SEM Basics:

A Supplement to Multivariate Data Analysis

Multivariate Data Analysis

Pearson Prentice Hall Publishing

Table of Contents

SEM Basics:

A Supplement to
Multivariate Data Analysis

LEARNING OBJECTIVES

In the course of completing this supplement, you will be introduced to the following:

- The basics of estimating path coefficients based on the specified path model.
- Determination of the direct and indirect effects implied in a path model, plus determination whether they can be characterized as causal or noncausal.
- Some additional absolute fit indices used in certain situations.
- Specification of the path model as a series of equations for both the structural model and the measurement model.
- Use of LISREL notation to represent these equations and the relationships in the path model.
- Testing for mean differences between latent constructs in different groups.
- Item parceling to reduce the number of items per construct.
- Assessment of measurement bias by introduction of an additional latent construct.
- Estimation of moderating effects for continuous multi-item constructs.

PREVIEW

This supplement to the text *Multivariate Data Analysis* provides additional coverage of some basic concepts that are the foundations for structural equation modeling (SEM). While there is considerable coverage of the technique in the text, the authors felt that readers may benefit from further review of certain topics not covered in the text, but issues addressed by many researchers. Moreover, there is a more comprehensive discussion of the notation used in SEM, particularly those associated with LISREL. There will be some overlap with material in the chapters so as to fully integrate the concepts.

The supplement is not intended to be a comprehensive "primer" on all of the SEM topics not covered in the test, but only those selected issues that may be encountered in the

course of behavioral research. We encourage readers to complement this supplement with other treatments and texts on these concepts as needed.

The supplement focuses on three broad areas related to SEM. The first area covers some of the fundamental concepts related the "basics" of path models – estimation of the path estimates and determining and interpreting direct and indirect effects. The second area is specification of the SEM model in more formal terms. The primarily involves discussion of what is termed LISREL notation. This involves the notation used in the LISREL software program which has become a common method of describing the relationships in both the structural and measurement models. Several examples, including the HBAT CFA and structural models, are used to illustrate how those models can be expressed in this notation. Included in the discussion are also some techniques to accomplish specialized tasks in LISREL, as well as a brief introduction to AMOS, another popular SEM software package. Finally, some more advanced topics are discussed to provide the user an introduction into some of the more complex, but often used, techniques available in SEM analyses.

FUNDAMENTALS OF STRUCTURAL EQUATION MODELING

The use of SEM is predicated on a strong theoretical model by which latent constructs are defined (measurement model) and these constructs are related to each other through a series of dependence relationships (structural model). The emphasis on strong theoretical support for any proposed model underlies the confirmatory nature of most SEM applications.

But many times overlooked is exactly how the proposed structural model is translated into structural relationships and how their estimation is interrelated. Path analysis is the process wherein the structural relationships are expressed as direct and indirect effects in order to facilitate estimation. The importance of understanding this process is not so that the research can understand the estimation process, but instead to understand how model specification (and respecification) impacts the entire set of structural relationships. We will first illustrate the process of using path analysis for estimating relationships in SEM analyses. Then we will discuss the role that model specification has in defining direct and indirect effects and classification of effects as causal versus spurious. We will see how this designation impacts the estimation of structural model.

ESTIMATING RELATIONSHIPS USING PATH ANALYSIS

What was the purpose of developing the path diagram? Path diagrams are the basis for path

analysis, the procedure for empirical estimation of the strength of each relationship (path) depicted in the path diagram. Path analysis calculates the strength of the relationships using only a correlation or covariance matrix as input. We will describe the basic process in the following section, using a simple example to illustrate how the estimates are actually computed.

Identifying Paths

The first step is to identify all relationships that connect any two constructs. Path analysis enables us to decompose the simple (bivariate) correlation between any two variables into the sum of the compound paths connecting these points. The number and types of compound paths between any two variables are strictly a function of the model proposed by the researcher.

A compound path is a path along the arrows of a path diagram that follow three rules:

1. After going forward on an arrow, the path cannot go backward again; but the path can go backward as many times as necessary before going forward.

2. The path cannot go through the same variable more than once.

3. The path can include only one curved arrow (correlated variable pair).

When applying these rules, each path or arrow represents a path. If only one arrow links two constructs (path analysis can also be conducted with variables), then the relationship between those two is equal to the parameter estimate between those two constructs. For now, this relationship can be called a direct relationship. If there are multiple arrows linking one construct to another as in X → Y → Z, then the effect of X on Z seem quite complicated but an example makes it easy to follow:

The path model below portrays a simple model with two exogenous constructs (X_1 and X_2) causally related to the endogenous construct (Y_1). The correlational path A is X_1 correlated with X_2, path B is the effect of X_1 predicting Y_1, and path C shows the effect of X_2 predicting Y_1.
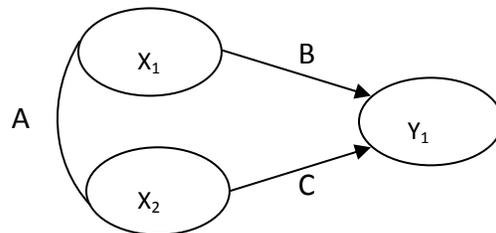
The value for Y_1 can be stated simply with a regression-like equation:

$$Y_1 = b_1 X_1 + b_2 X_1$$

We can now identify the direct and indirect paths in our model. For ease in referring to the paths, the causal paths are labeled A, B, and C.

Direct Paths	Indirect Paths
A = X_1 to X_2	
B = X_1 to Y_1	AC = X_1 to Y_1
C = X_2 to Y_1	AB = X_2 to Y_1

Estimating the Relationship

With the direct and indirect paths now defined, we can represent the correlation between each construct as the sum of the direct and indirect paths. The three unique correlations among the constructs can be shown to be composed of direct and indirect paths as follows:

$$Corr_{X1\ X2} = A$$
$$Corr_{X1\ Y1} = B + AC$$
$$Corr_{X2\ Y1} = C + AB$$

First, the correlation of X_1 and X_2 is simply equal to A. The correlation of X_1 and Y_1 ($Corr_{X1,Y1}$) can be represented as two paths: B and AC. The symbol B represents the direct path from X_1 to Y_1, and the other path (a compound path) follows the curved arrow from X_1 to X_2 and then to Y_1. Likewise, the correlation of X_2 and Y_1 can be shown to be composed of two causal paths: C and AB.

Once all the correlations are defined in terms of paths, the values of the observed correlations can be substituted and the equations solved for each separate path. The paths then represent either the causal relationships between constructs (similar to a regression coefficient) or correlational estimates.

Assuming that the correlations among the three constructs are as follows: $Corr_{X1\ X2}$ = .50, $Corr_{X1\ Y1}$ = .60 and $Corr_{X2\ Y1}$ = .70, we can solve the equations for each correlation (see below) and estimate the causal relationships represented by the coefficients b_1 and b_2. We know that A equals .50, so we can substitute this value into the other equations. By solving these two equations, we get values of B(b_1) = .33 and C(b_2) = .53. This approach enables path analysis to solve for any causal relationship based only on the correlations among the constructs and the specified causal model.

<div align="center">

Solving for the Structural Coefficients

.50 = A

.60 = B + AC

.70 = C + AB

Substituting A = .50

.60 = B + .50C

.70 = C + .50B

Solving for B and C

B = .33

C = .53

</div>

As you can see from this simple example, if we change the path model in some way, the causal relationships will change as well. Such a change provides the basis for modifying the model to achieve better fit, if theoretically justified.

With these simple rules, the larger model can now be modeled simultaneously, using correlations or covariances as the input data. We should note that when used in a larger model, we can solve for any number of interrelated equations. Thus, dependent variables in one relationship can easily be independent variables in another relationship. No matter how large the path diagram gets or how many relationships are included, path analysis provides a way to analyze the set of relationships.

UNDERSTANDING DIRECT AND INDIRECT EFFECTS

While path analysis plays a key role in estimating the effects represented in a structural model, it also provides additional insight into not only the direct effects of one construct versus another, but all of the myriad set of indirect effects between any two constructs. While direct effects can always be considered causal if a dependence relationship is specified, indirect effects require further examination to determine if they are causal (directly attributable to a dependence relationship) or non-causal (meaning that they represent relationship between constructs, but it cannot be attributed to a specific causal process).

Identification of Causal versus Non-Causal effects

The prior section discussed the process of identifying all of the direct and indirect effects between any two constructs by a series of rules for compound paths. Here we will discuss how to categorize them into causal versus non-causal and then illustrate their use in understanding the implications of model specification.

An important question is: Why is the distinction important? The parameter estimates are made without any distinction as described above. But the estimated parameters in the

structural model reflect only the direct effect of one construct on another. What about all of the indirect effects, which can be substantial? Just because a construct is not directly related to another construct does not mean that there is no impact. Thus, we need a method to distinguish between these myriad types of indirect effects and be able to understand if we can infer any dependence relationship (causal) attributable to them even though they are indirect.

Assume we are identifying the possible effects of A → B. Causal effects are of two types: a direct causal effect (A → B) or an indirect causal effect (A → C → B). In either the direct or indirect compound path, only dependence relationships are present and the direction of the relationships is never reversed. Non-causal (sometimes referred to as spurious effects) can arise from three conditions: common effects (C → A and C →B), correlated effects (C → A and C is correlated with B) and reciprocal effects (A → and B → A).

Decomposing Effects into Causal versus Noncausal

We will use a simple example of four constructs all related to each other as shown in the following path diagram.
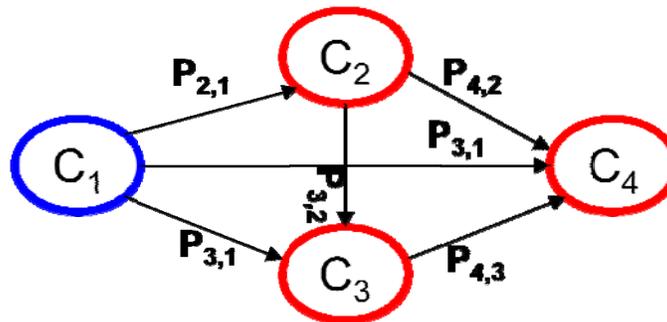
We can see that there are only direct dependence relationships in this example (i.e., no correlational relationships among exogenous constructs). So we might presuppose that all of the effects (direct and indirect) would be causal as well. But as we will see, that is not the case. The total set of effects for each relationship is shown in the table below.

C1 → C2. Let's start with the simple relationship between C_1 and C_2. Using the rules for identifying compound paths earlier, we can see that there is only one possible effect – the direct effect of C_1 on C_2. There are no possible indirect or noncausal effects, so the path estimate for $P_{2,1}$ represents the total effects of C_1 on C_2.

C1 → C3. The next relationship is C_1 with C_3. Here we can see two effects: the direct effect ($P_{3,1}$) and the indirect effect ($P_{3,2}$ x $P_{2,1}$). Since the direction of the paths never reverses in the indirect effect, it can be categorized as causal. So the direct and indirect effects are both causal effects.

C2 → C3. This relationship introduces the first noncausal effects we have seen. There is the direct effect of $B_{3,2}$, but there is also the noncausal effect (due to common cause) seen in $B_{3,1}$ x $B_{2,1}$. Here we see the result of two causal effects creating a noncausal effect since they both originate from a common construct (C_1 → C_2 and C_1 → C_3).

C1 → C4. In this relationship we will see the potential for numerous indirect causal effects in addition to direct effects. In addition to the direct effect ($B_{4,1}$), we see three other indirect effects that are also causal: $B_{4,2}$ x $B_{2,1}$; $B_{4,3}$ x $B_{3,1}$; and $B_{4,3}$ x $B_{3,2}$ x $B_{2,1}$.

C3 → C4. This final relationship we will examine has only one causal effects ($B_{4,2}$), but four different noncausal effects, all a result of C_1 or C_2 acting as common causes. The two noncausal effects associated with C_1 are $B_{4,1}$ x $B_{3,1}$ and $B_{4,1}$ x $B_{2,1}$ x $B_{3,2}$. The two other noncausal effects are associated with C_2 ($B_{4,2}$ x $B_{3,2}$ and $B_{4,2}$ x $B_{2,1}$ x $B_{3,1}$).

The remaining relationship is C_2 → C_4. See if you can identify the causal and noncausal effects. Hint: There are all three types of effects – direct and indirect causal effects and noncausal effects as well.

	Effects		
Relationship	**Direct (Causal)**	**Indirect (Causal)**	**Indirect (Noncausal)**
C_1 → C_2	$P_{2,1}$	None	None
C_1 → C_3	$P_{3,1}$	$P_{3,2}$ x $P_{2,1}$	None
C_2 → C_3	$P_{3,2}$	None	$P_{3,1}$ x $P_{2,1}$
C_1 → C_4	$P_{4,1}$	$P_{4,2}$ x $P_{2,1}$ $P_{4,3}$ x $P_{3,1}$ $P_{4,3}$ x $P_{3,2}$ x $P_{2,1}$	None
C_3 → C_4	$P_{4,3}$	None	$P_{4,1}$ x $P_{3,1}$ $P_{4,1}$ x $P_{2,1}$ x $P_{3,2}$ $P_{4,2}$ x $P_{3,2}$ $P_{4,2}$ x $P_{2,1}$ x $P_{3,1}$

Calculating Indirect Effects

In the previous section we discussed the identification and categorization of both direct and indirect effects for any pair of constructs. The next step is to calculate the amount of the effect based on the path estimates of the model. Assume this path model with estimates as follows:
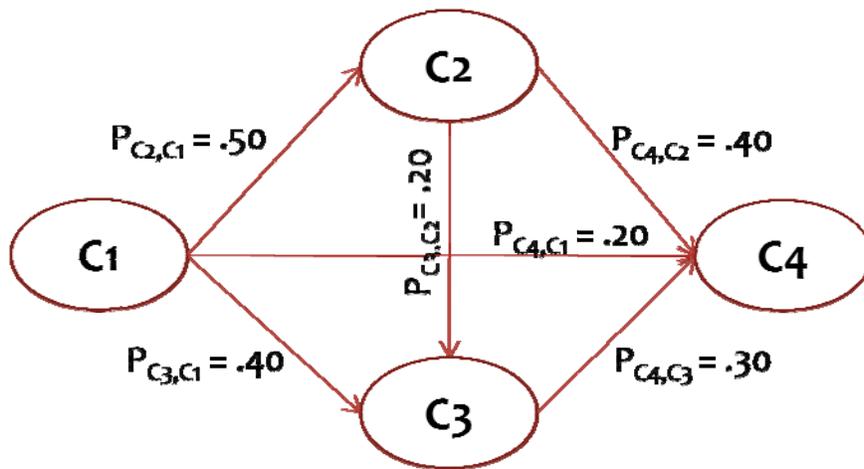
For the relationship between C1 → C4, there are both direct and indirect effects. The direct effects are shown directly by the path estimate $P_{C4,C1} = .20$. But what about the indirect effects? How are they calculated?

The size of an indirect effect is a function of the direct effects that make it up. SEM software typically produces a table showing the total of the indirect effects implied by a model. But to see the size of each effect you can compute them by multiplying the direct effects in the compound path.

For instance, one of the indirect effects for C1 → C4 is $P_{C2,C1} \times P_{C4,C2}$. We can then calculate the amount of this effect as .50 × .40 = .20. Thus, the complete set of indirect effects for C1 → C4 can be calculated as follows:

$P_{C2,C1} \times P_{C4,C2} =$ .50 × .40 = .20

$P_{C3,C1} \times P_{C4,C3} =$ .40 × .30 = .12

$P_{C2,C1} \times P_{C3,C2} \times P_{C4,C3} =$ .50 × .20 × .30 = .03

Total Indirect Effects = .20 + .12 + .03 = .35

So in this example, the total effect of C1 → C4 equals the direct and indirect effects, or .20 + .35 = .55. It is interesting to note that in this example the indirect effects are greater than the direct effects.

Impact of Model Respecification

The impact of model respecification on both the parameter estimates and the causal/noncausal effects can be seen in our example as well. Look back at the $C_3 \rightarrow C_4$ relationship. What happens if we eliminate the C1 $\rightarrow$ C4 relationship? Does it impact the C3 $\rightarrow$ C4 relationship in any way? If we look back at the indirect effects, we can see that two of the four noncausal effects would be eliminated ($B_{4,1} \times B_{3,1}$ and $B_{4,1} \times B_{2,1} \times B_{3,2}$). How would this impact the model? If these effects were substantial but eliminated when the $C_1 \rightarrow C_4$ path was eliminated, then most likely the $C_3 \rightarrow C_4$ relationship would be underestimated, resulting in a larger residual for this covariance and overall poorer model fit. Plus, a number of other effects that used this path would be eliminated as well. This illustrates how the removal or addition of a path in the structural model can impact not only that direct relationship (e.g., $C_1 \rightarrow C_4$), but many other relationships as well.

OTHER ABSOLUTE FIT INDICES

Most SEM programs today provide the user with many different fit indices. In the text we focused more closely on those that are most widely used. In this section, we briefly touch on a few other absolute fit indices that are sometimes reported:

- The expected cross-validation index (ECVI) is an approximation of the goodness-of-fit the estimated model would achieve in another sample of the same size. Based on the sample covariance matrix, it takes into account the actual sample size and the difference that could be expected in another sample. The ECVI also takes into account the number of estimated parameters for a given model. It is most useful in comparing the performance of one model to another.
- The actual cross-validation index (CVI) can be formed by using the computed covariance matrix derived from a model in one sample to predict the observed covariance matrix taken from a validation sample. Given a sufficiently large sample (i.e., $N > 500$ for most applications), the researcher can create a validation sample by splitting the original observations randomly into two groups.

- Gamma Hat also attempts to correct for both the sample size and model complexity by including each in its computation. Typical Gamma Hat values range between .9 and 1.0. Its primary advantage is that it has a known distribution [10].

SPECIFICATION ISSUES IN SEM PROGRAMS

In this section we provide an overview of specification issues in SEM for two software packages. We will first discuss the notation used in LISREL, a popular SEM program. This notation has become a standard "language" of SEM in referring to both measurement and structural model relationships. We will then examine how the formulation of the path model is "translated" into program commands while conforming to the LISREL notation. The first example will be a simple path model to illustrate the basic issues involved. The discussion then shifts to the HBAT analysis from the textbook for both the CFA and structural models. We then review these issues briefly for AMOS as well. In the final section several more complex issues in model specification are discussed.

THE MULTIVARIATE RELATIONSHIPS IN SEM

As we discussed in the text, SEM models are defined by two sub-models – the measurement model and the structural model. Each sub-model can be expressed is a set of multivariate equations. It isn't called structural "equations" modeling for nothing! Even though it is possible to learn how to run a SEM model without a full and complete understanding of its equations, knowing the basic equations can be helpful in understanding the distinction between measured variables and constructs and between exogenous and endogenous constructs. Moreover, the equations introduce the notation used in LISREL, which we will discuss in more detail in the following section. Finally, the equations also help show how SEM is similar to other techniques.

The Main Structural Equation

In regression, our goal was to build a model that predicted a single dependent variable. Here, we are trying to predict and explain a set of endogenous constructs. Therefore, we need equations that explain endogenous constructs (η) in addition to those explaining the measured items (individual x and y variables used as indicators). Not surprisingly, we find that these equations are similar to the multiple regression equation that explains the dependent variable (y) with multiple independent variables (i.e., x_1 and x_2). This fundamental equation for the structural model is as follows (refer to the abbreviation guide in the Appendix of this document for any needed help with pronunciations or definitions):

$$\eta = \mathrm{B}\eta + \Gamma\xi + \zeta$$

The η represents the endogenous constructs in a model. We will have a separate equation for each endogenous construct. The η appears on both sides of the equation because endogenous constructs can be dependent on one another (i.e., one endogenous construct can

be a predictor of another endogenous construct). The **B** represents the parameter coefficients that link endogenous constructs with other endogenous constructs. The **B** is a matrix consisting of as many rows and columns as there are endogenous constructs. So is there are two endogenous constructs in the model, **B** would be a 2×2 matrix, with 2 rows and 2 columns. The individual elements of **B** are designated by a β. The Γ is the corresponding matrix of parameter coefficients linking the exogenous constructs (ξ) with the endogenous constructs (η). It also is a matrix that has as many rows as there are exogenous constructs and as many columns as there are endogenous constructs. If there are three exogenous constructs and two endogenous constructs, there would be a $3 \times 2 \, \Gamma$ matrix. Its individual elements are designated by γ as shown in the figure. Finally, ζ represents the error in the prediction of η. It can be thought of as the residual or converse of the R^2 concept from regression (i.e., $1 - R^2$).

Another way to think of the structural equation is as a multiple regression equation predicting η (a construct) instead of y, with the other η values and the ξ values as predictors. The **B** ($\eta_{1,1}$, ...) and Γ ($\gamma_{1,1}$, ...) provide structural parameter estimates. In the regression equation, the predictor values were represented by x and the standardized parameter estimates by the regression coefficients. In both cases, the parameter estimate depicts the linear relationship between a predictor and an outcome. Thus, clear similarities exist between SEM and regression analysis.

Using Constructs to Explain Measured Variables: The Measurement Model

Once values for η are known, we can also predict the y variables using an equation of the form:

$$y_i = \lambda_{i,j}\eta_j + \varepsilon_i$$

Here, each measured variable y is predicted by its loadings on the endogenous constructs. Typically a measured variable only has a loading on one construct, but that can vary in certain situations. Predicted values for each x also can be computed in the same manner using the following equation:

$$x_i = \lambda_{i,j}\xi_j + \delta_j$$

The predicted values for each observed variable, (whether a predicted x or y) can be used to compute covariance estimates that could be compared to the actual observed covariance terms in assessing model fit. In other words, we can use the parameter estimates to model the actual observed variables. The estimated covariance matrix obtained by computing covariation among predicted values for the measured items is Σ_k. Recall that the difference between the actual covariance matrix for observed items (**S**) and the estimated covariance matrix is an important part of analyzing the validity of any SEM model.

Rarely is it necessary in most applications to actually list predicted values based on the values of the other variables or constructs. Although it is useful to understand how predicted

values can be obtained because it helps demonstrate the way SEM works, the focus in social science typically is on explaining relationships.

Complete Structural and Measurement Model Equations

As noted earlier, LISREL notation has become in some sense the "language" of SEM. A researcher, therefore, must have a basic understanding of the notation no matter what software program is being used. The example below illustrates the complete equations for a model consisting of three exogenous constructs, two endogenous constructs and four indicators each for the sets of endogenous and exogenous constructs.

Structural Model Equations

Endogenous Construct		Exogenous Construct		Endogenous Construct		Error
η_1	$=$	$\gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3$	$+$	$\beta_{11}\eta_1 + \beta_{12}\eta_3$	$+$	ζ_1
η_2	$=$	$\gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3$	$+$	$\beta_{21}\eta_2 + \beta_{22}\eta_2$	$+$	ζ_2

Measurement Model Equations

Exogenous Indicator		Exogenous Construct		Error
X_1	$=$	$\lambda^x_{11}\xi_1 + \lambda^x_{12}\xi_2 + \lambda^x_{13}\xi_3$	$+$	δ_1
X_2	$=$	$\lambda^x_{21}\xi_1 + \lambda^x_{22}\xi_2 + \lambda^x_{23}\xi_3$	$+$	δ_2
X_3	$=$	$\lambda^x_{31}\xi_1 + \lambda^x_{32}\xi_2 + \lambda^x_{33}\xi_3$	$+$	δ_3
X_4	$=$	$\lambda^x_{41}\xi_1 + \lambda^x_{42}\xi_2 + \lambda^x_{43}\xi_3$	$+$	δ_4

Endogenous Indicator		Endogenous Constructs		Error
Y_1	$=$	$\lambda^y_{11}\eta_1 + \lambda^y_{12}\eta_2$	$+$	ε_1
Y_2	$=$	$\lambda^y_{21}\eta_1 + \lambda^y_{22}\eta_2$	$+$	ε_2
Y_3	$=$	$\lambda^y_{31}\eta_1 + \lambda^y_{32}\eta_2$	$+$	ε_3
Y_4	$=$	$\lambda^y_{41}\eta_1 + \lambda^y_{42}\eta_2$	$+$	ε_4

Structural Equation Correlations Among Constructs

	Among Exogenous Constructs (Phi ϕ)				*Among Endogenous Constructs (Psi ψ)*	
	ξ_1	ξ_2	ξ_3		η_1	η_2
ξ_1	—			η_1	—	
ξ_2	ϕ_{21}	—		η_2	ψ_{21}	—
ξ_3	ϕ_{31}	ϕ_{32}	—			

Correlations Among Indicators

	Among Exogenous Indicators *(Theta-delta θ_δ)*					**Among Endogenous Indicators** *(Theta-epsilon $\theta\varepsilon$)*			
	X_1	X_2	X_3	X_4		Y_1	Y_2	Y_3	Y_4
X_1	—				Y_1	—			
X_2	$\theta_{\delta 21}$	—			Y_2	$\theta\varepsilon_{21}$	—		
X_3	$\theta_{\delta 31}$	$\theta_{\delta 32}$	—		Y_3	$\theta\varepsilon_{31}$	$\theta\varepsilon_{32}$	—	
X_4	$\theta_{\delta 41}$	$\theta_{\delta 42}$	$\theta_{\delta 43}$	—	Y_4	$\theta\varepsilon_{41}$	$\theta\varepsilon_{42}$	$\theta\varepsilon_{43}$	—

The path model not only represents the structural relationships between constructs, but also provides a means of depicting the direct and indirect effects implied in the structural relationships. A "working knowledge" of the direct and indirect effects of any path model gives the researcher not only the basis for understanding the foundations of model estimation, but also insight into the "total" effects of one construct upon another. Moreover, the indirect effects can be further subdivided into casual and non-causal/spurious to provide greater specificity into the types of effects involved. Finally, an understanding of the indirect effects allows for greater understanding of the implications of model respecification, either through addition or deletion of a direct relationship.

The following table provides an overview of the notation used for matrices, constructs and indicators commonly used in SEM. SEM terminology often is abbreviated with a combination of Greek characters and roman characters to help distinguish different parts of a SEM model. It is followed by a guide to aid in the pronunciation and understanding of common SEM abbreviations.

Matrices, Construct/Indicators and Model Equation Notation of the LISREL Model

LISREL Model Element	Description	Notation Matrix	Element
Matrices			
Structural Model			
Beta	Relationships of endogenous to endogenous constructs	B	β_{nn}
Gamma	Relationships of exogenous to endogenous constructs	Γ	γ_{nm}
Phi	Correlation among exogenous constructs	Φ	ϕ_{mm}
Psi	Correlation of structural equations or endogenous constructs	Ψ	φ_n
Measurement Model			
Lambda-X	Correspondence (loadings) of exogenous indicators	Λ_x	λ^x_{pm}
Lambda-Y	Correspondence (loadings) of endogenous indicators	Λ_y	λ^y_{qn}
Theta-delta	Matrix of prediction error for exogenous construct indicators	Θ_δ	δ_{pp}
Theta-epsilon	Matrix of prediction error for endogenous construct indicators	Θ_ε	ε_{qq}
Construct/Indicators			
Construct			
Exogenous	Exogenous construct		ξ
Endogenous	Endogenous construct		η
Indicator			
Exogenous	Exogenous indicator		X
Endogenous	Endogenous indicator		Y
Structural and Measurement Model Equations			
Structural Model	Relationships between exogenous and endogenous constructs	$\eta = \Gamma\xi + B\eta + \zeta$	
Measurement Model			
Exogenous	Specification of indicators for exogenous constructs	$X = \Lambda_x\xi + \delta$	
Endogenous	Specification of indicators for endogenous constructs	$Y = \Lambda_y\eta + \varepsilon$	

Pronunciation Guide to SEM Notation

| Symbol	Pronunciation	Meaning
ξ	xi (KSI or KZI)	An exogenous construct associated with measured *X* variables
η	eta ("eight-ta")	An endogenous construct associated with measured *Y* variables
λ_X	lambda "x"	A path representing the factor loading between an exogenous latent construct and a measured *x* variable
λ_Y	lambda "y"	A path representing the factor loading between an endogenous latent construct and a measured *y* variable
Λ	capital lambda	The set of loading estimates represented in a matrix where rows represent measured variables and columns represent latent constructs
ϕ	phi (fi)	An arced two-headed arrow denoting the covariation of two exogenous (ξ) constructs
Φ	capital phi	A way of referring to the covariance or correlation matrix between a set of exogenous (ξ) constructs
γ	gamma	A path representing a causal relationship from an exogenous construct (ξ) to an endogenous construct (η)
Γ	capital gamma	A way of referring to the entire set of γ relationships for a given model
β	beta ("bay-ta")	A path representing a causal relationship from one endogenous (η) construct to another η construct
B	capital beta	A way of referring to the entire set of β relationships for a given model
δ	delta	The error term associated with an estimated, measured *x* variable
θ_δ	theta ("they-ta") delta	The residual variances and covariances associated with the *x* estimates; the error variance items are the diagonal
ε	epsilon	The error term associated with an estimated, measured *y* variable
θ_ε	theta-epsilon	The residual variances and covariances associated with the *y* estimates; the error variance items are the diagonal
ζ	zeta ("zay-ta")	The covariation between η construct errors
τ	tau (like "now")	The intercept terms for a measured variable
κ	kappa	The intercept terms for a latent construct
χ^2	chi (ki)-squared	The likelihood ratio

SPECIFYING A MODEL IN LISREL NOTATION

For LISREL and AMOS, the user can either use the drop-down menus to generate the syntax that matches the measurement model, draw the measurement model using a path diagram, or write the appropriate program commands into a syntax window. If either of the first two alternatives is done correctly, the programs generate the program syntax automatically. We will discuss this third approach for LISREL since this best illustrates how to use LISREL notation in specifying the model.

Specification of a CFA Model with LISREL

Specification is quite different using CFA compared to EFA. The commands below illustrate how the simple CFA model shown below is communicated using LISREL program statements. Note that here we only provide the commands needed to define the model. The complete set of program commands are given in our HBAT example in a later section. Also, line numbers have been added to the commands for reference, but they are not needed as input to LISREL.

In our example we have four constructs, each with four indicators. See the CFA path model below:

(a) CFA Model
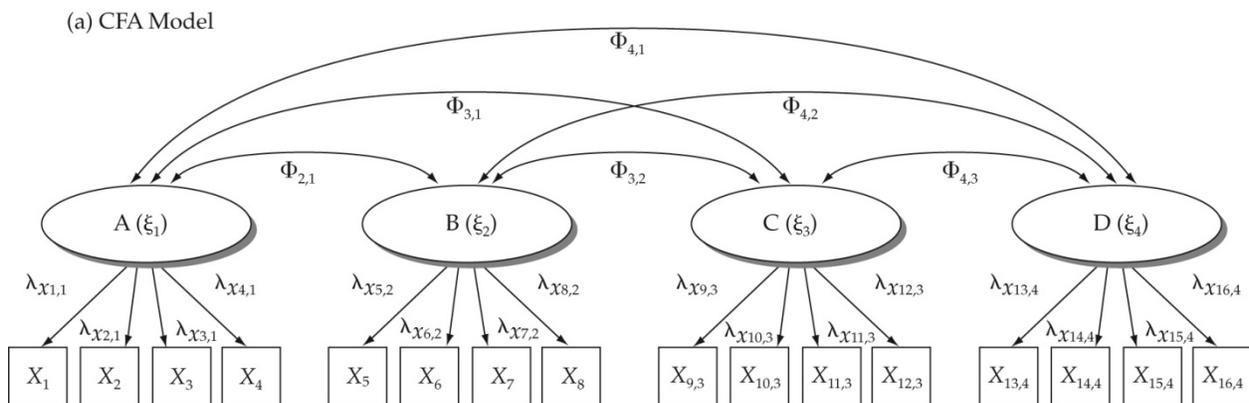
The LISREL commands for this CFA are as follows:

```
01    MO NX=16 NK=4 PH=SY,FR
02    VA 1.0 LX 1 1 LX 5 2 LX 9 3 LX 13 4
03    FR LX 2 1 LX 3 1 LX 4 1 LX 6 2 LX 7 2 LX 8 2
04    FR LX 10 3 LX 11 3 LX 12 3 LX 14 4 LX 15 4 LX 16 4
```

We begin with the Model command (MO) indicating the numbers of measured and latent

variables and descriptions of the key matrices of parameters. In a CFA model we only have exogenous constructs and thus of *x* variables. NX stands for number of *x* variables, in this case 16. NK stands for the number of exogenous (ξ) constructs, in this case 4. PH indicates that the matrix of covariances between the 4 constructs (Φ) will be symmetric (SY) and free (FR). In other words, the construct variances (the diagonal of Φ) and the covariance between each pair of constructs will be estimated.

Line 2 is a value statement (VA) where we assign a value to a fixed parameter. In this case, each of the parameters listed on this line is fixed to 1.0 to "set the scale" for the constructs. One item is fixed to 1.0 on each construct. LX1,1 represents the parameter for the first loading on the first construct ($\lambda_{x1,1}$). The L stands for lambda, the X is an *x* variable and 1 1 stand for the measured variable number and construct number, respectively. Thus, LX2,1 stands for the parameter representing the factor loading of the second measured variable (x_2) on the first latent construct (ξ_1), or $\lambda_{x2,1}$.

Lines 3 and 4 designate the free loading estimates (FR). The 12 loadings referred to on these lines will be estimated and shown as factor results in the output (in Λ_x). Thus, this model estimates a total of 16 loadings, one for each indicator (actually 12 are estimated and four fixed to a value of 1.0) as shown in the path diagram. This compares to EFA, where there would be a total of 64 loadings (one for each indicator on each construct).

Changing The CFA Setup in LISREL to a Structural Model Test

As discussed in the text, the CFA model forms the "foundation" from which the structural model is formulated. In making the "conversion" from a CFA to a structural model, the research must make two fundamental decisions: distinguish between exogenous and endogenous constructs and specify the structural relationships between constructs. Note that in most instances, the measurement model will be specified and analyzed in the CFA stage.

To illustrate the process, we utilize the CFA example discussed in the section above. As can be seen from the path model below, two of the constructs are defined as endogenous with relationships to the two remaining exogenous constructs.

(b) Structural Model
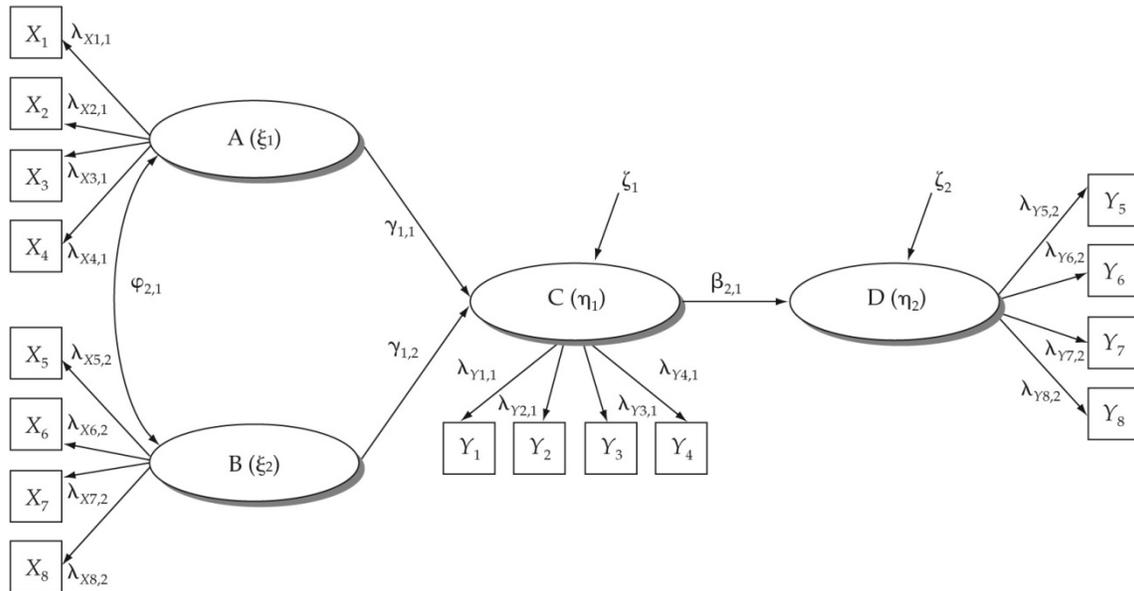
Shown below are the modified LISREL syntax changes corresponding to the structural model. As before, line numbers (not required in the actual LISREL syntax) are included for reference and only the commands relating to the model specification are shown.

```
MO NY=8 NE=2 NX=8 NK=2 PH=SY,FR PS=DI,FR GA=FU,FI BE=FU,FI
VA 1.0 LX 1 1 LX 5 2 LY 1 1 LY 5 2
FR LX 2 1 LX 3 1 LX 4 1 LX 6 2 LX 7 2 LX 8 2
FR LY 2 1 LY 3 1 LY 4 1 LY 6 2 LY 7 2 LY 8 2
FR GA 1 1 GA 1 2
FR BE 2 1
```

The structural model commands have several changes:

1. The MO statement now provides values for:

    a. The number of indicators of endogenous constructs (NY = 8)
    b. The number of endogenous constructs (NE = 2)
    c. The new number of indicators of exogenous constructs (NX = 8)
    d. The new number of exogenous constructs (NK = 2)
2. The MO statement now provides the parameter matrices for the structural parameter estimates:
    a. GA stands for the relationships between exogenous and endogenous constructs (Γ, or gamma). It is specified as full (FU) and fixed (FI). The convention is to specify individual free elements below.

b. BE stands for the relationships between endogenous constructs (η). It is also specified as FU and FI. The free elements will be specified with a FR statement below the MO line.

3. The scales are set for the exogenous and endogenous constructs by setting one loading to be equal to 1.0 as follows:

```
VA 1.0 LX 1 1 LX 5 2 LY 1 1 LY 5 2
```

4. The factor pattern for both λ_x and λ_y (the respective loading matrices) is specified in much the same way as in the CFA model. The only difference is that the item indicators for constructs C and D are now referred to by LY instead of LX.
5. The pattern of hypotheses between latent constructs is represented by freeing the appropriate elements of Γ and **B**, respectively.

```
FR GA 1 1 GA 1 2
FR BE 2 1
```

The remainder of the model syntax can be specified as in the CFA model setup with the exception of the optional label statements (LA). Separate label statements are needed for exogenous and endogenous constructs. These will be illustrated in the HBAT example in the following section.

HBAT: THE CFA MODEL

The examples above are for a simplified path model, plus they do not contain the complete LISREL command set. The following example uses the HBAT example from the textbook and demonstrates the process of specifying the CFA model in LISREL commands. As before, line numbers have been added for reference purposes, but are not required in LISREL. The commands for the CFA analysis are as follows:

```
01 TI HBAT CFA MODEL
02 DA NI=28 NO=399 MA=CM
03 CM FI=HBAT.COV
04 LA
05 ID JS1 OC1 OC2 EP1 OC3 OC4 EP2 EP3 AC1 EP4 JS2 JS3 AC2 SI1 JS4
   SI2 JS5 AC3 SI3 AC4 SI4
06 C1 C2 C3 AGE EXP JP
07 SE
08 JS1 JS2 JS3 JS4 JS5 OC1 OC2 OC3 OC4 SI1 SI2 SI3 SI4 EP1 EP2
   EP3 EP4 AC1 AC2 AC3 AC4/
09 MO NX=21 NK=5 PH=SY,FR TD=DI,FR
10 VA 1.00 LX 1 1 LX 6 2 LX 10 3 LX 14 4 LX 18 5
11 FR LX 2 1 LX 3 1 LX 4 1 LX 7 2 LX 8 2 LX 9 2
12 FR LX 11 3 LX 12 3 LX 13 3 LX 15 4 LX 16 4 LX 17 4
     LX 19 5 LX 20 5 LX 21 5
13 LK
14 'JS' 'OC' 'SI' 'EP' 'CA'
15 PD
16 OU RS SC ND=2
```

Line 01 is simply a title statement. The user can enter anything on this line that helps identify the analysis. Line 02 is a data statement. It must begin with DA and tells the SEM program that 28 variables are included in the data set of 399 observations. Although the data set originally contained 400 observations, one response point was deleted for being out of range and another was simply missing. Using pairwise deletion and the previous rule of thumb, the number of observations was set at the minimum number of observations for any covariance computation. In this case, at least 399 observations are involved in any single covariance computation. This number can be verified by examining the statistical output for the covariance computations. If listwise deletion had been used, then NO would be set at 398 since both cases with a missing response would be deleted from any computations. MA = CM denotes that the input matrix is a covariance matrix. Line 03 indicates that a covariance matrix (CM) is stored in a file (FI) named HBAT.COV. Line 04 is a labels statement and must begin with LA. The labels are listed beginning on the line below. Lines 05 and 06 show the labels for the 28 variables. Users can choose any labels the respective program will allow. In this case, HBAT labeled the variables with initials from the construct names like JS1, JS2,..., SI4. They could have used X1–X28 or V1–V28 or any other similar abbreviation. One label is necessary for each variable in the data set.

Line 07 is a select statement and must be denoted with SE. It indicates that the variables listed on the next line(s) are the ones to be used in the analysis. A / indicates the end of the selected variables list. The order is particularly important. Whatever is listed first will become the first observed variable. For example, the first measured variable in the CFA program, designated as x_1 (the small x with subscript here representing the first observed variable selected and corresponds to the loading estimate $\lambda_{x1,1}$), will be represented by the inputted variable labeled "JS1." "SI4," the 21st variable on the SE line, will become the 21st measured variable or x_{21}, and the loading estimates associated with this variable will be found in the 21st row of the factor loading matrix ($\lambda_{x21,5}$ of Λ_x in this case).

Only in rare circumstances will the variables be stored in the original data file in the exact order that would match the configuration corresponding to the theory being tested. Also, the user seldom includes all variables in the CFA because most data will also contain some classification variables or identifying variables as well as potential variables that were measured but not included in the CFA. The select process, whether through a statement or a drop-down menu, is the way the variables involved in the CFA are selected.

Line 09 is a model statement and must begin with MO. Model statements indicate the respective numbers of measured and latent variables and can include descriptions of the key matrices of parameters. The abbreviations shown here are relatively easy to follow. NX stands for number of x variables, in this case 21. NK stands for the number of ξ constructs, in this case 5. PH indicates that the matrix of covariances between the 5 constructs (Φ) will be symmetric (SY) and free (FR). In other words, the construct variances (the diagonal of Φ) and the covariance between each pair of constructs will be estimated. TD is the matrix of error variances and covariances. It is set as diagonal (DI) and free (FR), so only the error variances are estimated. Any parameter matrix not listed in the MO line is set at the program default value. The reader can consult the program documentation for other possible abbreviations and defaults.

Line 10 is a value statement (VA). Value statements assign a value to a fixed parameter. In this case, each of the parameters listed on this line is fixed to 1.0. This statement sets the scale for the constructs so that one item is fixed to 1.0 on each construct. LX1,1 represents the parameter for the first loading on the first construct ($\lambda_{x1,1}$). The L stands for lambda, the X is an x variable and 1 1 stand for the measured variable number and construct number, respectively. Thus, LX2,1 stands for the parameter representing the factor loading of the second measured variable (x_2) on the first latent construct (ξ_1), or $\lambda_{x2,1}$. Factor loadings in a reflective factor model can be expressed equally as causal paths. Using this terminology, LX21,5 stands for the path from construct ξ_5 to x_{21} ($\lambda_{x21,5}$).

Lines 11 and 12 start with FR and designate the free loading estimates. The 16 loadings referred to on these lines will be estimated and shown as factor results in the output (in Λ_x). With the five estimates fixed at 1 on line 10 and 16 loadings estimated, 84 elements remain in the factor pattern fixed at zero (21 variables $\times$ 5 constructs = 105 potential loadings; 105 – 16 – 5 = 84). Recall that EFA would produce an estimate for all 105 loadings. The pattern of free and fixed loadings matches the theoretical structure proposed in the measurement model. Consistent with the congeneric model proposed, only one loading estimate is free for each measured indicator variable. In other words, each measured indicator variable loads on only one construct.

Line 13 is another label line. It is where the labels for the latent constructs can be listed. LK stands for labels for ksi (ξ). The actual labels appear on the next line or lines if necessary. In this case, the labels match the construct abbreviations provided (JS, OC, SI, EP, and AC). Line 15, with the abbreviation PD, requests that a path diagram be drawn by the program depicting the specified model and path estimates. The OU line (16) is required and is where any one of numerous options can be requested. For example, the SC is requesting that completely standardized estimates be included in the output. RS requests that all model residuals resulting from estimating the model be shown, including both the standardized and nonstandardized residuals. ND = 2 means that results will be shown to two significant digits.

At times a researcher may wish to place additional constraints on a CFA model. For instance, it is sometimes useful to set two or more parameters as equal. It would produce a solution that requires the values for these parameters be the same. If tau-equivalence is assumed for instance, this constraint is needed. With LISREL, this task can be done using the EQ command line. Similarly, researchers sometimes wish to set a specific parameter to a specific value by using the VA command line. Additional information about constraints can be found in the documentation for the SEM program of choice.

HBAT: THE STRUCTURAL MODEL

The HBAT CFA can be "transformed" into the HBAT structural model as was done earlier in the example. The LISREL commands for the structural model are shown below, followed by a discussion of making the changes from the CFA to the structural model. Again, line numbers have been added to the far left to aid in describing the syntax.

```
01 TI HBAT EMPLOYEE RETENTION MODEL
02 DA NI=28 NO=399 NG=1 MA=CM
03 CM FI=HBAT.COV
04 LA
05 ID JS1 OC1 OC2 EP1 OC3 OC4 EP2 EP3 AC1 EP4 JS2 JS3 AC2 SI1 JS4
   SI2 JS5 AC3 SI3 AC4 SI4
06 C1 C2 C3 AGE EXP JP
07 SE
08 JS1 JS2 JS3 JS4 JS5 OC1 OC2 OC3 OC4 SI1 SI2 SI3 SI4 EP1 EP2
   EP3 EP4 AC1 AC2 AC3 AC4/
09 MO NY=13 NE=3 NX=8 NK=2 PH=SY,FR PS=DI,FR BE=FU,FI GA=FU,FI
   TD=DI,FR TE=DI,FR
10 VA 1.00 LX 1 1 LX 5 2 LY 1 1 LY 6 2 LY 10 3
11 FR LX 2 1 LX 3 1 LX 4 1 LX 6 2 LX 7 2 LX 8 2
12 FR LY 2 1 LY 3 1 LY 4 1 LY 5 1 LY 7 2 LY 8 2 LY 9 2 LY 11 3 LY
   12 3 LY 13 3
13 FR GA 1 1 GA 2 1 GA 1 2 GA 2 2
14 FR BE 2 1 BE 3 1 BE 3 2
15 LK
16 EP AC
17 LE
18 JS OC SI
19 PD
20 OU RS SC MI EF ND=2
```

The first change from the CFA setup is noted in line 09. The model statement must now specify a number of variables and constructs for both exogenous and endogenous constructs. Thus, the MO line specifies NY = 13 (5 items for JS, 4 items for OC, 4 items for SI). Even though these are the same items as represented by these constructs in the CFA model, they now become *y* variables because they are associated with an endogenous construct. Their loading parameters are now changed to be consistent with this to λ_y (LY). Next, the MO line specifies NE = 3, indicating three endogenous constructs. This process is repeated for the exogenous constructs (NX = 8 and NK = 2). PH and TD remain the same.

Several new matrices are specified. BE = FU,FI means that **B,** which will list all parameters linking endogenous constructs with one another (β), is set to full and fixed. It means we will free the elements corresponding to the following hypotheses. GA representing Γ, which will list all parameters linking exogenous constructs with endogenous constructs (γ), is treated in the same way. Because we now have endogenous constructs, the error variance terms associated with the 13 *y* variables are now shown in θ_ε, which is abbreviated with TE = DI,FR, meaning it is a diagonal matrix and the diagonal elements will be estimated.

Line 10 sets the scale for factors just as in the CFA model with the exception that three of the set values are for *y* variables (λ_y values: LY1,1; LY6,2; LY10,3). Lines 11 and 12 specify the free values for the measured items just as in the CFA. We are following the rule of thumb that the free factor loading parameters should be estimated rather than fixed even though we have some idea of their value based on the CFA results. Lines 13 and 14 specify the pattern of free structural parameters. Line 13 specifies the free elements of Γ. These correspond with $H_1 - H_4$

in Figure 12-6 ($\gamma_{1,1}$ is listed as GA1,1). Similarly, line 14 specifies the free elements of **B.** Lines 15 and 16 lists the labels for the ξ constructs (LK). Lines 17 and 18 do the same for the η constructs (LE). Line 19 contains a PD that tells the program to generate a path diagram from the input. Line 20 is the output line and is the same as in the CFA example except for the addition of EF, which will provide a separate listing of all direct and indirect effects.

If the user is using a graphical interface (e.g., AMOS or LISREL), the user will need to make the corresponding changes to the path diagram. These changes would include making sure the constructs are properly designated as exogenous or endogenous and that observed variables each have a corresponding error variance term. Then each of the curved two-headed arrows that designated covariance between constructs in CFA will have to be replaced with a single-headed arrow to represent hypothesized relationships. Arrows between constructs for which no relationship is hypothesized are unnecessary. Therefore, the two-headed paths between these constructs in the CFA can be deleted. Once these changes are made, the user can reestimate the model and the results should now reflect the structural model results. If the program syntax has been changed as indicated, the program will produce the appropriate path diagram automatically.

A visual diagram corresponding to the SEM can be obtained by selecting Structural Model from the view options and requesting that the completely standardized estimates be displayed by the SEM program. In LISREL, for example, the values on the path diagram can be requested so that either the estimates are shown on the diagram, the *t*-values for each estimate, or other key estimates are shown including the modification indices.

How to Fix Factor Loadings to a Specific Value in LISREL

If a researcher wished to fix the factor loadings of a SEM model to the values identified in the CFA, procedures such as those described here can be used. To specify the values shown in the path model below, the researcher would take the following steps if using the LISREL software.
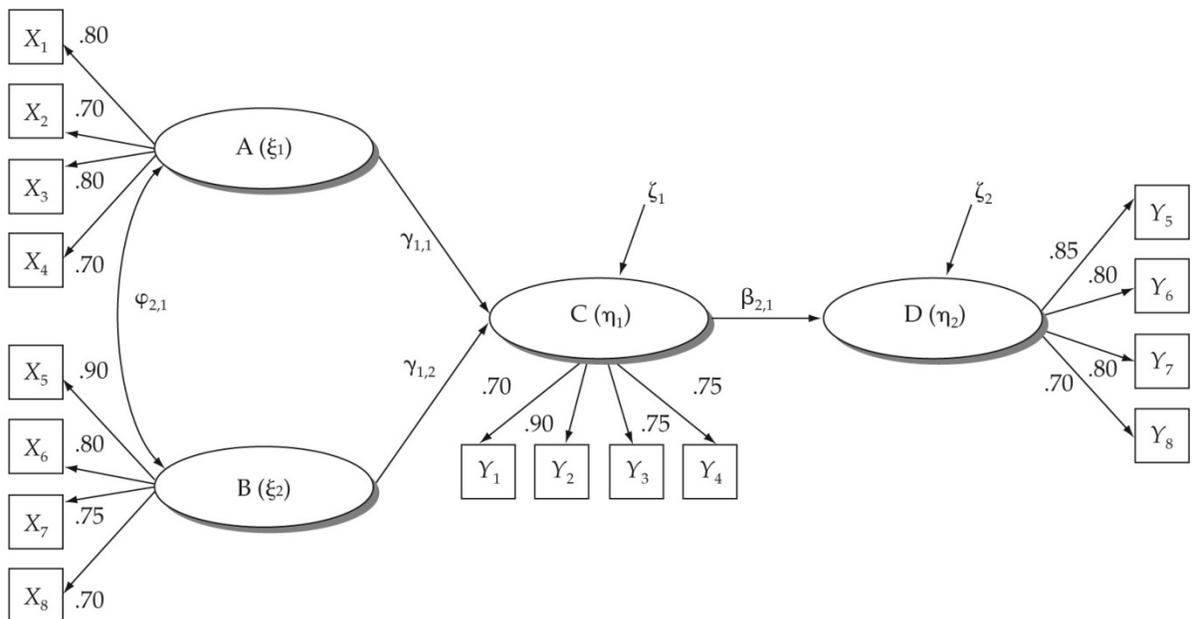
The following loading estimates would be fixed and their values set as follows:

```
FI LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 2 LX 6 2 LX 7 2 LX 8 2
FI LY 1 1 LY 2 1 LY 3 1 LY 4 1 LY 5 2 LY 6 2 LY 7 2 LY 8 2
VA .80 LX 1 1
VA .70 LX 2 1
VA .80 LX 3 1
VA .75 LX 4 1
VA .90 LX 5 2
VA .80 LX 6 2
VA .75 LX 7 2
VA .70 LX 8 2
VA .70 LY 1 1
VA .90 LY 2 1
VA .75 LY 3 1
VA .75 LY 4 1
VA .85 LY 5 2
VA .80 LY 6 2
VA .80 LY 7 2
VA .70 LY 8 2
```

The error variance terms also can be fixed to their CFA estimates as shown here:

```
FI  TD 1 1  TD 2 2  TD 3 3  TD 4 4  TD 5 5  TD 6 6  TD 7 7  TD 8 8
FI  TE 1 1  TE 2 2  TE 3 3  TE 4 4  TE 5 5  TE 6 6  TE 7 7  TE 8 8
VA  .36  TD 1 1
VA  .51  TD 2 2
VA  .36  TD 3 3
VA  .44  TD 4 4
VA  .19  TD 5 5
VA  .36  TD 6 6
VA  .44  TD 7 7
VA  .36  TD 8 8
VA  .51  TE 1 1
VA  .81  TE 2 2
VA  .44  TE 3 3
VA  .44  TE 4 4
VA  .28  TE 5 5
VA  .36  TE 6 6
VA  .36  TE 7 7
VA  .51  TE 8 8
```

The researcher could then proceed to specify the free elements of the structural theory.

MEASURED VARIABLE AND CONSTRUCT INTERCEPT TERMS

It often becomes necessary to use the measured variable and latent variable means in drawing conclusions about similarities and differences between groups. Until now, no SEM equation has shown a mean value. Now, however, the means may be considered.

One way that we could think about the mean value of any measured variable is to think of it as the sum of its zero-intercept term, plus the factor loading, times the average value of the latent construct. In equation form, it would look like the following expressed in terms of x_1:

$$\mu_{X1} = \tau_{X1} + \lambda_{X1}\kappa_{\xi 1}$$

The $\kappa_{\xi 1}$ represents the mean value for the first latent construct ξ_1, the μ_{x1} represents the mean of the measured variable x_1, and the τ_{X1} is the zero-intercept for x_1. More generally, κ represents the mean for any latent construct. Mathematically, it is also the zero-intercept term when solving for ξ. Even though the mathematics in this calculation may be difficult to follow, it is important to know that unless specific instructions are provided to the SEM program, it will *not* consider nor estimate construct means of any type.

This equation can be rearranged to solve for either τ_{X1} or κ. If any hypotheses concern differences between construct means, those differences can be found in the values for κ.

PATH MODEL SPECIFICATION WITH AMOS

Program statements can also be written for AMOS that would form the model in the same way as the LISREL statement. However, the assumption with AMOS is that the user will work with a path diagram. In essence, the path diagram provides the framework from which to build the

model. However, the user must assign variables to each rectangle, which represents a measured variable, and assign construct names to each oval. Likewise, the user must specific names for each measured variable error term. Then, the appropriate arrows must be drawn to form the model. The user must be careful that variables are assigned correctly. Drop-down windows can be used to add constraints to the model and to perform advanced applications like multiple group analysis. While it is possible to specify the SEM model through commands, AMOS is designed to be used through the graphical interface and this is the recommended route for most users.

RESULTS USING DIFFERENT SEM PROGRAMS

Although the input for different SEM programs varies, the results should be essentially the same. The algorithms may vary slightly, but a model that displays good fit using one SEM program also should display good fit in another. Each has its own idiosyncrasies that may prevent the same model specification from being estimated. For instance, some make it more or less difficult to use each of the missing variable options just mentioned. Each approach can be easily specified with LISREL, but AMOS uses EM alone. Listwise deletion, for example, can be performed with AMOS by screening observations with missing data prior to beginning the AMOS routine (e.g., with SPSS).

The overall model fit statistics, including the χ^2 and all fit indices, should not vary in any consequential way between the programs. Similarly, the parameter estimates should also not vary in any consequential way. Differences can be expected in two areas.

One area where differences in the numerical estimates may vary is in the residuals. In particular, some differences may be found between AMOS and the other programs. Without getting into the details, AMOS uses a different method for scaling the error terms of measured variables than do the other programs. This format has to do with setting the scale for the error terms, much as we set the scale for the latent constructs in a SEM model. This method may cause relatively small differences in the values for residuals and standardized residuals computed with AMOS. However, the differences do not affect the rules of thumb given in the text.

Another area where numerical estimates may vary is in the modification indices. Again, AMOS takes a different computational approach than do some of the other SEM programs. The difference lies in whether the change in fit is isolated in one or several parameters. Once again, although the user comparing results between AMOS and other programs may find some differences in MI, the differences should not be so large as to affect the conclusions in most situations. So, once again, the rules of thumb for the MI hold using any SEM program.

ADDITIONAL SEM ANALYSES

TESTING FOR DIFFERENCES IN CONSTRUCT MEANS

A final type of multigroup comparison is the test for differences in construct means. If at least partial scalar invariance is present, we can operationalize a value for the means of the latent constructs. In this way, ww tell the SEM program that we are interested in analyzing means. An earlier discussion showed the equation to represent latent construct means. In one way or another though, the SEM program must be told we are interested in the means of the latent constructs.

SEM programs compare means only in a relative sense. In other words, they can tell you whether the mean is higher or lower relative to another group. One reason for this limitation has to do with identification given that the intercept terms are now being estimated. A result is that the vector of latent construct means (contained in the kappa matrix) has to be fixed to zero in one group to identify the model. We refer to this group as group 1. It can be freely estimated in the other group(s) and the resulting values can be interpreted as how much higher or lower the latent construct means are in this group relative to group 1.

Assume we have a two-group model with three constructs in each group. The SEM output will now include estimates for the vector in group 2 (i.e., the comparison of group 2 relative to group 1). Typically, this output would include an estimated value, a standard error, and a t-value associated with each value. For instance, it may look like this:

| | KAPPA (κ) | |
Construct 1	Construct 2	Construct 3
2.6	0.09	−3.50
(0.45)	(0.60)	(1.55)
5.78	0.10	−2.25

These values suggest that the mean for the construct 1 is 2.6 greater in group 2 than in group 1. This difference is significant as evidenced by the t-value of 5.78 ($p < .001$). The mean for construct 2, differing by 0.09, is not significantly different ($t = 0.10$). The mean for construct 3 on the other hand, is significantly lower in group 2 relative to group 1 ($t = -2.25$; $p < .05$).

ITEM PARCELING IN CFA AND SEM

Item parceling refers to combining measured variables into sets of several variables by either summing or averaging the items. Parceling provides a way of dealing with an unmanageable number of measured reflector variables per construct. For example, some psychological

scales may contain more than 100 items to capture only two or three basic personality dimensions. Thus, even with a few constructs one could end up with far more than 100 measured items. SEM applications are difficult to manage with so many measured variables.

Using item parceling a single latent construct with 40 measured items (x_1–x_{40}) could be represented by eight parcels, each consisting of 5 of the 40 measured items. A *parcel* is a mathematical combination summarizing multiple variables into one. In the extreme, all measured items for a construct can be combined into one average or sum of those variables. In Chapter 3, we discussed how to create a summated construct in this fashion. The term *composite indicator* is generally used to refer to parceling resulting in only one parcel from all the measured items for a construct.

Numerous issues are associated with item parceling. These issues include the appropriateness of parceling, which items should be combined into a parcel, and what the effects of parceling are on evaluating models. Parceling has the potential to improve model fit simply because it reduces the complexity of the model, and models with fewer variables have the potential for better fit. Better fit alone, however, is not a sufficient rationale for combining multiple items into one because the primary goal is creating a model that best represents the actual data. Further, item parcels can often mask problems with item measures and suggest a better fit than actually exists in reality. Parceling also can hide other latent constructs that exist in the data. So, a covariance matrix that actually contains five latent constructs may be adequately but falsely represented by three latent constructs using parceling.

When Is Parceling Appropriate?

Item parceling should only be considered when a construct has a large number of measured variable indicators. For instance, applications involving fewer than 15 items do not call for parceling. Similarly, parceling is not used with formative models because it is important that all causes of a formative factor be included. Parceling is appropriate when all the items for a construct are unidimensional. That is, even with a large number of measured items, they should all load highly on only one construct and it should display high reliability (.9 or better). Most importantly, parceling is appropriate when information is not lost by using parcels instead of individual items. Thus, some simple checks prior to parceling would involve running a CFA on the individual factor to check for unidimensionality and to see whether the construct reflected by all individual items relates to other constructs in the same way as a

construct reflected by a smaller number of parcels.

How Should Items Be Combined into Parcels?

Traditionally, little thought was given to how items should be combined. However, the combination strategy can affect the likelihood that a CFA is actually supporting a false measurement theory. Although many intricacies are associated with the combination strategies, two simple considerations lead to the best performance when a researcher must use item parcels. One consideration is empirical and the other is theoretical. Given that the individual items suggest unidimensionality, the best parcels are formed by items that display approximately the same covariance, which should lead them to have approximately the same factor loading estimates. Further, the parcels should contain groups of items with the most conceptual similarity. That is, items with the closest content validity. Thus, parcels with items showing approximately the same amount of covariance and that share a conceptual basis will tend to perform well and represent the data most accurately.

MEASUREMENT BIAS

Researchers sometimes become concerned that survey responses are biased based on the way the questions are asked. For instance, it could be argued that the order in which questions are asked could be responsible for the covariance among items that are grouped close together. If so, a nuisance factor based on the physical proximity of scale items may be explaining some of the inter-item covariance.

Similarly, researchers often are faced with resolving the question of constant methods bias. Constant methods bias would imply that the covariance among measured items is driven by the fact that some or all of the responses are collected with the same type of scale. A questionnaire using only semantic differential scales, for instance, may be biased because the opposing terms response form becomes responsible for covariance among the items. Thus, the covariance could be explained by the way respondents use a certain scale type in addition to or instead of the content of the scale items. Here, a simple illustration is provided using the HBAT example. It shows how a CFA model can be used to examine the possibility of measurement bias in the form of a nuisance factor.

The HBAT employee questionnaire consists of several different types of rating scales. Although it could be argued that respondents prefer a single format on any questionnaire, several advantages come with using a small number of different formats. One advantage is that the extent to which any particular scale type is biasing the results can be assessed using CFA.

In this case, HBAT is concerned that the semantic differential items are causing measurement bias. The analyst argues that respondents have consistent patterns of responses to semantic differential scales no matter what the subject of the item is. Therefore, a semantic differential factor may help explain results. A CFA model can be used to test this proposition. One way to do so is to create an additional construct that is also hypothesized as causing the semantic differential items. In this case, items EP4, JS2, JS3, AC2, and SI4 are measured with semantic differential scales. Thus, the model needs to estimate paths between this new construct and these measured items. The addition of a nuisance factor of this type violates the principles of good measurement and so the new model will not have congeneric measurement properties.

We will modify the original HBAT CFA model shown in the text. A sixth construct is introduced (ξ_6). Next, dependence paths (causal in this case) would be estimated (drawn if using a path diagram) from ξ_6 to EP4, JS2, JS3, AC2, and SI4. Thus, the factor pattern no longer exhibits simple structure because each of these measured variables is now determined both by its conceptual factor and by the new construct ξ_6.

The analyst tests this model and observes the following fit statistics. The χ^2 = 232.6 with 174 degrees of freedom and the RMSEA, PNFI, and CFI are .028, .80, and .99, respectively. The added paths have not provided a poor overall fit although the RMSEA has increased slightly and the PNFI has decreased. However, the $\Delta\chi^2$ = 4.0 (236.6 − 232.6) with 5 (179 − 174) degrees of freedom, is insignificant. In addition, none of the estimates associated with the bias factor (ξ_6) are significant. The completely standardized estimates of lambda (factor loadings) and associated t-values are shown here:

Parameter	Estimate	t-Value
$\lambda_{x2,6}$	0.14	1.19
$\lambda_{x3,6}$	−0.01	−0.08
$\lambda_{x17,6}$	0.16	1.32
$\lambda_{x19,6}$	0.07	0.84
$\lambda_{x21,6}$	0.20	1.48

Also, the values for the original parameter estimates remain virtually unchanged as well. Thus, based on the model fit comparisons, the insignificant parameter estimates, and the parameter stability, no evidence supports the proposition that responses to semantic differential items are biasing results. The HBAT analyst concludes, therefore, that this case is not subject to measurement bias. Another factor could be added to act as a potential nuisance cause for the items representing another scale type, such as all Likert items. The test would proceed in much the same way. The end result of all of these tests is that the researcher can proceed to test more specific hypotheses about employee retention and

related constructs.

CONTINUOUS VARIABLE INTERACTIONS

An approach for handling a continuous moderator which does not involve creating groups from the continuous moderator is to create an interaction between the moderator and the predictor. Single-variable interactions were treated in the text, so we focus here on a moderating construct that would be measured by multiple indicators. Consider a SEM model with two exogenous constructs predicting a single endogenous construct. Each construct is indicated by four measured items. If the first construct (ξ_1) is hypothesized as the predictor construct and the second construct (ξ_2) is hypothesized as a moderator, then an interaction construct can be created to represent the moderating effect by multiplying the indicators of the predictor and moderator constructs together. Using this rationale, the indicators for the third interaction construct (ξ_3) can be computed as follows:

$$x_9 = x_1 \times x_5$$
$$x_{10} = x_2 \times x_6$$
$$x_{11} = x_3 \times x_7$$
$$x_{12} = x_4 \times x_8$$

These computed variables can then be added to the actual data containing 12 measured variables and the covariance terms between these computed variables and the others can be calculated. Now, the covariance matrix for this model would change from 12×12 to 16×16. This can be shown in a path model form by the following specification.

The following values would be set as follows:

$$\lambda_{x9,3} = \lambda_{x5,2}\lambda_{x1,1} \qquad \theta_{\delta9,1} = \tau_1\,\theta_{\delta1,1}$$
$$\lambda_{x10,3} = \lambda_{x6,2}\lambda_{x2,1} \qquad \theta_{\delta9,5} = \tau_5\,\theta_{\delta5,5}$$
$$\lambda_{x11,3} = \lambda_{x7,2}\lambda_{x3,1} \qquad \theta_{\delta10,2} = \tau_2\,\theta_{\delta2,2}$$
$$\lambda_{x12,3} = \lambda_{x8,2}\lambda_{x4,1} \qquad \theta_{\delta10,6} = \tau_6\,\theta_{\delta6,6}$$
$$\lambda_{x9,1} = \tau_1\lambda_{x1,1} \qquad \theta_{\delta11,3} = \tau_3\,\theta_{\delta3,3}$$
$$\lambda_{x10,1} = \tau_2\lambda_{x2,1} \qquad \theta_{\delta11,7} = \tau_7\,\theta_{\delta7,7}$$
$$\lambda_{x11,1} = \tau_3\lambda_{x3,1} \qquad \theta_{\delta12,4} = \tau_4\,\theta_{\delta4,4}$$
$$\lambda_{x12,1} = \tau_4\lambda_{x4,1} \qquad \theta_{\delta12,8} = \tau_8\,\theta_{\delta8,8}$$
$$\lambda_{x9,2} = \tau_5\lambda_{x5,2}$$
$$\lambda_{x10,2} = \tau_6\lambda_{x6,2}$$
$$\lambda_{x11,2} = \tau_7\lambda_{x7,2}$$
$$\lambda_{x12,2} = \tau_8\lambda_{x8,2}$$
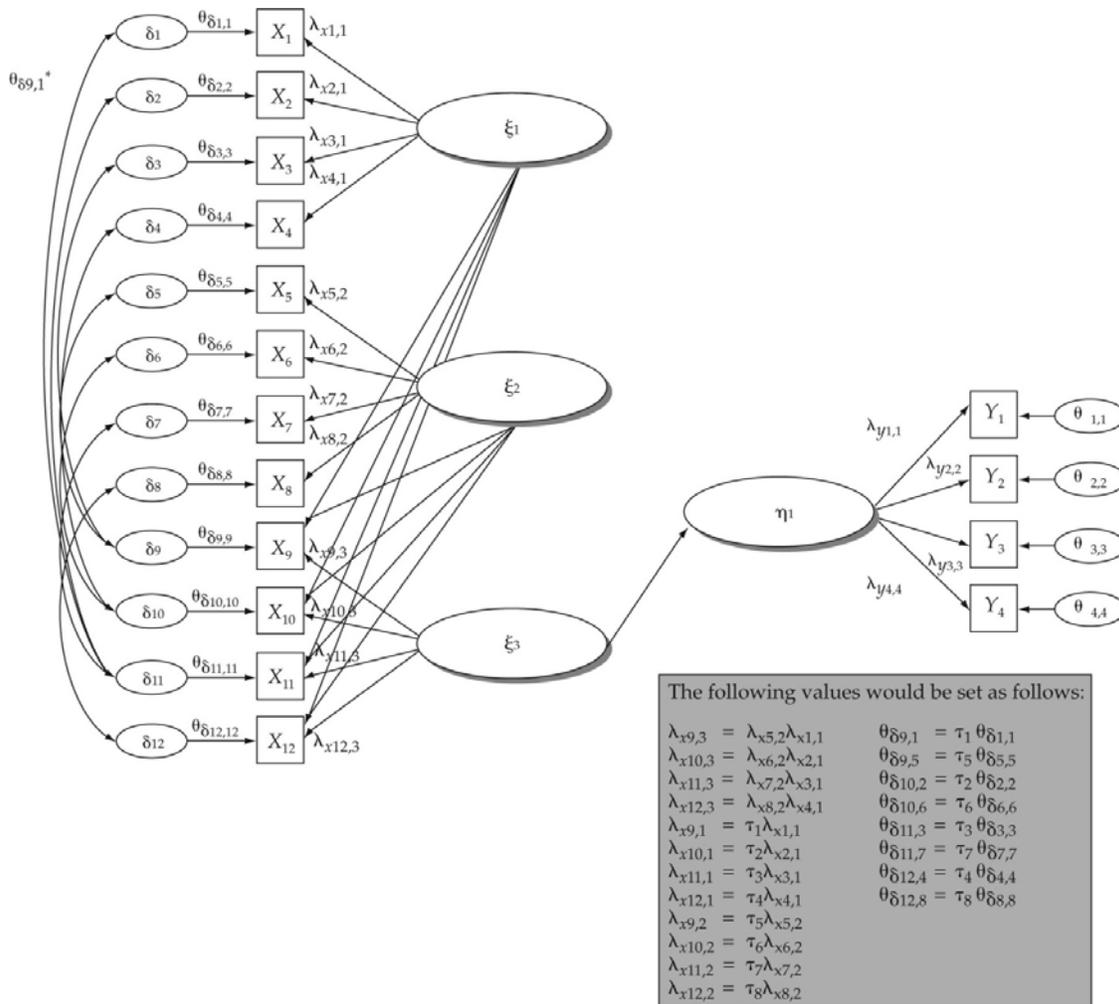
Estimating this model is complicated by several factors. These factors include the fact that the assumption that error terms are uncorrelated is no longer feasible because the loadings for the third construct (ξ_3) are a mathematical function of those for constructs 1 and 2 (ξ_1 and ξ_2). This fact leads to a quite complex SEM model setup that is recommended for advanced users only. So it is only briefly described here. This setup requires that the intercept terms for the measured items (τ_x) be estimated as described earlier. The exogenous factor loading pattern can no longer exhibit simple structure. Even though the loading estimates for the third construct can be computed by multiplying the loading estimates corresponding to the variables that created each interaction indicator, cross-construct loadings also exist for the interaction term. They are computed by crossing the matching τ_x terms with loading estimates. Again, this process is rather complex to follow, but as an example, the 12th row of Λ_x would end up as:

$$\left| \tau_4\lambda_4 \quad \tau_8\lambda_8 \quad \lambda_4\lambda_8 \right|$$

In addition, the error variance–covariance matrix for the *x* variables (θ_δ) must now include terms for the appropriate error covariance items that exist due to the computational nature of the intercept construct. These items need not be estimated because they are determined mathematically as the intercept term for the measured item used to compute the interaction indicator times the error variance for a construct. This concept is more easily illustrated by an example. δ_9 is the error variance term for x_9, the first indicator for the moderator construct. Because it is computed as x_1 times x_5, an error covariance term is needed for both $\theta_{\delta 9,1}$ and $\theta_{\delta 9,5}$. The values would be set as τ_1 times $\theta_{\delta 1,1}$ and τ_5 times $\theta_{\delta 5,5}$, respectively.

After finishing a setup following along these lines, the model can be estimated specifying only the structural path between the interaction construct and the outcome. If moderation is supported, the corresponding estimate, $\gamma_{3,1}$ in this case, would be significant. Realize that the effects of ξ_1 and ξ_2 on η_1 are of questionable validity in the presence of a significant interaction. Therefore, they should be estimated and interpreted only if the structural interaction term ($\gamma_{3,1}$) is insignificant.

Interaction terms sometimes cause problems with model convergence and distortion of the standard errors. Therefore, larger samples are often required to minimize the distortion. An absolute minimum sample size would be 300 for this type of analysis with a sample size of more than 500 recommended.