio	Standardized estimate
Uniqueness autocorrelations			
e_1, e_{11}	.09	16.04***	.49
e_2, e_{12}	.02	5.98***	.21
e_3, e_{13}	.02	5.11***	.28
e_4, e_{14}	.12	14.25***	.51
e_5, e_{15}	.14	18.26***	.58
e_6, e_{16}	.22	15.01***	.51
e_7, e_{17}	.02	4.88***	.29
e_8, e_{18}	.03	9.18***	.46
e_9, e_{19}	.55	12.30***	.62
e_{10}, e_{20}	.01	4.55***	.28
Autoregressive paths			
F1,F4	.64	5.16***	.68
F2,F5	.83	9.83***	.81
F3,F6	.28	1.70	.27
Cross-lagged paths			
F1,F5	.05	0.81	.08
F1,F6	−.27	−2.79**	−.46
F2,F4	−.24	−2.43*	−.13
F2,F6	.06	0.75	.05
F3,F4	−.02	−0.13	−.01
F3,F5	.07	0.82	.08
Variances, disturbances			
d_4, d_4	.05	9.75***	.60
d_5, d_5	.02	6.07***	.60
d_6, d_6	.02	5.58***	.66
Covariances, latent variables			
F1,F2	−.07	−11.09***	−.77
F1,F3	−.08	−11.22***	−.88
F2,F3	.03	9.77***	.68
Covariances, disturbances			
d_4, d_5	−.02	−7.76***	−.67
d_4, d_6	−.03	−7.51***	−.82
d_5, d_6	.01	5.93***	.54

Notes. Parameter estimates obtained using maximum likelihood. The standard errors for critical ratios were robust standard errors, rescaled to account for non-normality in the data. *F1* = Time 1 Protection. *F2* = Time 1 Risk. *F3* = Time 1 Problem Behavior. *F4* = Time 2 Protection. *F5* = Time 2 Risk. *F6* = Time 2 Problem Behavior.

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Moving now to the cross-lagged paths, we find that only two coefficients are significantly different from zero. The effect of Time 1 protection on Time 2 problem behavior is significant and negative. The standardized estimate indicates that, for each standard deviation increase in Time 1 protection, there is almost half a standard deviation decrease in problem behavior at Time 2, controlling for other paths in the model. A second significant path is between Time 1 risk and Time 2 protection. The negative coefficient, in standardized terms, indicates that for each standard deviation increase in risk-related personality at Time 1, there is a modest decrease in protective attitudes, values, and perceptions at Time 2. These findings, coupled with the nonsignificant cross-lagged paths suggest that these protective factors exert a directional, potentially causal, influence on problem behavior and that problem behavior neither begets problem behavior during this period of time nor erodes protective attitudes, values, and perceptions. The addition of a third time point would allow the test of an intriguing hypotheses suggested by the findings: risk-related personality is associated with less protective attitudes, values, and perceptions, which results in greater problem behavior.

I began the section by outlining Jöreskog's (1993) discussion of the different approaches to using SEM. To this point, it would appear that I am using SEM in a strictly confirmatory manner. I could, however, consider alternatives to the specified model that might serve to increase my confidence that the model affords the best account of the data. Use of this alternative models approach is most effective when the models to be compared are nested. One model is nested in another when its free parameters are a subset of the free parameters in another model. I can produce a model nested in our example model by fixing some of its parameters. To illustrate this process, I consider two alternatives. In one alternative, I include only the autoregressive, or stability, paths; this model assumes no directional relation between the variables. To produce this model, I would fix the six cross-lagged paths in the model to zero. In so doing, I reduce the number of free parameters from 65 to 59, thereby increasing the degrees of freedom from 145 to 151. Adding restrictions always increases the value of the fitting function and, therefore, the value of χ^2 for a model. As with comparisons between R^2s in hierarchical multiple regression, the question is whether the increase is significant given the increase in degrees of freedom. Like R^2s, χ^2s are additive; thus, I can subtract the value for the less restrictive model from the value for the nested model and test it by referring to the χ^2 distribution for the difference in degrees of freedom. When normal theory estimates and statistics can be used, this test is as simple as subtracting the two χ^2s to produce $\Delta\chi^2$. When the scaling adjustment for non-normality is used, as in the present case, the subtraction is a bit more involved because the influence of non-normality is not identical for the two models (Satorra & Bentler, 2001). The scaled χ^2 for the model with no cross-lagged paths is 705.75. The scaled difference between this value and the scaled value of 679.35 for the specified model (computed using a executable file downloaded from http://www.abdn.ac.uk/~psy086/dept/sbdiff.htm) is highly significant, $\Delta\chi^2 = 26.79$, $p < .001$, indicating that (as we know), one or more of the cross-lagged paths is necessary to account for the data.

An alternative means of producing a nested model is through the use of equality constraints. Equality constraints are used in SEM as a means of testing whether two parameter estimates are significantly different. If, for example, I constrain two loadings to equality in a measurement model, I reduce the number of parameters to be estimated because the parameters constrained to be equal are treated as one parameter. This results in an increase in the number of degrees of freedom by one and allows for a statistical comparison of the two models using the χ^2-difference test. As noted, for the measurement component of the example model, corresponding loadings on the factors at the two time points appear to be equivalent. Indeed, it is important for hypothesis testing using the measures that I show their relation to the latent variables of interest remains constant across the 8-month period of time of the study. I can use equality constraints and a nested model comparison to test for this property of the indicators. Although there are 10 indicators, the loadings for three (one on each latent variable) are fixed and therefore not estimated. I can produce a nested model by constraining loadings for the remaining seven to be equal at the two time points. This yields seven degrees of freedom for testing the difference. The χ^2-difference test was highly significant, $\Delta\chi^2 = 43.27$, $p < .001$, indicating that at least one of the equality constraints was inconsistent with the data. Returning to the parameter estimates in Table 19.4, it appears that the loading of the second indicator of protection varies more from Time 1 to Time 2 and any of the other loadings. Relaxing this constraint yields a model with six more additional degrees of freedom than the original model and nested in it. Although the comparison between these two models yields a significant χ^2 difference, $\Delta\chi^2 = 13.72$, $p = .033$, the remaining differences are, practically speaking, trivial and would only be found different because of the high level of statistical power resulting from the large sample size.

MODEL MODIFICATION

Specified models, including the alternatives to which they are compared do not always provide an acceptable account of the data. That is, they yield values for the fit indexes of choice that fail with reference to a priori criteria (e.g., CFI < .95, RMSEA > .10). In such cases, if the researcher is using SEM in a strictly confirmatory or alternative models manner, the analysis is completed and the model or models declared inadequate as a description of the processes that produced the observed data. Yet, data typically are acquired at considerable effort and expense, leading most researchers to respond to evidence of unacceptable fit by asking, "What if?" which shifts them to using SEM for model generation. The what-if question might concern the observed variables: What if certain variables were dropped from the model or new variables added to it? It is generally the case, however, that the variables in a model are of sufficient interest that dropping one or more would leave a set that does not map well onto the original research question or conceptual model guiding the research. Typically, the what-if question concerns the relations between variables in a model and can be phrased in terms of fixed and free parameters. What if certain parameters that were originally fixed are freed? Or, what if one or more free parameters were fixed?

The consideration of whether model fit might be improved by changing the status of parameters in a model is termed specification searching. As the term indicates, the goal of specification searching is to find and address the misspecification in the original model that resulted in its lack of acceptable fit to the data. Specification searching typically takes one of two forms, each of which involves consulting the results from the initial estimation and testing. In *manual specification searching*, the researcher consults the tests of specific parameter to determine whether parameters that were free in the model could be fixed to zero or consults the residual matrix to determine whether one or more observed covariances were not sufficiently accounted for by the free parameters in a model. In *automated specification searching*, a computer program evaluates all fixed or free parameters and returns a list of those parameters that are contributing to the size of the model χ^2 relative to its degrees of freedom

Although the example model fits the data at an acceptable label, you will recall that the value of *CFI*, when scaled, dropped just below the criterion of .95. Although the value of RMSEA indicates a close fit, it might nonetheless prove useful to evaluate whether any adjustments to the model would result in a better account of the data. I first examine tests of individual parameters. Referring to Table 19.5, four of the six cross-lagged paths were nonsignificant, suggesting that fixing them to zero would simplify the model without significantly reducing fit. The resultant model does not affect the value of CFI but results in a slightly reduced value of RMSEA (which rewards parsimony). In addition to the trivial improvement in fit, these adjustments to the model result in the autoregressive path for problem behavior achieving statistical significance. This result illustrates the fact that parameters in a model are dependent. As such, changing the status of one will likely change the estimate of others. For this reason, post hoc adjustments are best made one at a time. Because the adjustments do not substantially improve the fit of the model, the model already offers a suitable account of the data, and there is a good chance the modifications would not replicate (MacCallum et al., 1992), I do not retain them. Although the final value of CFI falls just short of the ideal value of .95, the value of .942 for a model that includes no post hoc modifications is more impressive than a value greater than .95 for a model that includes modifications based on specification searching.

I next consult the residual matrix to determine whether there is evidence that any covariances have been underestimated because certain parameters were fixed at zero. Recall that the residual matrix is constructed by subtracting corresponding elements in the observed and implied covariance matrixes. Relatively large values raise the value of the fitting function and, thereby, reduce the favorability of fit indexes. Manually evaluating the residuals can be relatively straightforward for small models; however, for a model based on 20 observed variables, scanning the 190 residual covariances, even their more readily interpretable standardized form, often does not suggest obvious ways in which a model might be modified to improve its fit. Indeed, the only thing clear from scanning the residuals is that the relation between low self-esteem and a number of the remaining observed variables is underestimated. Given the number of these residuals and the position of relevant parameters across the model, I see no obvious fixed parameters that, if freed, would provide the maximum improvement in the fit of the model.

An alternative strategy, one better suited to larger models such as our example, is automated specification searching. Automated searching typically focuses on fixed parameters in a model (although some computer programs allow a focus on free parameters as well). The computer program, in effect, frees each fixed parameter and returns, in the form of a *modification index*, the amount by which the model χ^2 would be reduced if the parameter were freed (Bentler, 1986a; Sörbom, 1989). Some programs order these from highest to lowest, allowing a quick determination of which modifications would produce the greatest improvement in fit. As noted, however, each modification ripples through the model, affecting estimates of some, if not all, of the other parameters. Thus, for example, it may appear that model fit would improve if two fixed parameters were freed; however, freeing the first yields a model in which freeing the second offers no additional improvement in fit. Because of the dependency among parameters, a better approach is to use the multivariate approach to automated searching, in which the full set of modification indexes is evaluated in search of the set of modifications that, together, would produce the maximum improvement in fit. Although automated specification searching for the example models suggests a number of fixed parameters that, if freed, would improve the fit of the model, one stands out as particularly influential. Referring back to the covariance matrixes, the observed covariance between bonding to school and low self-esteem at Time 2 is −.328, but the value implied by the model is only −.051. Thus, the model only accounts for about 16% of the covariance between these variables. One means of accounting for the residual covariance in the model is to allow the uniquenesses for these two variables to covary. Indeed, the automated search indicates that freeing this parameter would result in a drop in the χ^2 for the model of 143.51. And respecifying the model to include this parameter results in a value of CFI that now exceeds the criterion, .952, and an improved value of RMSEA, .043. The improvement can be seen in the residual covariance matrix, for which the value of this covariance falls to zero (i.e., it is fully accounted for by the model).

Because adjustments to the original specification were made with reference to the data, the likelihood of Type I error is unacceptably high, and therefore fit statistics and indexes cannot be taken at face value (MacCallum et al., 1992). Thus, although researchers working in model generation mode might be tempted to confidently interpret the results from estimation of respecified models that produce acceptable values of fit indexes, they should instead proceed with caution because the likelihood is unacceptably high (i.e., > .05) that the model that fits the data in hand would not fit another set of data from the same population. As such, the results can only be interpreted with reference to the current sample. To infer with confidence that the parameter I freed on the basis of consulting the data applies to the population and not just my sample, I would need to demonstrate satisfactory fit of the modified model to data from a cross-validation sample from the same population.

INTERPRETATION

Once support has been obtained for an interesting and useful model, the focus moves to the final step in using SEM—interpretation and reporting. Interpretation begins with presentation. Although what is presented and, to some extent, how it is presented will be constrained by the medium and forum, a number of principles apply whether the findings are presented in a journal article, a book, or an oral presentation. Particularly when the findings are presented in written format, the observed data should be provided either in the form of a covariance matrix or a correlation matrix with standard deviations (which computer programs can use to derive the variances and covariances). For small data sets, this information can be presented in a table in the text of the document. For larger data sets, the information can be included in an appendix or a public access computer, typically accessed through a website (the covariance matrix for the example used in this chapter can be accessed at http://www.duke.edu/~rhoyle/data/Hoyle_HRM_chapter.dat). In addition, an accurate accounting of the focal model as well as any alternatives and modifications should be included. For many models, a path diagram will suffice. Regardless of the format in which the model is presented, a thorough accounting for degrees of freedom should be offered. The estimation method should be made explicit and its use justified. A set of fit indexes and criteria for their interpretation should be offered (see Hu & Bentler, 1999, for a useful overview). Parameter estimates should be provided. For simple models, these can be included on a path diagram and flagged for statistical significance. For more complex models, a table such as Tables 19.4 and 19.5 should be used. In my presentation of findings from the example analysis, I listed all parameters and provided unstandardized and standardized values as well as statistical test information. I would be unlikely to provide that level of detail in a journal article; however, the estimates and tests of the focal parameters (e.g., the autoregressive and cross-lagged paths) should be given in full. When multiple alternative models are specified a priori, it is useful to table the fit statistics for each model, which facilitates comparisons between models. Finally, if modifications are made to the originally specified model or an alternative specified alternative, it should be made clear to the reader the basis on which the model was modified, including how many consultations of the data were required to produce the model to be interpreted. Additional detail regarding these aspects of presentation can be found in a number of articles and chapters focused specifically on the presentation of results from SEM analyses (e.g., Boomsma, 2000; Hoyle & Panter, 1995; McDonald & Ho, 2002).

The substantive interpretation of SEM results refers to information provided in the presentation. Care should be taken to distinguish between comparative and absolute fit, unstandardized and standardized parameters estimates, and a priori and post hoc components of a model. As with any multivariate model, specific relations should be described with reference for other relations in the model.

Beyond these specific guidelines for presenting and interpreting the statistical results of an SEM analysis, two broader considerations should receive attention. A key consideration in interpretation is the extent to which a statistically defensible model provides a *uniquely* satisfactory account of the data. That is, for many well-fitting models, it is possible to generate one or more alternative models that are, statistically speaking, equivalent to the specified model (Breckler, 1990; MacCallum, Wegener, Uchino, & Fabrigar, 1993). For example, in a simple study involving two variables, *x* and *y*, the model that specifies *x* as a cause of *y* cannot be distinguished in terms of fit from a model that specifies *y* as a cause of *x* or *x* and *y* as simply correlated. A set of basic respecification rules can be used to generate possible alternatives to a specified model when they are less obvious (e.g., Stelzl, 1986). Sometimes the inferential conundrum produced by equivalent models can be resolved through design (e.g., *x* is manipulated) or a consideration of what the variables represent (e.g., *x* is a biological characteristic). Otherwise, the researcher can only infer that the results provide necessary, but not sufficient, support for the focal model.

Perhaps the most controversial aspect of interpreting results from SEM analyses is the inference of causality (e.g., Baumrind, 1983). Although the judicious application of SEM can strengthen causal inferences when they are otherwise warranted, SEM cannot overcome the significant limitations of nonexperimental designs, particularly when all data are gathered at one point in time. Because of its capacity for isolating putative causal variables and modeling data from longitudinal designs, SEM offers a stronger basis for inferring causality than commonly used statistical techniques. In the end, however, statistics yield to design when it comes to causal inferences, and therefore data generated by experimental or carefully designed quasi-experimental designs with a longitudinal component are required (Hoyle & Robinson, 2003).

PROTOTYPIC MODELS

Except for limitations associated with research design (e.g., temporal order of variables in longitudinal models) and model identification, there is considerable flexibility in model specification for SEM analyses. Indeed, as the number of observed variables increases, the number of models one might construct from them increases dramatically (e.g., Raykov & Marcoulides, 2001). For that reason, it is not possible to review all, or even a substantial proportion, of the models psychological scientists might construct and evaluate using SEM. Rather, to give a sense of the sort of models psychological scientists might evaluate using SEM, I briefly describe a number of prototypic models. These generally fall into one of two categories: (a) models focused primarily on latent variables without particular concern for how they relate to each other, and (b) models focused primarily on structural paths, whether they involve observed or latent variables.

At several points in the presentation of these models, I refer to comparisons between groups. Because such comparisons can be undertaken for any model evaluated using SEM, I describe the basic strategy for making such comparisons before presenting the models. In multigroup SEM, data are available on all of the observed variables for two or more groups of cases for which a comparison would be of interest. For instance, the example data set referenced throughout the chapter includes female and male middle-school students in sufficient numbers that parameters in the model could be estimated separately for the two groups and compared individually or in sets. To compare a model across I groups, I first need to divide the data according to group membership and produce (or allow the computer program to produce) separate covariance matrices. A single model is simultaneously fit to these matrixes, permitting the use of equality constraints to compare parameters between groups. For example, I could fit the example model to data from female and male students, constraining corresponding autoregressive and cross-lagged paths to be equal. That model is nested in and has nine more degrees of freedom than a model in which the parameters are free to vary between groups. The χ^2-difference test evaluates whether the set of constraints are consistent with the data. If the difference is significant, adding one or more of the equality constraints resulted in a decline in fit—that is, one or more parameters differ between the groups. Such multiple group comparisons are the equivalent of tests of statistical interaction because a significant between-group difference indicates that the relation between two variables differs across levels of a third variable—in this case, the grouping variable (for a general treatment of between-group comparisons, see Hancock, 2004).

MODELS FOCUSED ON LATENT VARIABLES

Although the model used to illustrate the steps involved in using SEM included latent variables, the primary focus of the model was the directional relations between variables from Time 1 to Time 2. Such is not always the case. For example, it might be the case the latent variables—protection, risk, and problem behavior—are not sufficiently well-defined to warrant tests of hypotheses about the relations between them. Or perhaps the constructs are generally well-defined, but there is reason to believe the relations between the indicators and latent variables differ across groups or time. These and a host of additional questions of potential interest to psychological scientists can be addressed using the measurement model. I illustrate the possibilities by describing two latent-variable focused models.

Measurement Invariance

When latent variables are included in a model and mean levels on those variables or their relations with

APA-COOPER_V2-11-0601-019.indd 360 12/13/11 7:11:09 PM

other variables are to be compared across samples or within a sample across time, a key concern is whether the meaning of the latent variables is consistent across the units to be compared. To the extent that the measurement model for a latent variable is consistent across samples or time, it is *invariant* with respect to measurement.

SEM is an appealing strategy for evaluating measurement invariance, because of the flexibility with which models can be specified and estimated, and the ease with which parameters can be compared across groups or time. For instance, consider the question of measurement invariance with respect to the risk construct in the example used throughout the chapter. Perhaps our concern is the degree to which dispositional risk influences problem behavior for young women and men. A meaningful comparison of this influence assumes that the indicators of risk function similarly for females and males. We can simultaneously estimate the measurement model for risk for females and males, using equality constraints to compare any or all parameters in the model. For instance, it would be important to show that responses on the indicators (e.g., impulsive decision making) are influenced by the latent variable to a similar degree for the two groups (i.e., the factor loadings are comparable). This evaluation is done within a multigroup context in which the models are simultaneously fit to separate covariance matrixes. If we constrain all of the factor loadings to be equal for females and males, and the fit is equivalent to the fit of a model in which the loadings are allowed to vary, we infer that the loadings are equivalent. The same strategy could be used to compare uniquenesses, the variance of the latent variable, or any parameter in the measurement model. Although it would involve a single covariance matrix, the same strategy would be used to evaluate the invariance of the construct for a given sample at two or more points in time as in the example. This form of invariance would be particularly important for indicators measured on multiple occasions across a period of development during which considerable change is expected (e.g., puberty).

The evaluation of measurement invariance can extend beyond the parameters estimated in standard measurement models. Specifically, the invariance of parameters that implicate the means of the observed variables can be included in the evaluation. This focus on metric invariance (Widaman & Reise, 1997) requires expanding the observed covariance matrix to include the observed means (which typically are zero because variables are mean centered), producing the *augmented moment matrix*. The specification expands as well to include intercepts in the measurement equations, which in the standard application included only the slopes as well as the means of the latent variables. Thus, in addition to the typical covariance structure, the model includes the mean structure (Curran, 2003). The addition of the mean structure allows for the comparison of additional aspects of the measurement model that are relevant to the question of consistency across groups or time and allows for the comparison of *structured means* on the latent variables even when the measurement model is not fully invariant (Byrne, Shavelson, & B. Muthén, 1989).

Latent Growth Models

The addition of the mean structure to a model makes possible another class of specialized models of relevance to phenomena studied by researchers in psychological scientists—latent growth models. Like trend analysis in repeated measures ANOVA, latent growth models focus on modeling patterns of means over time. For instance, returning to the illustrative example, if a sample of adolescents completed our bonding to school measure at the beginning of each of their last 2 years in middle school and their first 2 years of high school, we could examine the trajectory, or growth curve, of school belonging during these transitional years. As in repeated measures ANOVA, we can evaluate the first $k - 1$ order curves, with k indicating the number of repeated assessments.

A virtue of latent growth curve modeling in the SEM context is the ability to focus on individual growth (Singer & Willett, 2003; Willett & Sayer, 1994). For instance, in the school-bonding example, let us assume that, generally speaking, the trajectory during the period we are studying is linear. This outcome would be determined by observing acceptable fit of relevant portions of a model in which the means are fit to a straight line. If estimation yields support for this model, then the focus turns to variability in growth parameters—the slope and intercept of the linear trajectory. These growth parameters are modeled as variances of latent variables, hence, the label *latent* growth modeling. If, for instance, there is in effect no variability in the slope parameter, we might infer that the observed trajectory is normative. If there is variability in this parameter or the intercept parameter, then we can move to an interesting set of questions that concern the explanation or consequences of this variability. At this point, the model is referred to as a conditional growth model, because we are acknowledging that the growth parameters vary across individuals as a function of some characteristic of those individuals or the circumstances in which they live. This general strategy—determining the shape of the trajectory of change in a construct over time and then attempting to explain individual variability in trajectories—is particularly useful for studying development or the time course of a process.

MODELS FOCUSED ON STRUCTURAL PATHS

As with the example used throughout the chapter, SEM frequently is used to model the relations between variables, Although, as in path analysis, the relations can involve only measured variables, the most beneficial use of SEM is to model directional relations between latent variables. I illustrate the benefits of modeling relations in this way.

Cross-Lagged Panel Models

Panel models are those in which the same variables are assessed at multiple points in time. Our example is an instance of this type of model. The simplest case, two variables and two waves of data, illustrates the logic and benefits of cross-lagged panel models as analyzed using SEM. Modeling the two variables as latent variables ensures that no path coefficients are attenuated because of measurement error. This is a particular advantage in this context because path coefficients are to be compared, and it is important that any observed differences in coefficients be attributable to differences in the actual strength of the relations between constructs as opposed to differential attenuation of path coefficients because of measurement error. Another concern is that the cross-lagged path coefficients are estimated controlling for autoregression. By controlling for autoregression, we ensure that the cross-lagged paths do not reflect the covariation between stable components of the two constructs.

In the prototypic application of SEM to cross-lagged panel designs, we are interested in the absolute and the relative magnitudes of the coefficients associated with the cross-lagged paths. In absolute terms, we are interested in whether, after controlling for stability in the constructs, there is evidence of an association between them. This is determined by testing the departure of the estimates of the path coefficients from zero. In relative terms, we are interested in whether one cross-lagged path coefficient is larger than the other. This is determined by constraining the coefficients to be equal and determining whether the fit of the model declines significantly. If it does, then the constraint must be relaxed and the inference is that the two coefficients differ. If one cross-lagged coefficient is larger than the other, particularly if the smaller coefficient is not significantly different from zero, then the evidence supports a causal relation in the direction of the path associated with the larger coefficient. Thus, for instance, although it is not clear a priori whether problem behaviors influence, or are influenced by, protective attitudes, values, and perceptions, the findings in our example provide strong evidence that the effect is directional and, during the developmental period covered by the study, runs from protection to problem behavior.

Mediation

A distinguishing feature of theories in psychological science is their prescription of the processes or mechanisms by which one construct exerts a causal influence on another. Competing theories may agree about the causal relation but offer differing accounts of the process or mechanisms, prompting research focused specifically on the explanation for the effect. Variables that capture the putative explanations are mediators, and models in which their explanation of an effect is estimated are mediation models. Although such models can be estimated and tested by a series of multiple regression equations (Baron & Kenny, 1986), the inability to deal effectively with measurement error in multiple regression analysis, although always a concern, is of particular problematic for tests of mediation.

The simplest mediation model includes three directional effects: the direct effect of a causal variable on an outcome, the direct effect of the cause on a mediator, and the direct effect of the mediator on the outcome. The latter two effects, together, constitute the indirect effect of the cause on the outcome. The product of the path coefficients for these two paths indexes the indirect, or mediated, effect. In this simple model, the direct and indirect paths are perfectly correlated. As one increases, the other decreases. Moreover, the strongest evidence in support of a putative mediator is a significant indirect effect and a nonsignificant direct effect of the cause on the outcome. Although a number of conditions might lead to the underestimation of the indirect effect, one that is addressed well by SEM is unreliability in the mediator (Hoyle & Kenny, 1999). As the degree of unreliability in the mediator increases, the indirect effect is increasingly underestimated and the direct effect is overestimated. SEM offers a means of estimating these effects with unreliability removed from the mediator. When multiple indicators are available, this is accomplished by modeling the mediator as a latent variable. When only one indicator is available, unreliability can be removed from the mediator by modeling it as a latent variable with a single indicator whose uniqueness is fixed to reflect an estimate of unreliability in the indicator. Using either strategy ensures that the indirect effect is not underestimated, the direct effect is not overestimated, and, as a result, evidence for mediation by the proposed process or mechanism is not missed when it should be found.

Latent Interaction

Measurement error is similarly problematic for tests of moderation, or interaction, in which the effect of one variable on another is assumed to vary across levels of a third variable. The effects of measurement error are compounded in tests of moderation because those tests often involve the product of two variables, each of which may be measured with error. The reliability of the product term typically is lower than the reliability of either of the individual variables, increasing the likelihood that a model that includes the main effects and interaction will yield support only for the main effects. Fortunately, unreliability can be effectively removed from the individual variables and their product by modeling them as latent variables using SEM.

Although the basic approach to using SEM in this way was documented in the mid-1980s (Kenny & Judd, 1984), it has seen limited use. In part, this is due to the fact that specification of the latent interaction variable is complex—increasingly so as the number of indicators of each individual variable increases beyond two. This complexity stems from the fact that indicators of the latent interaction variable are the cross-products of the indicators of the two variables (e.g., two latent variables with three indicators each yields nine indicators of the latent interaction variable) and their relation to both the latent variable and their uniqueness term is nonlinear. It is now apparent that this complexity can be substantially reduced with little or no impact on the integrity of the estimate of the interaction effect. One approach involves fixing most of the parameters (i.e., loadings and uniquenesses) in the latent interaction variable using values calculated from parameter estimates obtained from a model that includes only the individual latent variables (i.e., no latent interaction variable; Ping, 1995). The nonlinearity is conveyed in the fixed parameters, substantially simplifying its specification and estimation. Recent work supports the use of an even simpler approach to specifying the latent interaction term that yields results that differ trivially from those obtained using the full specification. In this approach, each indicator of the two variables is used in only one product variable (e.g., two latent variables with three indicators each—x_1, x_2, x_3, and y_1, y_2, y_3—yield only three indicators of the latent interaction variable—x_1y_1, x_2y_2, x_3y_3), and the relation of these products to their uniqueness and the latent variable is modeled as linear (Marsh, Wen, Nagengast, & Hau, in press). By extracting unreliability from the interaction term, these methods significantly increase the likelihood of detecting moderation when it is present.

APA-COOPER_V2-11-0601-019.indd 363 12/13/11 7:11:09 PM

CONCLUSION

The flexibility and increasing generality of SEM make it an attractive alternative to traditional statistical methods such as ANOVA, multiple regression analysis, and exploratory factor analysis for psychological scientists. Although those methods have been, and will continue to be, well suited to many hypothesis tests in psychology, they significantly limit the range of hypotheses than can be considered. SEM, with its focus on modeling rather than simply analyzing data, affords new ways of studying the measurement of complex constructs and the relations between them. The ability to embed latent variables in any model is perhaps the most important feature of SEM and, as I have demonstrated, can be used to considerable benefit when the observed variables are measured with error. If the possibilities described and illustrated in this chapter have piqued your interest, I recommend building on the foundation provided here by consulting one or more of the detailed treatments offered in the growing array of textbooks (e.g., Bollen, 1989; Kaplan, 2009; Kline, 2010; Schumacker & Lomax, 2004) and edited volumes (e.g., Hancock & Mueller, 2006; Hoyle, 1995, 2012; Marcoulides & Schumacker, 1996).

References

Arbuckle, J. L. (1995). AMOS for Windows Analysis of Moment Structures (Version 3.5) [Computer software]. Chicago, IL: SmallWaters Corp.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Baumrind, D. (1983). Specious causal attribution in the social sciences: The reformulated stepping-stone theory of heroin use. *Journal of Personality and Social Psychology*, *45*, 1289–1298. doi:10.1037/0022-3514.45.6.1289

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1986a). *Lagrange multiplier and Wald tests for EQS and EQS/PC*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1986b). Structural equation modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika*, *51*, 35–51. doi:10.1007/BF02293997

Bentler, P. M. (1989). *EQS: Structural equations program manual*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606. doi:10.1037/0033-2909.88.3.588

Bentler, P. M., & Speckart, G. (1979). Models of attitude–behavior relations. *Psychological Review*, *86*, 452–464. doi:10.1037/0033-295X.86.5.452

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289–308. doi:10.1007/BF02293905

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Blunch, N. J. (2008). *Introduction to structural equation modelling using SPSS and AMOS*. Thousand Oaks, CA: Sage.

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

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling*, *7*, 461–483. doi:10.1207/S15328007SEM0703_6

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Lawner & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 199–217). New York, NY: Academic Press.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, *107*, 260–273. doi:10.1037/0033-2909.107.2.260

Browne, M. W. (1984). Asymptotic distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Thousand Oaks, CA: Sage.

Burks, B. S. (1928). The relative influence of nature and nurture upon mental development: A comparative study of foster parent–foster child resemblance and true parent–true child resemblance. In G. M. Whipple (Ed.), *The twenty-seventh yearbook of the National Society for the Study of Education* (pp. 219–316). Bloomington, IL: Public School Publishing Company.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts,* *applications, and programming*. New York, NY: Psychology Press.

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

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Taylor & Francis.

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466. doi:10.1037/0033-2909.105.3.456

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

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529–569. doi:10.1207/s15327906mbr3804_5

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292–307. doi:10.1037/1082-989X.2.3.292

Denis, D. J., & Legerski, J. (2006). Causal modeling and the origins of path analysis. *Theory and Science, 7*. Retrieved from http://theoryandscience.icaap.org/content/vol7.1/denis.html

Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, *72*, 177–182. doi:10.1037/h0027876

Duncan, O. D. (1975). *Introduction to structural equation models*. New York, NY: Academic Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155–174. doi:10.1037/1082-989X.5.2.155

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. doi:10.1037/1082-989X.9.4.466

Fox, J. (2006). Structural equation modeling with the SEM package in R. *Structural Equation Modeling*, *13*, 465–486. doi:10.1207/s15328007sem1303_7

Goldberger, A. S. (1971). Econometrics and psychometrics: A survey of commonalities. *Psychometrika*, *36*, 83–107. doi:10.1007/BF02291392

Goldberger, A. S., & Duncan, O. D. (Eds.). (1973). *Structural equation models in the social sciences*. New York, NY: Academic Press.

Hancock, G. R. (2004). Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables. In D. Kaplan (Ed.), *The handbook of quantitative methodology for the social sciences* (pp. 317–334). Thousand Oaks, CA: Sage.

Hancock, G. R., & Mueller, R. O. (Eds.). (2006). *Structural equation modeling: A second course*. Greenwich, CT: Information Age.

Harrington, N. G., Giles, S. M., Hoyle, R. H., Feeney, G. J., & Yungbluth, S. C. (2001). Evaluation of the All Stars Character Education and Problem Behavior Prevention Program: Effects on mediator and outcome variables for middle school students. *Health Education and Behavior*, *28*, 533–546. doi:10.1177/109019810102800502

Hershberger, S. L. (2003). The growth of structural equation modeling: 1994–2001. *Structural Equation Modeling*, *10*, 35–46. doi:10.1207/S15328007SEM1001_2

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Hoyle, R. H. (Ed.). (2012). *Handbook of structural equation modeling*. New York, NY: Guilford Press.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hoyle, R. H., & Robinson, J. I. (2003). Mediated and moderated effects in social psychological research: Measurement, design, and analysis issues. In C. Sansone, C. Morf, & A. T. Panter (Eds.), *Handbook of methods in social psychology* (pp. 213–233). Thousand Oaks, CA: Sage.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–5. doi:10.1080/10705519909540118

Huba, G. J., & Bentler, P. M. (1982). On the usefulness of latent variable causal modeling in testing theories of naturally occurring events (including adolescent drug use): A rejoinder to Martin. *Journal of Personality and Social Psychology*, *43*, 604–611. doi:10.1037/0022-3514.43.3.604

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, *48*, 398–407. doi:10.2307/2095231

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York, NY: Academic Press.

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.

Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.8 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, *44*, 457–467. doi:10.1037/0012-1649.44.2.457

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: Sage.

Keesling, J. W. (1972). *Maximum likelihood approaches to causal analysis* (Unpublished doctoral dissertation). University of Chicago, Chicago, IL.

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, CA: Sage.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210. doi:10.1037/0033-2909.96.1.201

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

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. doi:10.1037/0033-2909.111.3.490

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

Mann, H. B., & Wald, E. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, *11*, 173–220. doi:10.2307/1905674

Marcoulides, G. A., & Schumacker, R. E. (Eds.). (1996). *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Erlbaum.

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, *16*, 191–225. doi:10.1080/10705510902751010

Marsh, H. W., Wen, Z., Nagengast, B., & Hau, K.-T. (in press). Latent-variable approaches to tests of interaction effects. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*. New York, NY: Guilford Press.

Maruyama, G., & McGarvey, B. (1980). Evaluating causal models: An application of maximum-likelihood analysis of structural equations. *Psychological Bulletin*, *87*, 502–512. doi:10.1037/0033-2909.87.3.502

McArdle, J. J. (in press). Some ethical issues in factor analysis. In A. T. Panter & S. Sterba (Eds.), *Handbook of ethics in quantitative methodology*. New York, NY: Taylor & Francis.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64–82. doi:10.1037/1082-989X.7.1.64

Mueller, R. O. (1999). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York, NY: Springer.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide*. Los Angeles, CA: Authors.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered, categorical and continuous latent variable indicators. *Psychometrika*, *49*, 115–132. doi:10.1007/BF02294210

Myung, J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100. doi:10.1016/S0022-2496(02)00028-7

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

Ping, R. A., Jr. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research, 32*, 336–347. doi:10.2307/3151985

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190. doi:10.1007/BF02295939

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. doi:10.1207/S15328007SEM0801_8

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*, 1–12. doi:10.1037/a0018326

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

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514. doi:10.1007/BF02296192

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371–384. doi:10.1007/BF02294623

Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods*, *7*, 210–227. doi:10.1037/1082-989X.7.2.210

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

Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, *21*, 309–331. doi:10.1207/s15327906mbr2103_3

Tomer, A. (2003). A short history of structural equation models. In B. H. Pugesek, A. Tomer, & A. von Eye (Eds.), *Structural equation modeling: Applications in ecological and evolutionary biology* (pp. 85–124). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511542138.005

Tremblay, P. F., & Gardner, R. C. (1996). On the growth of structural equation modeling in psychological journals. *Structural Equation Modeling*, *3*, 93–104. doi:10.1080/10705519609540035

Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin*, *74*, 193–212. doi:10.1037/h0029778

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association. doi:10.1037/10222-009

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69–83). New York, NY: Academic Press.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381. doi:10.1037/0033-2909.116.2.363

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79. doi:10.1037/1082-989X.12.1.58

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences, 6*, 320–332. doi:10.1073/pnas.6.6.320

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161–215. doi:10.1214/aoms/1177732676