Structural equation models and the quantification of behavior

Kenneth A. Bollen[1] and Mark D. Noble

Department of Sociology, University of North Carolina, Chapel Hill, NC 27599

Quantifying behavior often involves using variables that contain measurement errors and formulating multiequations to capture the relationship among a set of variables. Structural equation models (SEMs) refer to modeling techniques popular in the social and behavioral sciences that are equipped to handle multiequation models, multiple measures of concepts, and measurement error. This work provides an overview of latent variable SEMs. We present the equations for SEMs and the steps in modeling, and we provide three illustrations of SEMs. We suggest that the general nature of the model is capable of handling a variety of problems in the quantification of behavior, where the researcher has sufficient knowledge to formulate hypotheses.

error in variables | factor analysis | path analysis | LISREL | covariance structures

Among the many problems in quantifying behavior are the challenges presented by the multiple equations to study and the difficulty of accurately measuring key concepts. For instance, scientists might be interested in general arousal and how it relates to more specific forms of arousal, or researchers might want to analyze quality of sleep and it relationship to different forms and intensities of pain. Arousal, quality of sleep, pain intensity, and numerous other variables are difficult to measure without considerable measurement error. Additionally, studying how several or more of these difficult to measure latent variables relate to each other is an even more arduous task when the multiequation nature of the problem is included.

Ignoring these issues leads to inaccuracy of findings. Ignoring the measurement error in arousal or pain data, for instance, leads to inaccurate assessments of effects. Therefore, our assessment of perception of pain on quality of sleep is unlikely to be correct if we do not take account of the measurement error. We might have several ways to measure the same latent arousal variable and not be sure how to incorporate these ways into the model. Also, if we ignore the indirect effects of one variable on another variable and concentrate only on the direct effect, we are more likely to be mistaken in our assessment of how one variable affects another variable.

Structural equation models (SEMs) refer to modeling techniques popular in the social and behavioral sciences that are equipped to handle multiequation models, multiple measures of concepts, and measurement error. This general model incorporates more familiar models as special cases. For instance, multiple regression is a special form of SEM, where there is a single dependent variable and multiple covariates and the covariates are assumed to be measured without measurement error. ANOVA is a another specialization where the covariates are assumed to be dichotomous variables. Factor analysis is yet another special form of the latent variable SEM. Here, we assume that we have multiple indicators that measure one or more factors and that the factors are permitted to correlate or not correlate. Recursive models, nonrecursive models, growth curve models, certain fixed and random effects models, etc. can all be incorporated as special cases of the general latent variable SEM. However, the structural component in SEM reflects that the researcher is bringing causal assumptions to the model, whereas multiple regression, ANOVA, etc. might be applied purely for descriptive purposes without any causal assumptions (1).

The purpose of this work is to give a brief overview of latent variable SEMs and to illustrate them with several hypothetical examples. It is meant to give the reader a sense of the potential of these procedures for research on the quantification of behavior. The next section explains SEMs with a brief verbal description followed by a more formal model of latent variable SEMs and its assumptions. This section is followed by a section that contains three illustrations. These illustrations are examples meant to give the reader a flavor of the type of applications that might be done. The section on illustrations is followed by the conclusions.

What Are SEMs?

SEMs are traceable at least back to the path analysis work of Wright (1, 2). SEMs did not receive much attention until they were introduced into sociology in the 1960s by Blalock (3) and Duncan (4). From sociology, they spread to the other social sciences and psychology. A turning point was the development of the LISREL model by Jöreskog (5) and the LISREL SEM software. There have been numerous contributors to SEM from several disciplines, and we cannot fully describe these here. However, historic reviews are available from several sources (6, 7). SEMs have continued to diffuse through numerous disciplines and are reaching beyond the social and behavioral sciences into biostatistics, epidemiology, and other areas. Furthermore, there are a variety of SEM software packages, including LISREL (8), Mplus (9), AMOS (10), EQS (11), and the SEM procedure in R (12).

SEMs are marked by typically including two or more equations in the model. This process differs from the usual single equation regression model that has a single dependent variable and multiple covariates. In SEMs, it is not unusual to have a number of equations with several explanatory variables in each equation. The usual terms of dependent variable and independent variable make less sense in this context, because the dependent variable in one equation might be an independent variable in another equation. For this reason, the variables in a model are called either endogenous or exogenous variables. At the risk of oversimplifying, endogenous variables are variables that appear as dependent variables in at least one equation. Exogenous variables are never dependent variables and typically, are allowed to correlate with each other, although explaining the source of their associations is not part of the model.

Another division among the variables is between latent and observed variables. Latent variables are variables that are important to the model but for which we have no data in our dataset (13). Observed variables are variables that are part of our analysis but for which we have values in our dataset. For in-

stance, sharp pain might be a latent variable of interest, and the self-reports of perceived sharp pain would be an observed variable to measure it. We recognize that the subjective measure is a less than perfect measure of the latent variable of sharp pain. Indeed, we could devise several differently worded questions to try to tap the latent sharp pain variable. With SEMs, we would build a measurement model of the relationship between each indicator and the latent sharp pain variable. This model would enable us to estimate the relationship between each indicator and the latent variable and determine which measure is the most closely related to the latent variable.

Similarly, if interest lies in another latent variable such as quality of sleep, then we could follow a similar procedure to develop several indicators of sleep quality and build a measurement model with each indicator related to the sleep quality latent variable. Combining the measurement model of sleep quality with that for sharp pain, we could construct a latent variable model that examines whether the latent sharp pain variable influences the latent sleep quality variable while controlling for the measurement error in the indicators.

The preceding discussion illustrates that the latent variable SEM permits us to study the relationships of latent variables to each other and between latent and observed measures of these latent variables. By so doing, the analysis can study the relationships among the theoretical variables of primary interest without having the effects confounded with measurement error. In addition, the analysis permits assessment of which measures seem to be most closely related to the latent variables.

Model. The last section gave more of a verbal than formal description of the latent variable SEMs. In this subsection, we present the model and assumptions more formally. The latent variable SEM is conveniently divided into two parts: the latent variable model and the measurement model. The latent variable model (sometimes called the structural model) is (Eq. **1**)

$$\eta_i = \alpha_\eta + \mathbf{B}\eta_i + \mathbf{\Gamma}\xi_i + \zeta_i, \qquad [1]$$

where η_i is a vector of latent endogenous variables for unit i, α_η is a vector of intercept terms for the equations, $\mathbf{B}$ is the matrix of coefficients giving the expected effects of the latent endogenous variables (η) on each other, ξ_i is the vector of latent exogenous variables, $\mathbf{\Gamma}$ is the coefficient matrix giving the expected effects of the latent exogenous variables (ξ) on the latent endogenous variables (η), and ζ_i is the vector of disturbances. The i subscript indexes the ith case in the sample. We assume that $E(\zeta_i) = 0$, $COV(\xi_i', \zeta_i) = 0$, and $(\mathbf{I} - \mathbf{B})$ is invertible. As we stated, exogenous variables are variables that are not explained within the model and that are uncorrelated with all disturbances in the system. Endogenous variables are variables that are directly influenced by other variables in the system other than its disturbance. Two covariance matrices are part of the latent variable model: $\Sigma_{\xi\xi}$ is the covariance matrix of the exogenous latent variables (ξ) and $\Sigma_{\zeta\zeta}$ is the covariance matrix of the equation disturbances (ζ). The mean of ξ is μ_ξ.

The latent variable model reflects the hypotheses about how the different concepts such as perceived pain and quality of sleep relate to each other. In its general form, it incorporates any number of endogenous or exogenous latent variables.

The measurement model links the latent to the observed responses (indicators). It has two equations (Eqs. **2** and **3**),

$$\mathbf{y}_i = \alpha_y + \mathbf{\Lambda}_y \eta_i + \varepsilon_i \text{ and} \qquad [2]$$

$$\mathbf{x}_i = \alpha_x + \mathbf{\Lambda}_x \xi_i + \delta_i, \qquad [3]$$

where $\mathbf{y}_i$ and $\mathbf{x}_i$ are vectors of the observed indicators of η_i and ξ_i, respectively, α_y and $\alpha_\mathbf{x}$ are intercept vectors, $\mathbf{\Lambda}_y$ and $\mathbf{\Lambda}_x$ are ma-

trices of factor loadings or regression coefficients giving the impact of the latent η_i and ξ_i on $\mathbf{y}_i$ and $\mathbf{x}_i$, respectively, and ε_i and δ_i are the unique factors of $\mathbf{y}_i$ and $\mathbf{x}_i$. We assume that the unique factors (ε_i and δ_i) have expected values of zero, have covariance matrices of $\Sigma_{\varepsilon\varepsilon}$ and $\Sigma_{\delta\delta}$, respectively, and are uncorrelated with each other and with ζ_i and ξ_i.

Using these equations, we can illustrate the generality of the model. First, researchers can do all confirmatory factor analysis models using only Eq. **3**. This model includes the observed variables or measures in the vector $\mathbf{x}_i$, the factor loadings are in the $\mathbf{\Lambda}_x$ matrix, and the intercepts of the equation are in α_x. The δ_i vector contains the unique factors. We could have just as easily used Eq. **2** to present the confirmatory factor analysis model.

Multiple regression provides another example. By introducing the restrictions of assuming no measurement error in $\mathbf{x}_i$ ($\alpha_x = 0$, $\mathbf{\Lambda}_x = \mathbf{I}$, $\Sigma_{\delta\delta} = \mathbf{0}$) and no error in $\mathbf{y}_i$ ($\alpha_y = 0$, $\mathbf{\Lambda}_y = \mathbf{I}$, $\Sigma_{\varepsilon\varepsilon} = 0$), using a single dependent variable ($\mathbf{y}_i$ is a scalar), and setting $\mathbf{B} = 0$, we get a multiple regression model where the implicit assumption of no measurement error is made explicit. ANOVA follows if we let $\mathbf{x}_i$ consist only of dummy variables.

Econometrics developed the idea of simultaneous equation models, which is a multiequation model where a series of endogenous variables are related to each other as well as to a series of exogenous variables. Unlike regression, simultaneous equations have multiple equations rather than a single one. However, like multiple regression, measurement error is assumed absent. By assuming that there is a single observed variable for each latent variable and that there is no measurement error (i.e., $\alpha_x = 0$, $\mathbf{\Lambda}_x = \mathbf{I}$, $\Sigma_{\delta\delta} = 0$, $\alpha_y = 0$, $\mathbf{\Lambda}_y = \mathbf{I}$, and $\Sigma_{\varepsilon\varepsilon} = 0$), the latent variable model becomes (Eq. **4**)

$$\mathbf{y}_i = \alpha_y + \mathbf{B}\mathbf{y}_i + \mathbf{\Gamma}\mathbf{x}_i + \zeta_i, \qquad [4]$$

and this equation is equivalent to the simultaneous equation model from econometrics that is used in some of the other social sciences as well. Furthermore, nonlinear models that relate dichotomous, ordinal, or censored observed variables to the latent variables also are possible (14), although we do not cover these models in this overview. Special cases of this model include fully recursive and nonrecursive models that also assume no measurement errors.

The general model also incorporates growth curve models, random and fixed effects models, and a large number of other models that are used in statistical modeling. Below, *Illustrations*, *Illustration 2* will show a growth curve model; however, other applications of the general model are not shown here (6).

Modeling Steps. Latent variable SEM occurs in a series of steps. These steps include model specification, model-implied moments, identification, estimation, model fit, and respecification. We give an overview of each step.

Model specification. Model specification is the term that refers to the formulation of the model. It involves specifying the primary latent variables and how they relate to each other. It also involves constructing the measurement model that specifies the relationship between the latent and observed variables and whether there are any correlated unique factors predicted. Another way of understanding this step is that it gives the specific composition and form of the matrices that are in the latent variable and measurement models.

The information necessary for model specification comes from the subject matter experts and their knowledge of theory and prior research in this area. If the information in an area leads to two or more different models, the researcher should use all plausible models. Later in the process, the fit of these different models can be compared to determine which is more consistent with the data.

The model specification step makes it difficult for statisticians or biostatisticians to use SEM unless they have substantive expertise or are collaborating with a colleague who has this knowledge.

Model specification is done using the latent variable and measurement models from the SEM or sometimes, is represented in a path diagram. A path diagram is a figure that represents the latent variable and measurement models for a particular application. Later, we provide examples of path diagram, and we will explain their symbols.

Model-implied moments. After a model is specified, it has implications for the observed variables. One set of implications has to do with predicting the means, variances, and covariances of the observed variables using the parameters from the model specification. These means, variances, and covariances are the moments of the observed variables.

The parameters of the model are placed in a vector called θ. If a model specification is valid, then the following mean and covariance structures hold (Eqs. **5** and **6**):

$$\mu = \mu(\theta) \text{ and} \tag{5}$$

$$\Sigma = \Sigma(\theta). \tag{6}$$

The implication is that if the model is true, then the parameters (θ) of the model should enable us to exactly predict the means, variances, and covariances of the observed variables. In other words, there should be an exact match between the population means, variances, and covariances and those values implied by the parameters in the model.

These implied moments are useful in estimating and assessing the fit of a model. They also are useful in establishing model identification, a topic to which we now turn.

Identification. Identification concerns whether it is possible to uniquely estimate values of all of the parameters in an SEM. When we can, the model is identified. When we cannot, it is underidentified. In multiple regression or ANOVA, the assumptions of the model lead to an exactly identified model, and therefore, this issue is rarely discussed in these contexts. However, when we are dealing with multiequation models with multiple indicators and latent variables, we cannot assume that all model parameters are identified. For instance, suppose that we had two observed variables, y_{1i} and y_{2i}, that had a feedback relationship such that y_{1i} caused y_{2i} and y_{2i} caused y_{1i}. The errors of these two equations are correlated. Having just the means, variances, and covariance of these two variables would be insufficient to identify the coefficients in this feedback relationship. Even with data from the population, we could not find the true coefficient values.

If a parameter in a model is not identified, then there will be two or more values of the parameters that are equally consistent with the data, and the researcher will not be able choose among them empirically, even if population data were available. Hence, knowing the identification status of a model is important to proper estimation and interpretation of a model. In general, the means (μ), variances, and covariances (Σ) of the observed variables are identified parameters for virtually all observed variables that are part of a model. In the most common situation, the identification question is whether the parameters of the model (e.g., factor loadings, regression coefficients, and variances of latent variables) can be written as unique functions of the known to be identified means, variances, or covariances of the observed variables. If yes, then the parameters and model are identified. If no, the model is not identified, even if some equations or individual parameters are identified.

There are rules of identification that apply to some models (15, 16). Empirical techniques for model identification are part of nearly all SEM software. The empirical procedures generally are accurate but not always, because they examine a more limited type of identification (17).

Models that are overidentified (that is, models that have more than enough information to identify all parameters) provide a way to test the overall model structure. However, before we test the model, we need to estimate it.

Estimation. We review two classes of estimators for SEM. We refer to these as the model-implied moment (MIM) and the model-implied instrumental variable (MIIV) estimators. The MIM estimators are the most common, and therefore, we treat them first.

MIM estimators. In a previous section, we examined the MIMs, means, variances, and covariances of the observed variables and the fact that these values should equal the corresponding population means, variances, and covariances of the observed variables if the model is valid. The MIM estimators are based on this relationship in that they choose values of the model parameters such that, when they are substituted into the implied moments, they will come as close as possible to reproducing the means, variances, and covariances of the observed variables. These estimators are also called full information estimators. They are full information in that all relationships in the system, including the covariances, variances, and means of exogenous variables and disturbances, play a role in developing the estimates. There are a number of full information estimators, but the dominant full information estimator is the maximum likelihood (ML) estimator. The ML fitting function in SEMs is (Eq. **7**)

$$F_{ML} = \ln|\Sigma(\theta)| - \ln|S| + \text{tr}[\Sigma^{-1}(\theta)S] - P_z$$
$$+ (\bar{z} - \mu(\theta))'\Sigma^{-1}(\theta)(\bar{z} - \mu(\theta)), \tag{7}$$

where **S** is the sample covariance matrix, $\bar{z}$ is the vector of the sample means of the observed variables ($\bar{y}$ and $\bar{x}$ stacked in a vector), P_z is the number of observed variables, ln is the natural log, |·| is the determinant, and tr is the trace of a matrix. The ML estimator, $\hat{\theta}$, is chosen so as to minimize F_{ML}. Like all ML estimators, $\hat{\theta}$ has several desirable properties. It is consistent, asymptotically unbiased, asymptotically efficient, and asymptotically normally distributed, and the asymptotic covariance matrix of $\hat{\theta}$ is the inverse of the expected information matrix.

The classic derivation of the ML estimator assumes that all observed variables come from a multinormal distribution. However, a variety of studies have proved that this estimator maintains its properties under less restrictive assumptions. For instance, the ML properties still hold if the observed variables come from a continuous distribution with no excess multivariate kurtosis (18). Robustness studies provide conditions under which the significance tests based on the ML estimator maintain their desirable asymptotic properties even when none of the preceding conditions holds (19). Finally, there are bootstrapping methods (20, 21) and corrected significance tests (22, 23) that enable significance testing even when the robustness conditions fail. Thus, there are numerous options to handle observed endogenous variables that come from nonnormal distributions.

A negative consequence of being a full information estimator such as ML is that, in the likely situation that there are structural misspecifications in the model, there is a great potential for the structural errors in one part of the system to spread their effects elsewhere. This spread can happen even if the other parts of the model are correctly specified. Given the high likelihood of at least some structural error in models, this spread is a drawback of full information MIM estimators. This potential for structural errors or model misspecifications is motivation to use more robust estimators such as the MIIV estimators discussed below.

MIIV estimators. The MIIVs come in a couple of different forms. Here, we concentrate on the MIIV two-stage least squares (MIIV-2SLS) estimator from Bollen (24, 25).

The main requirement of this MIIV-2SLS estimator is that each latent variable have a scaling indicator such that (Eqs. **8** and **9**)

$$\mathbf{y}_{1i} = \eta_i + \varepsilon_{1i} \text{ and} \qquad [8]$$

$$\mathbf{x}_{1i} = \xi_i + \delta_{1i}, \qquad [9]$$

where $\mathbf{y}_{1i}$ contains the scaling indicators of η_i and $\mathbf{x}_{1i}$ contains the scaling indicators of ξ_i. These equations imply that $\eta_i = \mathbf{y}_{1i} - \varepsilon_{1i}$ and $\xi_i = \mathbf{x}_{1i} - \delta_{1i}$, and by substituting the right-hand sides of these equations, we can eliminate the latent variables from Eqs. **1–3**, resulting in Eqs. **10–12**:

$$\mathbf{y}_{1i} = \alpha_\eta + \mathbf{B}\mathbf{y}_{1i} + \mathbf{\Gamma}\mathbf{x}_{1i} + \varepsilon_{1i} - \mathbf{B}\varepsilon_{1i} - \mathbf{\Gamma}\delta_{1i} + \zeta_i, \qquad [10]$$

$$\mathbf{y}_{2i} = \alpha_{\mathbf{y}_2} + \mathbf{\Lambda}_{\mathbf{y}2}\mathbf{y}_{1i} - \mathbf{\Lambda}_{\mathbf{y}2}\varepsilon_{1i} + \varepsilon_{2i}, \text{ and} \qquad [11]$$

$$\mathbf{x}_{2i} = \alpha_{\mathbf{x}_2} + \mathbf{\Lambda}_{\mathbf{x}2}\mathbf{x}_{1i} - \mathbf{\Lambda}_{\mathbf{x}2}\delta_{1i} + \delta_{2i}. \qquad [12]$$

This simple substitution transforms a system of latent and observed variables into a system of observed variables and a more complex error structure. The end result is a system of equations that consists of observed variables and composite disturbance terms. Generally, most of the equations in this model cannot be consistently estimated with ordinary least squares regression, because one or more of the variables in the composite disturbances correlate with the right-hand side variables. An instrumental variable (IV) approach is used to take account of this problem.

We illustrate an IV using a simple regression equation, $y_{1i} = \alpha_1 + \beta_{11}x_{1i} + \varepsilon_{1i}$, where $COV(x_{1i}, \varepsilon_{1i}) \neq 0$. The ordinary least squares (OLS) estimator of this equation would result in an inconsistent and biased estimator of β_{11}. However, if there were another variable (x_{2i}) that had a moderate to strong correlation with x_{1i} but no correlation with ε_{1i}, then we could regress x_{1i} on x_{2i} and form the predicted value of $\widehat{x}_{1i}$. The OLS regression of y_{1i} on $\widehat{x}_{1i}$ would provide a consistent estimator of β_{11}. The resulting estimator would be the IV estimator where x_{2i} is the IV.

In the general latent variable SEM, the researcher chooses the first equation to estimate from the transformed system that eliminates the latent variables (Eqs. **10–12**). The researcher then chooses the MIIV for this equation. The MIIVs are observed variables that are part of the model and based on the model structure, should not correlate with the disturbance of the equation but do have a significant association with the problematic explanatory variable(s). The MIIV selection contrasts with the more typical applications of IV, where the selection of IVs is somewhat ad hoc. In the MIIV-2SLS latent variable approach in the work by Bollen (24), the model comes first, and the observed variables that satisfy the conditions of IVs follow from the model structure. Bollen and Bauer (26) describe an algorithm that finds the MIIVs for a specific model structure, or the researcher can determine the MIIVs by inspection of the model. After the MIIVs are determined for an equation, a 2SLS estimator is applied. The estimator is consistent, asymptotically unbiased, and asymptotically normally distributed, and an estimate of the asymptotic covariance matrix of the coefficients is readily available. As stated above, this estimator seems more robust to misspecified structures than the MIM estimators and could prove helpful when structural misspecification is likely. The works by Bollen (24, 25) have the formula and more technical details.

Both the MIM and MIIV estimators are justified based on their asymptotic properties. Asymptotic refers to large sample properties. It is impossible to give a single sample size N at which these properties take hold, because they depend on many things, such as the complexity of the model and the strength of relationships. However, most SEM researchers would consider an N less than 100 small for all but simple models and an N of 1,000 as sufficiently large for most models. However, even here, there can be exceptions depending on the estimator, distribution of variables, number of parameters, and strength of relationships.

Model fit. After estimation, the researcher turns to assessing the fit of the model. It is convenient to describe model fit as having two aspects. One concerns the component fit of the model. Overall model fit is the second aspect of fit. In evaluating the component fit of the model, the MIM and MIIV estimators are similar in that both produce asymptotic SEs of the coefficients and intercepts so that the researcher can test individual parameters for statistical significance. In addition, groups of parameters can be tested using, for example, Wald tests. The MIIV-2SLS estimator also has a test of individual equations for all equations that are overidentified in that they have more than the minimum number of MIIVs needed for identification. The Sargan test (27) is a good test to use for this purpose. Its null hypothesis is that all MIIVs are valid in that they are uncorrelated with the disturbance of the equation. The alternative hypothesis is that at least one MIIV is correlated with the disturbance, signifying a problem with the model structure.

Overall fit is a second way to gauge model fit. When using ML and other full information estimators, a χ^2 test of overall fit is generally available. For instance, for the ML estimator, the test statistic, T_{ML}, is $(N-1)F_{ML}$, where F_{ML} is evaluated at the final estimates $(\widehat{\theta})$ of the parameters. When the assumptions of the ML estimator are satisfied, T_{ML} follows an asymptotic χ^2 distribution with degrees of freedom of $df = \left(\dfrac{1}{2}\right)P_z(P_z + 3) - t$, where P_z is the number of observed variables and t is the number of free parameters estimated in the model. The null hypothesis of this χ^2 test is $H_o: \mu = \mu(\theta)$ and $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$. A statistically significant test statistic casts doubt on the implied moment structure and the model that gave rise to it. A nonsignificant test statistic is consistent with the model structure. One aspect of overall fit to be mindful of is the existence of equivalent models. These equivalent models are models that have different structures but the same overall model fit (28). For instance, a factor analysis model that has four indicators depending on a single latent variable with uncorrelated errors has the same overall fit as the exact same model, except that the first indicator causes rather than depends on the latent variable. Subject matter expertise must be relied on to discriminate between such models, because their χ^2 fit and degrees of freedom are the same.

In practice, $H_o: \mu = \mu(\theta)$ and $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$ is too strict for most models, because the test is intolerant of even slight misspecifications, and hence, in situations with sufficient statistical power (e.g., when N is large), the null hypothesis is nearly always rejected. This finding has given rise to a wide variety of fit indices that supplement the χ^2 test and to controversy about the best ways to assess overall fit. Works by Bollen and Long (29) and Hu and Bentler (30) have rationales, formulas, and discussions of these additional fit indices. Current practice is to report the χ^2 test statistic, degrees of freedom, and P value along with several other fit indices when assessing a model's overall fit. Alternative methods for model evaluation are available as well (31). If the overall fit or component fit of a model is judged to be inadequate, then this finding suggests the need to respecify the model.

Respecification. It is common that researchers' first specified model is found wanting and in need of modification. Respecifications refer to revisions of an initial model. These revisions can range from minor (e.g., introducing a secondary path) to major (e.g., changing number of latent variables and their relationships). Subject matter expertise is the best guide to respecifica-

tions. A researcher who had considered alternative structures before choosing the tested one could turn again to these structures to see if fit improves.

Empirical procedures also exist to suggest respecifications of the initial model. For the MIM estimators, the Lagrangian Multiplier (modification indexes) test statistic is the most popular empirical guide. It is an estimate of the decrease in the χ^2 test statistic that would result by freeing a previously fixed parameter in the model. It considers freeing just a single parameter at a time, which can be misleading when multiple modifications are required. Reliance on the modification index can lead to the introduction of nonsense parameters and difficulty in recovering the true generating model (32, 33). The modification index is most useful if used in conjunction with substantive expertise. Other diagnostic tools include covariance residual analysis (34) and individual case diagnostics (35, 36). There also are recent developments in exploratory SEM that might prove useful (37). The MIIV-2SLS test of IVs for an equation is another alternative (38).

Empirically based respecifications move the analysis more in the direction of exploratory than confirmatory analysis. The significance tests from such an approach must be interpreted cautiously, because they do not consider that modifications were made in response to patterns found in a specific sample and might not replicate in a new sample from the same population. When possible, replication of the models on fresh data provides some protection against capitalizing on a specific sample. Another desirable strategy is to start the analysis with several plausible models for the same data and compare their fit in a more confirmatory manner. This method contrasts with starting with a single model and modifying it several times in response to its empirical fit. We also note that a researcher approaching modeling from a purely confirmatory perspective would not necessarily respecify a model. The researcher might be more interested in determining whether the model fits or not and regardless of the outcome, would not attempt to modify the original model; the researcher would just report the results of the initial specification.

Illustrations

In this section, we provide several hypothetical examples of SEMs as a way of illustrating the types of structures that are part of this model.

Illustration 1. The first illustration is a confirmatory factor analysis (CFA), a special case of an SEM that only includes the relationship between a latent variable (or factor) and its measures. This example is taken from Friedman et al. (39); it concerns childhood sleeping patterns and executive functioning at late adolescence and includes three latent variables, each representing a different executive function: inhibiting, updating, and shifting. Each of these executive functioning latent variables is measured by three manifest variables derived from performances

on tasks designed to tap into the concepts represented by the latent variables. Using any one of these measures as a covariate in a regression would lead to bias because of the measurement error in each indicator. The benefit of using a latent variable for each of these three executive functioning skills over a single measure is that the degree of measurement error can be estimated and controlled in the model.

The first model is a measurement model. Measurement models can inform the researcher about the degree of measurement error in each indicator and whether all of the indicators behave as hypothesized in relation to the latent variable. Furthermore, the measurement model permits us to determine the closeness of association of different latent variables after we take account of measurement error and see whether the latent variables are empirically separable from each other. Here, we have three correlated latent variables representing different aspects of executive functioning, each with a latent variable measured with three indicators. The latent variables from this measurement model will appear as endogenous (dependent) variables in Illustration 2.

The latent variable for the executive function, inhibiting, represents the concept of individual restraint while holding back a dominant response, and it is measured here by the three variables: antisac, stop-signal, and stroop. The updating executive function latent variable represents the process of discarding less relevant information as an individual acquires more relevant information. Its three measures are keep track, letter, and S2back. Finally, the executive function of shifting is the transitioning from one task to another efficiently and is measured by the variables number, color, and category.

Either equations or path diagrams can represent the relationship between the measures and latent variables. For instance, the equations for the three indicators of inhibiting are (Eqs. **13–15**)

$$
\begin{aligned}
x_1 &= \xi_1 + \delta_1, \\
x_2 &= \lambda_{21}\xi_1 + \delta_2, \; and, \\
x_3 &= \lambda_{31}\xi_1 + \delta_3
\end{aligned}
\qquad \textbf{[13-15]}
$$

where ξ_1 is the latent variable for inhibiting and λ_{21} and λ_{31} are the coefficients (factor loadings) that give the expected difference in x_2 and x_3, respectively, for a one-unit difference in ξ_1. The coefficient of the x_1 is implicitly set to one to provide a metric for the latent variable ξ_1. An alternative way to scale the latent variable is to set its variance to one rather than a factor loading. The equations for the measures of updating and shifting are similar but include their respective indicators and latent variables.

Fig. 1 is an alternative way to represent the CFA model—with a path diagram that is equivalent to the equations. The path diagram consists of three latent variables and nine observed variables. The three unobserved or latent variables mentioned
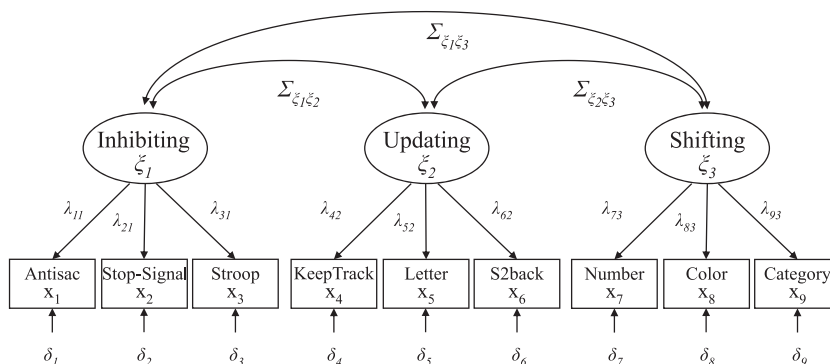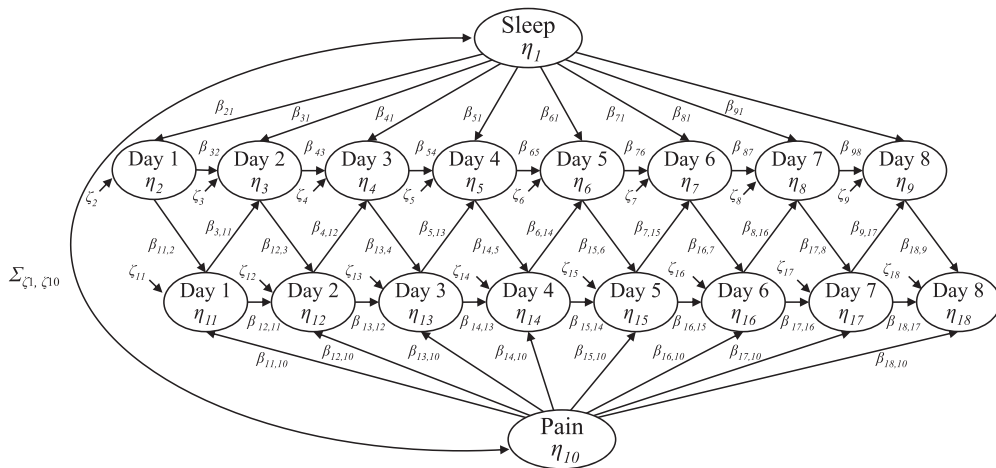

Fig. 1. Confirmatory factor analysis of inhibiting, updating, and shifting behaviors.

above, inhibiting, updating, and shifting, are represented by ovals. Curved, double-headed arrows between each pair of these latent variables allow for covariances between the latent variables in recognition of their likely association with each other.

The indicators, also called observed or manifest variables, are represented in the diagram by the nine boxes and include, for example, antisac, stop-signal, and stroop for the first latent variable of inhibiting. The single-headed straight arrows that originate with the latent variables and terminate in the indicators represent direct relationships from the latent to the observed variables. The small arrow at the bottom of these boxes, whose origin is not pictured, symbolizes the error's effect on the indicator. The error or disturbance includes all other variables that influence the indicator besides its respective latent variable. Each observed variable has a single error term. In the factor analysis tradition, this error term has two parts, with the first part being purely random error and the second part being the specific component. The specific component is a systematic variable that is specific to the particular measure. For simplification, we ignore this distinction and use the term error or disturbance to refer to this variable. Furthermore, the lack of any double-headed curved arrows between these errors identifies that observed variables have no additional covariation beyond that caused by their shared latent variable.

The path diagram in Fig. 1 represents pictorially the relationship between the three latent variables of executive function. We see from the diagram that three latent variables are allowed to covary, each latent variable has effects on only its three indicators and none of the errors from these observed variables covaries with other errors. Of course, more elaborate models are possible where dictated by substantive hypotheses.

Illustration 2. Illustration 2 is an extension of Fig. 1 by including this model with additional variables in a longitudinal application of the SEM framework known as a latent curve model (LCM). This model will test the hypothesis that a child's sleep problems from ages 4 to 16 y will be useful in predicting his/her executive functioning at age 17 y. The purpose of the LCM is to determine an individual trajectory of sleep problems for each child based on the reported number of sleep problems at 11 different time points in the development of a child. These 11 repeated measures of sleep problems serve as a window into the sleep problem trajectory of a child between the ages of 4 and 16 y. Ultimately, the parameters governing each individual child's sleep trajectory

are then used as a predictor of the executive functioning of that individual child at age 17 y. In practice, two unique parameters, the intercept and slope for each child, are used to predict their executive functioning at age 17 y. These two correlated parameters are called growth parameters, and the intercept for the sleep problem trajectory is at age 4 y.

The path diagram in Fig. 2 shows the three latent variables (inhibiting, updating, and shifting) from Fig. 1 in the lower part of the diagram. The upper part of Fig. 2 includes 11 repeated measures of sleep problems at different ages (4–16 y) of children. Each child can have a different starting point or intercept, and this variable is captured by the random intercept latent variable labeled intercept. Each child also can have a different random rate of change in sleep problems, and the random slopes are represented by the latent slope variable. Together, the intercepts and slopes represent the growth curves of sleep problems for the children in the sample. The model in Fig. 2 shows the intercept and slopes with direct effects on the three executive functioning latent variables of inhibiting, updating, and shifting. The errors of these latter three latent variables are allowed to correlate, which is represented by the curved double-headed arrows that connect them. The random intercepts and random slopes also are allowed to correlate, which is represented by the curved double-headed arrow that connects them.

This model illustrates an SEM approach to growth curve models, where the parameters of the growth curves have effects on other outcome variables (40). To avoid overcrowding the diagram, the Greek letters representing the latent variables and coefficients are not shown. However, the ovals and boxes clearly show which variables are latent and observed. Associated with each path is a single-headed arrow from one variable to another. The coefficients give the expected impact of the variable at the base of the arrow on the variable at the head of the arrow while holding constant any other variables affecting that target variable.

There are several benefits of using an LCM to predict executive functioning of children at age 17 y. It allows each child to have a unique trajectory of sleep problems described by their growth parameters. Researchers can test the functional form of the growth trajectory of sleep problems to find the best representation of the growth trajectory of sleep disruptions. Sometimes, nonlinear growth trajectories are better representations of the underlying growth trajectory, and researchers can fit these as well. This method also allows for the usual overall fit indices

Fig. 2. Latent curve model of childhood sleep problems on inhibiting, updating, and shifting behaviors at adolescence.

Fig. 3. Structural equation model of repeated measures of sleep and pain over 8 d.

from the SEM framework to be incorporated in addition to individual component fit measures to assess how well this model fits the data. Furthermore, it allows for estimation of the effect that the trajectory of sleep problems has on the three executive functioning tasks at age 17 y while controlling for the measurement error in inhibiting, updating, and shifting. This model also allows for an assessment of the correlation of the disturbances of inhibiting, updating, and shifting after their prediction by the sleep problem trajectory parameters.

Illustration 3. The third illustration of SEMs originates in the literature on the relationship of sleep and pain (Fig. 3). The first series of day 1–8 latent variables is the amount of sleep for each person over 8 d. The latent sleep variable that influences all 8 d represents the stable component of sleep across the 8 d.

Similarly, the second series of day 1–8 latent variables is the daily reports of pain. The latent pain variable that influences all of the daily pain reports is a stable component of the pain reports. The higher-order latent sleep (η_1) and latent pain (η_{10}) variables are allowed to correlate. The daily sleep and pain latent variables have an autoregressive relation and some cross-lagged effects from sleep to pain and pain to sleep with a 1-d lag. Prior values often influence current values of the same variable, and this type of autoregressive relationship is included (e.g., sleep on day 2 is regressed on sleep from day 1, sleep from day 3 is regressed on the sleep from day 2, etc.). A pure cross-lagged effect involves two variables and is when a previous measure of one variable predicts a subsequent measure of the other variable. Not shown in the diagram are the measures of the sleep and pain daily latent variables.

This illustration is an adaptation of the SEM from the work by Edwards et al. (41). Notice that this model permits stable individual differences in the latent sleep and pain variables that hold across the days, but it also allows each day's sleep to depend on the prior day's sleep and pain and vice versa. The model hypothesizes that the number of hours that an individual sleeps in one night predicts the daily reports of pain the following day. This model also hypothesizes that both the previous night's sleep and the daily reports of pain during the day are predictors of the amount of sleep that the individual will have that night. Furthermore, this model hypothesizes that the pain reports of a previous day help to predict the pain reports of the current day.

This model is a step to trying to untangle the finding that pain and sleep seem to have a nonrecursive relationship (41). To illustrate this model in equation form, consider the equation for the latent sleep variable for day 2 (η_3) (Eq. **16**),

$$\eta_3 = \beta_{31}\eta_1 + \beta_{32}\eta_2 + \beta_{3,11}\eta_{11} + \zeta_3, \qquad \textbf{[16]}$$

or the equation for the day 5 pain variable (η_{15}) (Eq. **17**),

$$\eta_{15} = \beta_{15,10}\eta_{10} + \beta_{15,6}\eta_6 + \beta_{15,14}\eta_{14} + \zeta_{15}. \qquad \textbf{[17]}$$

In a similar fashion, all of the equations could be read from the path diagram.

The whole model could be estimated simultaneously to answer a number of questions about the sleep–pain relationship. Are there individual differences in the overtime levels of pain or sleep? What are the relative strengths of the autoregressive relations among the daily sleep and daily pain variables after controlling for individual differences in sleep and pain higher-order latent variables? Are the lagged effects from sleep to pain and pain to sleep both significant? These questions and others would be addressed while controlling for the other variables in the model and taking account of measurement error.

Conclusions.
Among the many challenges facing studies that quantify behavior are the measurement errors in the behavioral and other measures, the need to formulate multiple equations to characterize the behavioral system, and the desire to understand the direct and indirect effects of variables as they work their way through the equation system. Latent variable SEMs provide the tools to address these challenges. They have the capability to allow quantification and testing of the hypothesized relationships among latent and observed variables. They provide tests of the consistency and plausibility of the assumed model compared with the observed data. Additionally, they enable a researcher to analyze direct as well as mediated relationships. Although SEMs cannot replace sound substantive knowledge in formulating a model, they can provide information on the match between the model and the data, and they do provide tools to further trace the implications of this structure.

1. Wright S (1934) The method of path coefficients. *Ann Math Stat* 5:161–215.
2. Wright S (1918) On the nature of size factors. *Genetics* 3:367–374.
3. Blalock HM (1964) *Causal Inferences in Nonexperimental Research* (University of North Carolina Press, Chapel Hill, NC).
4. Duncan OD (1966) Path analysis: Sociological examples. *Am J Sociol* 72:1–16.
5. Jöreskog KG (1973) A general method for estimating a linear structural equation system. *Structural Equation Models in the Social Sciences*, eds Goldberger AS, Duncan OD (Seminar Press, New York), pp 85–112.
6. Bollen KA (1998) Structural equation models. *Encyclopedia of Biostatistics*, eds Armitage P, Colton T (Wiley, New York), pp 4363–4372.

7. Matsueda RL (2012) Key advances in the history of structural equation modeling. *Handbook of Structural Equation Modeling*, ed Hoyle R (Guilford, New York).
8. Jöreskog KG, Sórbom D (2001) *LISREL 8: User's Reference Guide* (Scientific Software International, Uppsala, Sweden).
9. Muthén LK, Muthén BO (1998) *Mplus User's Guide* (Muthén and Muthén, Los Angeles), 5th Ed.
10. Arbuckle JL (2009) *Amos 18 User's Guide* (Amos Development Corporation, Crawfordville, FL).
11. Bentler PM (1995) *EQS Structural Equations Program Manual* (Multivariate Software, Encino, CA).
12. Fox J (2006) Structural equation modeling with the SEM package in R. *Struct Equ Modeling* 13:465–486.
13. Bollen KA (2002) Latent variables in psychology and the social sciences. *Annu Rev Psychol* 53:605–634.
14. Muthén B (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49:115–132.
15. Bollen KA (1989) *Structural Equations with Latent Variables* (Wiley, New York).
16. Davis WR (1993) The FC1 rule of identification for confirmatory factor analysis: A general sufficient condition. *Sociol Methods Res* 21:403–437.
17. Bollen KA, Bauldry S (2010) Model identification and computer algebra. *Sociol Methods Res* 39:127–156.
18. Browne MW (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *Br J Math Stat Psychol* 37:62–83.
19. Satorra A (1990) Robustness issues in structural equation modeling: A review of recent developments. *Qual Quant* 24:367–386.
20. Bollen KA, Stine RA (1990) Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociol Methodol* 20:115–140.
21. Bollen KA, Stine RA (1992) Bootstrapping goodness-of-fit measures in structural equation models. *Sociol Methods Res* 21:205–229.
22. Arminger G, Schoenberg R (1989) Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika* 54:409–425.
23. Satorra A, Bentler PM (1994) Corrections to test statistics and standard errors in covariance structure analysis. *Latent Variables Analysis: Applications for Developmental Research*, eds von Eye A, Clogg CC (Sage, Newbury Park, CA), pp 399–419.
24. Bollen KA (1996) An alternative 2SLS estimator for latent variable models. *Psychometrika* 61:109–121.
25. Bollen KA (2001) Two-stage least squares and latent variable models: Simultaneous estimation and robustness to misspecifications. *Structural Equation Modeling: Present and Future, A Festschrift in Honor of Karl Jöreskog*, eds Cudeck R, Du Toit S, Sörbum D (Scientific Software, Lincoln, IL), pp 119–138.
26. Bollen KA, Bauer DJ (2004) Automating the selection of model-implied instrumental variables. *Sociol Methods Res* 32:425–452.
27. Sargan JD (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26:393–415.
28. Hershberger SL (2006) The problem of equivalent structural models. *Structural Equation Modeling: A Second Course*, eds Hancock GR, Mueller RO (IAP, Greenwich, CT.), pp 13–41.
29. Bollen KA, Long JS, eds (1993) *Testing Structural Equation Models* (Sage, Newbury Park, CA).
30. Hu L, Bentler PM (1998) Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol Methods* 3:424–453.
31. Saris WE, Satorra A, van der Veld WM (2009) Testing structural equation models or detections of misspecifications? *Struct Equ Modeling* 16:561–582.
32. MacCallum LC (1986) Specification searches in covariance structure modeling. *Psychol Bull* 100:107–120.
33. MacCallum RC (2003) Working with imperfect models. *Multivariate Behav Res* 38:113–139.
34. Costner HL, Schoenberg R (1973) Diagnosing indicator ills in multiple indicator models. *Structural Equation Models in the Social Sciences*, eds Goldberger AS, Duncan OD (Seminar Press, New York), pp 167–200.
35. Bollen KA, Arminger G (1991) Observational residuals in factor analysis and structural equation models. *Sociol Methodol* 21:235–262.
36. Cadigan NG (1995) Local influence in structural equations models. *Struct Equ Modeling* 2:13–30.
37. Marsh HW, et al. (2009) Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Struct Equ Modeling* 16:439–476.
38. Kirby JB, Bollen KA (2009) Using instrumental variable (IV) tests to evaluate model specification in latent variable structural equation models. *Sociol Methodol* 39:327–355.
39. Friedman NP, Corley RP, Hewitt JK, Wright KP, Jr. (2009) Individual differences in childhood sleep problems predict later cognitive executive control. *Sleep* 32:323–333.
40. Bollen KA, Curran PJ (2004) *Latent Curve Models: A Structural Equation Perspective. Wiley Series in Probability and Mathematical Statistics* (Wiley, New York).
41. Edwards RR, Almeida DM, Klick B, Haythornthwaite JA, Smith MT (2008) Duration of sleep contributes to next-day pain report in the general population. *Pain* 137:202–207.