5

Specification

The specification of path analysis (PA) models, confirmatory factor analysis (CFA) measurement models, and structural regression (SR) models is the topic of this chapter. Outlined first are the basic steps of SEM and graphical symbols used in model diagrams. Some straightforward rules are suggested for counting the number of observations (which is not the sample size) in the analysis and the number of model parameters. Both of these quantities are needed for checking model identification (next chapter). Actual research examples dealt with in more detail in later chapters are also introduced. The main goal of this presentation is to give you a better sense of the kinds of hypotheses that can be tested with core structural equation models.

STEPS OF SEM

Six basic steps are followed in most analyses, and two additional optional steps, in a perfect world, would be carried out in every analysis. Review of these steps will help you to understand (1) the relation of specification, the main topic of this chapter, to later steps of SEM and (2) the utmost importance of specification.

Basic Steps

The basic steps are listed next and then discussed afterward, and a flowchart of these steps is presented in Figure 5.1. These steps are actually iterative because problems at a later step may require a return to an earlier step. (Later chapters elaborate specific issues at each step beyond specification for particular SEM techniques.)

1. Specify the model.
2. Evaluate model identification (if not identified, go back to step 1).

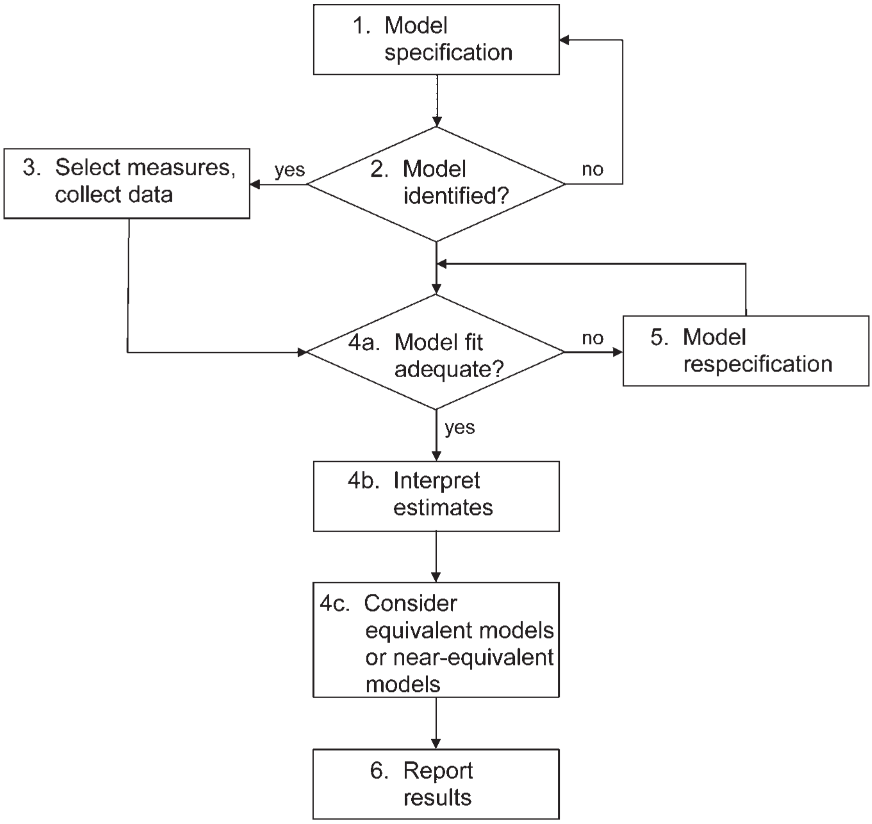

FIGURE 5.1. Flowchart of the basic steps of SEM.

3. Select the measures (operationalize the constructs) and collect, prepare, and screen the data.
4. Estimate the model:
   a. Evaluate model fit (if poor, skip to step 5).
   b. Interpret parameter estimates.
   c. Consider equivalent or near-equivalent models (skip to step 6).
5. Respecify the model (return to step 4).
6. Report the results.

Specification

The representation of your hypotheses in the form of a structural equation model is **specification**. Many researchers begin the process of specification by drawing a model diagram using a set of more or less standard graphical symbols (defined later), but the model can alternatively be described by a series of equations. These equations define the model's parameters, which correspond to presumed relations among observed or latent variables that the computer eventually estimates with sample data. *Specification is the most important step.* This is because results from later steps assume that the model is basically correct. I also suggest that you make a list of possible changes to the initial model that would be justified according to theory or empirical results. This is because it is often necessary to respecify models (step 5), and respecification should respect the same principles as specification.

Identification

If life were fair, the researcher could proceed directly from specification to collection of the data to estimation. Unfortunately, the analysis of a structural equation model is not always so straightforward. The problem that potentially complicates the analysis is that of **identification**. A model is identified if it is *theoretically* possible for the computer to derive a unique estimate of every model parameter. Otherwise, the model is not identified. The word "theoretically" emphasizes identification as a property of the model and not of the data. For example, if a model is not identified, then it remains so regardless of the sample size (N = 100, 1,000, etc.). Therefore, models that are not identified should be respecified (return to step 1); otherwise, attempts to analyze them may be fruitless. Different types of structural equation models must meet the specific requirements for identification that are described in Chapter 6.

Measure Selection and Data Collection

The various activities for this step—select good measures, collect the data, and screen them—were discussed in Chapter 3.

Estimation

This step involves using an SEM computer tool to conduct the analysis. Several things take place at this step: (1) Evaluate model fit, which means determine how well the model explains the data. Perhaps more often than not, researchers' initial models do not fit the data very well. When (not if) this happens to you, skip the rest of this step and go to the next, respecification, and then reanalyze the respecified model using the same data. Assuming satisfactory model fit, then (2) interpret the parameter estimates. In written summaries, too many researchers fail to interpret the parameter estimates for specific effects. Perhaps concern for overall model fit is so great that relatively little attention is paid to whether estimates of its parameters are meaningful (Kaplan, 2009). Next, (3) consider equivalent or near-equivalent models. Recall that an equivalent model explains the data just as well as the researcher's preferred model but does so with a different configuration of hypothesized relations among the same variables (Chapter 1). For a given model, there may be many—and in some cases infinitely many—equivalent versions. Thus, the researcher needs to explain why his or her preferred model should not be rejected in favor of statistically equivalent ones. Too many authors of SEM studies fail to even acknowledge the existence of equivalent models (MacCallum & Austin, 2000). There may also be near-equivalent models that fit the same data just about as well as the researcher's preferred model, but not exactly so. Near-equivalent models are often just as critical a validity threat as equivalent models, if not even more so.

Respecification

A researcher usually arrives at this step because the fit of his or her initial model is poor. In the context of model generation, now is the time to refer to that list of theoretically justifiable possible changes I suggested you make when you specified the initial model. We will deal with respecification in more detail in Chapter 8, but a bottom line of that discussion is that a model's respecification should be guided more by rational considerations than purely statistical ones. Any respecified model must be identified; otherwise, you will be "stuck" at this step until you have an estimable model.

Reporting the Results

The final step is to accurately and completely describe the analysis in written reports. The fact that too many published articles that concern SEM are seriously flawed in this regard was previously discussed. These blatant shortcomings are surprising considering that there are published guidelines for reporting results of SEM (e.g., Boomsma, 2000; McDonald & Ho, 2002; Schreiber, Nora, Stage, Barlow, & King, 2006). An integrated set of suggestions for reporting the results of SEM analyses is presented in Chapter 10.

Optional Steps

Two optional steps in SEM could be added to the basic ones just described:

7. Replicate the results.
8. Apply the results.

Replication

Structural equation models are seldom estimated across independent samples either by the same researchers who collected the original data (internal replication) or by other researchers who did not (external replication). The need for large samples in SEM complicates replication. Nevertheless, it is critical to eventually replicate a structural equation model if it is ever to represent anything beyond a mere statistical exercise.

Application

Kaplan (2009) notes that despite about 40 years of application of SEM in the behavioral sciences, rarely are results from SEM analyses used for policy or clinically relevant prediction studies. Neglecting to properly carry out the basic steps (1–6) may be part of the problem.

The ultimate goal of SEM—or any other type of model-fitting technique—is to attain what I refer to as **statistical beauty**, which means that the final retained model (if any):

1. Has a clear theoretical rationale (i.e., it makes sense).
2. Differentiates between what is known and what is unknown—that is, what is the model's range of convenience, or limits to its generality?
3. Sets conditions for posing new questions.

That most applications of SEM fall short of these goals should be taken as a positive incentive for all of us to do better. These issues are elaborated in Chapter 8 about hypothesis testing in SEM.

MODEL DIAGRAM SYMBOLS

Model diagrams are represented in this book by using symbols from the McArdle–McDonald reticular action model (RAM). The RAM symbolism explicitly represents every model parameter. This property has pedagogical value for learning about SEM. It also helps you to avoid mistakes when you are translating a diagram to the syntax of a particular SEM computer tool. Part of RAM symbolism is universal in SEM. This includes the representation in diagrams of

1. Observed variables with squares or rectangles (e.g., □, ▭).
2. Latent variables with circles or ellipses (e.g., ○, ⬭).
3. Hypothesized directional effects of one variable on another, or **direct effects**, with a line with a single arrowhead (e.g., →).
4. Covariances (in the unstandardized solution) or correlations (in the standardized one) between independent variables—referred to in SEM as **exogenous variables**—with a curved line with two arrowheads ().

The symbol described in (4) also designates an **unanalyzed association** between two exogenous variables. Although such associations are estimated by the computer, they are unanalyzed in the sense that no prediction is put forward about *why* the two exogenous variables covary (e.g., does one cause the other?—do they have a common cause?). In RAM symbolism (this next symbol is not universal), two-headed curved arrows that exit and reenter the same variable () represent the variance of an exogenous variable. Because the causes of exogenous variables are not represented in model diagrams, the exogenous variables are considered free to both vary and covary. The symbols and , respectively, reflect these assumptions. Specifically, the symbol

$\smile$ will connect every pair of observed exogenous variables, and the symbol $\frown$ will connect every observed or latent exogenous variable to itself in RAM symbolism.

This is not so for dependent (outcome, criterion) variables in model diagrams, which are referred to as **endogenous variables**. Unlike exogenous variables, the presumed causes of endogenous variables are explicitly represented in the model. Accordingly, endogenous variables are not free to vary or covary. This means in model diagrams that the symbol for an unanalyzed association, or $\smile$, does not directly connect two different endogenous variables, and the symbol for a variance $\frown$ will not originate from and end with any endogenous variable. Instead, the model as a whole represents the researcher's account about *why* endogenous variables covary with each other and also with the exogenous variables. During the analysis, this "explanation" based on the model is compared with the sample covariances (the data). If the two sets of covariances, predicted and observed, are similar, the model is said to fit the data; otherwise, the "explanation" is rejected.

Model parameters in RAM symbolism are represented with only three symbols: $\rightarrow$, $\smile$, and $\frown$. The following rule for defining parameters in words parallels these symbols and is consistent with the Bentler–Weeks representational system for SEM that underlies the EQS computer program:

Parameters of structural equation models when means are not analyzed include (1) direct effects on endogenous variables from other variables, either exogenous or endogenous; and (2) the variances and covariances of exogenous variables.	(Rule 5.1)

That's it. The simple rule just stated applies to all of the core SEM models described in this chapter when means are not analyzed (i.e., the model has a covariance structure only, not also a mean structure). An advantage of RAM symbolism is that you can quickly determine the number of model parameters simply by counting the number of $\rightarrow$, $\smile$, and $\frown$ symbols in its diagram. Several examples and exercises in counting parameters are presented later.

As mentioned in the previous chapter on SEM computer tools, model diagrams in Amos and Mx Graph are based on RAM symbolism. In other programs, such as LISREL and Mplus, error terms are represented by a line with a single arrowhead that points to the corresponding endogenous variables. This representation is more compact, but do not forget that error terms have parameters (variances) that are typically estimated in the analysis. This is one advantage of RAM symbolism: what you see is what you get concerning model parameters that require statistical estimates.

SPECIFICATION CONCEPTS

Considered next are key issues in model specification.

What to Include

The following is a basic specification issue: given a phenomenon of interest—health status, unemployment, and so on—what variables affect it? Because the literature for newer research areas can be limited, so decisions about what to include in the model must sometimes be guided more by the researcher's experience than by published reports. Consulting with experts in the field about plausible specifications may also help. In more established areas, sometimes there is too *much* information. That is, so many potential causal variables may be mentioned in the literature that it is virtually impossible to include them all. To cope, the researcher must again rely on his or her judgment about the most crucial variables.

The specification error of omitting causal variables that covary with others in the model has the same general consequences in SEM as in multiple regression (MR) (Chapter 2). However, it is unrealistic to expect all causal variables to be measured. Given that most structural equation models may be misspecified in this regard, the best way to minimize potential bias is preventive: make an omitted variable an included one through careful review of extant theory and research.

How to Measure the Hypothetical Construct

The selection of measures is a recurrent problem in research, and this is no less true in SEM (Chapter 3). Score reliability is especially important in the SEM technique of PA, which is characterized by **single-indicator measurement**. This means that there is only one observed measure of each construct. Therefore, it is critical that each measure have good psychometric characteristics. It is also assumed in PA that the exogenous variables are measured without error (r_{XX} = 1.00). The potential consequences of measurement error in PA are basically the same as those in MR (Chapter 2). Recall that disattenuating correlations for measurement error is one way to take score reliability into account (Equation 3.7), but this is not a standard part of PA. However, a method to do so for single-indicator measurement is described in Chapter 10.

Another approach is **multiple-indicator measurement**, in which more than one observed variable is used to measure the same construct. Suppose that a researcher is interested in measuring reading skill among Grade 4 children. In a single-indicator approach, the researcher would be forced to select a sole measure of reading skill, such as a word recognition task. However, a single task would reflect just one facet of reading, and some of its score variance may be specific to that task, not to general reading ability per se. In a multiple-indicator approach, additional measures can be selected and administered. In this example, a second measure could be a comprehension task, and a third measure could involve word attack skills. Use of the three tasks together may reflect more aspects of reading, and the reliability of factor measurement tends to be higher with multiple indicators.

Each measure in a multiple-indicator approach is represented in the model as a separate indicator of the same underlying factor. This representation assumes convergent validity. Specifically, scores from multiple indicators presumed to measure a common construct should be positively correlated. Otherwise, the measurement model for these indicators may be rejected. The technique of CFA and the analysis of SR models both feature multiple-indicator measurement. The analysis of an SR model in particular can be seen as a type of latent-variable PA that accommodates multiple-indicator measurement.

Directionality

The specification of directionalities of presumed causal effects, or **effect priority**, is an important part of SEM. In the technique of PA, specifications about directionality concern observed variables only. In path diagrams, direct effects represented by the symbol $\rightarrow$ (i.e., paths) correspond to the researcher's hypotheses about effect priority. For example, if X and Y are two observed variables, the specification $X \rightarrow Y$ implies that X is causally prior to Y (X affects Y). This specification does not rule out other causes of Y. If other variables are believed to also affect Y, then the corresponding direct effects (e.g., $W \rightarrow Y$) can be added to the model, too.

Five general conditions must be met before one can reasonably infer a cause–effect relation (e.g., Mulaik, 2009; Pearl, 2000):

1. *Temporal precedence*. The presumed cause (e.g., X) must occur before the presumed effect (e.g., Y).[1]
2. *Association*. There is an observed covariation; that is, variation in the presumed cause must be related to that in the presumed effect.
3. *Isolation*. There are no other plausible explanations (e.g., extraneous variables) of the covariation between the presumed cause and the presumed effect.
4. *Correct effect priority*. The direction of the causal relation is correctly specified. That is, X indeed causes Y ($X \rightarrow Y$) instead of the reverse ($Y \rightarrow X$) or X and Y cause each other in a reciprocal manner ($X \rightleftarrows Y$).
5. *Known distributional form*. When dealing with **probabilistic causality** instead of **deterministic causality**, the forms of the distributions of the parameters are specified. Deterministic causality assumes that given a change in the causal variable, the same consequence is observed in all cases for the affected variable. It is probabilistic causality that is modeled in SEM, and it allows for changes to occur in affected variables at some probability < 1.0.[2] Estimation of these probabilities (effects) with sample data are typically based on specific distributional

[1]See Rosenberg (1998) for a discussion of Immanuel Kant's arguments about the possibility of simultaneous causation.

[2]Kenny (1979) suggested that probabilistic causality models are compatible with the view that some portion of unexplained variance is fundamentally unknowable because it reflects, for lack of a better term, free will—the ability of people to act on occasion outside of external influences on them.

assumptions. If these assumptions are not reasonable, then the estimates may be incorrect.

The second and third conditions just listed require that the association between *X* and *Y* is not spurious when controlling for common causes or when other causes of *Y* are included in the model (e.g., *W*). Temporal precedence is established in experimental or quasi-experimental designs when treatment begins (and perhaps ends, too) before outcome is measured. In nonexperimental designs, the hypothesis that *X* causes *Y* would be bolstered if *X* is measured before *Y*; that is, the design is longitudinal. But the expected value of the covariance between *X* and *Y* in a longitudinal design could still be relatively large even if *Y* causes *X* and the effect (*X*) is measured before the cause (*Y*) (Bollen, 1989, pp. 61–65). This could happen because *X* would have been affected by *Y* before either variable was actually measured in a longitudinal study. This phenomenon explains the fourth requirement for correct specification of directionality: Even if *X* actually causes *Y*, the magnitude of their association may be low if the interval between their measurements is either too short (effects take time to materialize) or too long (temporary effects have dissipated). The fifth requirement explains the importance of distributional assumptions: Estimates of causal effects may be biased if assumptions about their distributional forms, such as normality, across random samples are not tenable.

The assessment of variables at different times provides a measurement framework consistent with the specification of directional effects. But longitudinal designs pose potential difficulties, such as case attrition and extra resource demands. This is probably why most SEM studies feature concurrent rather than longitudinal measurement. If all variables are measured simultaneously, however, it is not possible to demonstrate temporal precedence. Therefore, the researcher needs a clear, substantive rationale for specifying that *X* causes *Y* instead of the reverse (or that *X* and *Y* mutually influence each other) when all variables are measured at once. This process relies heavily on the researcher to rule out alternative explanations of the association between *X* and *Y* and also to measure other presumed causes of *Y*. Both require strong knowledge about the phenomena under study. If the researcher cannot give a cogent account of directionality specifications, then causal inferences in nonexperimental designs are unwarranted. This is why many researchers are skeptical about inferring causation in nonexperimental designs. An example follows.

Lynam, Moffit, and Stouthamer–Loeber (1993) hypothesized that poor verbal ability is a cause of delinquency, but both variables were measured simultaneously in their sample. This hypothesis raises some questions: Why this particular direction of causation? Is it not also plausible that certain behaviors associated with delinquency, such as truancy, could impair verbal ability? What about other causes of delinquency? Some arguments offered by Lynam et al. are summarized next: Their participants were relatively young (about 12 years), which may preclude delinquent careers long enough to affect verbal ability. They cited the results of prospective studies which indicated that low verbal ability precedes antisocial acts. Lynam et al. measured other presumed causes of delinquency, including social class and motivation, and controlled for these variables in the analysis. The particular arguments given by Lynam et al. are not above criticism (e.g., Block, 1995), but they exemplify the types of arguments that researchers should provide to justify directionality specifications. Unfortunately, too few authors of nonexperimental studies give such detailed explanations.

Given a single SEM study in which hypotheses about effect priority are tested, it would be almost impossible to believe that all of the logical and statistical requirements had been satisfied for interpreting the results as indicating causality. This is why the interpretation that direct effects in structural equation models correspond to true causal relations is typically without basis. It is only with the accumulation of the following types of evidence that the results of SEM analyses *may* indicate causality (Mulaik, 2000): (1) replication of the model across independent samples; (2) elimination of plausible equivalent or near-equivalent models; (3) corroborating evidence from empirical studies of variables in the model that are manipulable; and (4) the accurate prediction of the effects of interventions.

Although as students we are told time and again that *correlation does not imply causation*, too many researchers seem to forget this essential truth. For example, Robinson, Levin, Thomas, Pituch, and Vaughn (2007) reviewed about 275 articles published in five different journals in the area of teaching and learning. They found that (1) the proportion of studies based on experimental or quasi-experimental designs declined from about 45% in 1994 to 33% in 2004. Nevertheless, (2) the proportion of nonexperimental studies containing claims for causality increased from 34% in 1994 to 43% in 2004. It seems that researchers in the teaching-and-learning area—and, to be fair, in other areas, too—may have become less cautious than they should be concerning the inference of causation from correlation. Robinson et al. (2007) noted that more researchers in the teaching-and-learning area were using SEM in 2004 compared with 1994. Perhaps the increased use of SEM explains the apparent increased willingness to infer causation in nonexperimental designs, but the technique does not justify it.

There are basically three options in SEM if a researcher is uncertain about directionality: (1) specify a structural equation model but without directionality specifications between key variables; (2) specify and test alternative models, each with different causal directionalities; or (3) include reciprocal effects in the model as a way to cover both possibilities. The first option just mentioned concerns exogenous variables, which are basically always assumed to covary (e.g., $X_1 \leftrightarrow X_2$), but there is no specification about direct effects between exogenous variables. The specification of unanalyzed associations between exogenous variables in SEM is consistent with the absence of hypotheses of direct or indirect effects between such variables. A problem with the second option is that it can happen in SEM that different models, such as model 1 with $Y_1 \rightarrow Y_2$ and model 2 with $Y_2 \rightarrow Y_1$, may fit the same data equally well (they are equivalent), or nearly so. When this occurs, there is no statistical basis for choosing one model over another. The third option concerns the specification of reciprocal effects (e.g., $Y_1 \rightleftarrows Y_2$), but the specification of such effects is *not* a simple matter. This point is elaborated on later, but the inclusion of even one reciprocal effect in a model can make it more difficult to analyze. So there are potential costs to the inclusion of reciprocal effects as a hedge against uncertainty about directionality. If you are fundamentally uncertain about directionality, then you may not be ready to use SEM. In this case, conduct a **minimally sufficient analysis**, or use the simplest technique that will get the job done (Wilkinson & the Task Force on Statistical Inference, 1999). Simpler methods include regression techniques, such as canonical correlation when there are multiple predictor and outcome variables. A canonical correlation analysis requires no directionality assumptions among the variables in either set, predictor or outcome. There is no "embarrassment" in using a simpler statistical technique over a more complicated one, especially if the simpler technique is sufficient to test your hypotheses *and* if your comprehension of the more complex method is not strong. In general, it is better to resist the temptation to use the "latest and greatest" (i.e., more complicated) statistical technique when a simpler method will accomplish the task.

Model Complexity

There is another limit that must be respected in specification. It concerns the total number of parameters that can be estimated, or model complexity. This total is limited by the number of **observations** available for the analysis. In this context, the number of observations is *not* the sample size. Instead, it is literally the number of entries in the sample covariance matrix in lower diagonal form.[3] The number of observations can be calculated with a simple rule:

> If v is the number of observed variables, then the number of observations equals $v(v + 1)/2$ when means are not analyzed. (Rule 5.2)

Suppose that $v = 4$ observed variables are represented in a model. The number of observations is 4(5)/2, or 10. This count (10) equals the total number of variances (4) and unique covariances (below the diagonal, or 6) in the data matrix. With $v = 4$, the greatest number of parameters that could be estimated by the computer is 10. Fewer parameters can be estimated in a more parsimonious model, but not > 10. The number of observations has nothing to do with sample size. If four variables are measured for 100 or 1,000 cases, the number of observations is still 10. Adding cases does not increase the number of observations; only adding *variables* can do so.

The difference between the number of observations and the number of its parameters is the **model degrees of freedom**, or

$$df_M = p - q \tag{5.1}$$

where p is the number of observations (Rule 5.2) and q is the number of estimated

[3]Confusingly, LISREL uses the term *number of observations* in dialog boxes to refer to sample size, not the number of variances and unique covariances.

parameters (Rule 5.1). The requirement that there be at least as many observations as parameters can be expressed as the requirement that $df_M \geq 0$.

A model with more estimated parameters than observations ($df_M < 0$) is not amenable to empirical analysis. This is because such a model is not identified. If you tried to estimate a model with negative degrees of freedom, an SEM computer tool would likely terminate its run with error messages. Most models with zero degrees of freedom ($df_M = 0$) perfectly fit the data. But models that are just as complex as the data are not interesting because they test no particular hypothesis. Models with positive degrees of freedom generally do not have perfect fit. This is because $df_M > 0$ allows for the *possibility* of model–data discrepancies. Raykov and Marcoulides (2000) describe each degree of freedom as a dimension along which the model can *potentially* be rejected. Thus, retained models with greater degrees of freedom have withstood a greater potential for rejection. The idea underlies the **parsimony principle**: given two models with similar fit to the same data, the simpler model is preferred, assuming that the model is theoretically plausible.

Parameter Status

Each model parameter can be free, fixed, or constrained depending on its specification. A **free parameter** is to be estimated by the computer with the data. In contrast, a **fixed parameter** is specified to equal a constant. The computer "accepts" this constant as the estimate regardless of the data. For example, the hypothesis that X has no direct effect on Y corresponds to the specification that the coefficient for the path $X \rightarrow Y$ is fixed to zero. It is common in SEM to test hypotheses by specifying that a previously fixed-to-zero parameter becomes a free parameter, or vice versa. Results of such analyses may indicate whether to respecify a model by making it more complex (an effect is added—a fixed parameter becomes a free parameter) or more parsimonious (an effect is dropped—a free parameter becomes a fixed parameter).

A **constrained parameter** is estimated by the computer within some restriction, but it is not fixed to equal a constant. The restriction typically concerns the *relative* values of other constrained parameters. An **equality constraint** means that the estimates of two or more parameters are forced to be equal. Suppose that an equality constraint is imposed on the two direct effects that make up a feedback loop (e.g., $Y_1 \rightleftarrows Y_2$). This constraint simplifies the analysis because only one coefficient is needed rather than two. In a multiple-sample SEM analysis, a **cross-group equality constraint** forces the computer to derive equal estimates of that parameter across all groups. The specification corresponds to the null hypothesis that the parameter is equal in all populations from which the samples were drawn. How to analyze a structural equation model across multiple samples is explained in Chapter 9.

Other kinds of constraints are not seen as often. A **proportionality constraint** forces one parameter estimate to be some proportion of the other. For instance, the coefficient for one direct effect in a reciprocal relation may be forced to be three times the value of the other coefficient. An **inequality constraint** forces the value of a parameter estimate to be either less than or greater than the value of a specified constant. The specification that the value of an unstandardized coefficient must be > 5.00 is an example of an inequality constraint. The imposition of proportionality or inequality constraints generally requires knowledge about the relative magnitudes of effects, but such knowledge is rare in the behavioral sciences. A **nonlinear constraint** imposes a nonlinear relation between two parameter estimates. For example, the value of one estimate may be forced to equal the square of another. Nonlinear constraints are used in some methods to estimate curvilinear or interactive effects of latent variables, a topic covered in Chapter 12.

PATH ANALYSIS MODELS

Although PA is the oldest member of the SEM family, it is not obsolete. About 25% of roughly 500 articles reviewed by MacCallum and Austin (2000) concerned path models, so PA is still widely used. There are also times when there is just a single observed measure of each construct, and PA is a single-indicator technique. Finally, *if you master the fundamentals of PA, you will be better able to understand and critique a wider variety of structural equation models*. So read this section carefully even if you are more interested in latent variable methods in SEM.

Elemental Models

Presented in Figure 5.2 are the diagrams in RAM symbolism of three path models. Essentially, all more complex models can be constructed from these elemental models. A **path model** is a structural model for observed variables, and a **structural model** represents hypotheses about effect priority. The path model of Figure 5.2(a) represents the hypothesis that *X* is a cause of *Y*. By convention, causally prior variables are represented in the left part of the diagram, and their effects are represented in the right part. The line in the figure with the single arrowhead (→) that points from *X* to *Y* represents the corresponding direct effect. Statistical estimates of direct effects are **path coefficients**, which are interpreted just as regression coefficients in MR.

Variable *X* in Figure 5.2(a) is exogenous because its causes are not represented in the model. Accordingly, the symbol represents the fact that *X* is free to vary. In contrast, variable *Y* in Figure 5.2(a) is endogenous and thus is not free to vary. Each endogenous variable has a **disturbance**, which for the model of Figure 5.2(a) is an error (residual) term, designated as *D*, that represents unexplained variance in *Y*. It is the presence of disturbances in structural models that signal the assumption of probabilistic causality. Because disturbances can be considered latent variables in their own right, they are represented with circles in RAM symbolism. Theoretically, a disturbance can be seen as a "proxy" or composite variable that represents all unmeasured causes of the corresponding endogenous variable. Because the nature and number of these omitted causes is unknown as far as the model is concerned, disturbances can be viewed as

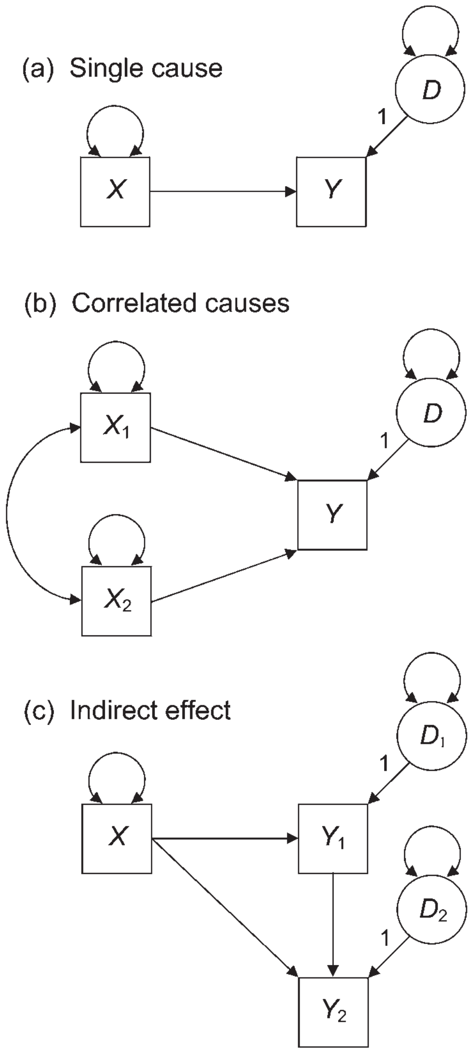

FIGURE 5.2. Elemental path models.

unmeasured (latent) exogenous variables. Accordingly, the symbol for the variance of an exogenous variable ( ) appears next to the disturbance in Figure 5.2(a).

Measurement error in the endogenous variable Y is manifested in its disturbance, so disturbances typically reflect both omitted causes and score unreliability. If scores on Y are unreliable, then its disturbance will be relatively large, which would be confounded with omitted causes. The path that points from the disturbance to the endogenous variable in Figure 5.2(a), or $D \rightarrow Y$, represents the direct effect of all unmeasured causes on Y. The numeral (1) that appears in the figure next to this path is a **scaling constant** that represents the assignment of a scale to the disturbance. This is necessary because disturbances are latent, and latent variables need scales before the computer can estimate anything about them. A scaling constant for a disturbance is also called an **unstandardized residual path coefficient**. The concept behind this specification for scaling a disturbance is explained in the next chapter, but it is required for identification. In contrast, exogenous variables do not have disturbances (e.g., X in Figure 5.2(a)). Therefore, it is generally assumed in PA that scores on exogenous variables are perfectly reliable. This assumption is just as unrealistic in PA as it is in MR.

Path coefficients are calculated holding all omitted causes constant (pseudoisolation; Chapter 2), which requires the assumption that all unmeasured causes represented by the disturbance are uncorrelated with measured causes of the corresponding endogenous variable. In Figure 5.2(a), it is assumed that D and X are uncorrelated. This is a strong assumption, one that is directly analogous to the assumption of uncorrelated residuals and predictors in MR.

The path model of Figure 5.2(b) represents the hypothesis of correlated causes. In this case, it is hypothesized that (1) both X_1 and X_2 are causes of Y, and (2) these exogenous variables covary. However, the model gives no account about *why* X_1 and X_2 covary. Accordingly, the curved line with two arrowheads that represents an unanalyzed association ($\smile$) connects the squares for the two measured exogenous variables in Figure 5.2(b). Together, the symbols $\frown$ and $\smile$ in the figure represent the assumptions that X_1 and X_2 are free to, respectively, vary and covary, but for reasons that are unknown, at least according to the model. Measured exogenous variables are basically always assumed to covary, so the symbol $\smile$ routinely connects every pair of such variables in structural models.

Path coefficients for the two direct effects in Figure 5.2(b), $X_1 \rightarrow Y$ and $X_2 \rightarrow Y$, are each estimated controlling for the covariation between X_1 and X_2, just as in MR. This model assumes that all unmeasured causes of Y are uncorrelated with both X_1 and X_2. A natural question is: If measured exogenous variables can have unanalyzed associations, can a disturbance have an unanalyzed association with a measured exogenous variable, such as $X_1 \smile D$? Such an association would imply the presence of an omitted cause that is correlated with X_1. This seems plausible, but, no, it is not generally possible to estimate covariances between and measured and unmeasured exogenous variables. (See Kenny, 1979, pp. 93–94 for conditions required to do so.) The only realistic way to cope with the restrictive assumption of uncorrelated measured and unmeasured causes is through careful specification.

Observe in the path model of Figure 5.2(c) that there are two direct effects on the endogenous variable Y_2 from other observed variables, one from the exogenous variable X and another from the other endogenous variable, Y_1. The latter specification gives Y_1 a dual role as, in the language of regression, both a predictor and a criterion. This dual role is described in PA as an **indirect effect** or a **mediator effect**.[4] Indirect effects involve one or more **intervening variables**, or **mediator variables**, presumed to "transmit"

[4]Note that the separate concept of a "moderator effect" refers to an interaction effect. Likewise, a "moderator variable" is one variable involved in interaction effect with another variable. Chapter 12 deals with the estimation of interaction effects in SEM.

some of the causal effects of prior variables onto subsequent variables. For the model of Figure 5.2(c), variable X is specified to affect Y_2 both directly and indirectly first by affecting Y_1, and then Y_1 in turn is presumed to have an effect on Y_2. The entire indirect effect just described corresponds to the three-variable chain $X \rightarrow Y_1 \rightarrow Y_2$.

Here is a concrete example: Roth, Wiebe, Fillingim, and Shay (1989) specified a path model of factors presumed to affect illness. Part of their model featured the indirect effect

$$\text{Exercise} \rightarrow \text{Fitness} \rightarrow \text{Illness}$$

The fitness variable is the mediator, one that, according to the model, is affected by exercise (more exercise, better fitness). In turn, fitness affects illness (better fitness, less illness). Just as direct effects are estimated in SEM, so too are indirect effects. The estimation of indirect effects is so straightforward in SEM that such effects are routinely included in structural models, assuming such specifications are theoretically justifiable.

Finally, the model of Figure 5.2(c) assumes that (1) the omitted causes of both Y_1 and Y_2 are uncorrelated with X and (2) the omitted causes of Y_1 are unrelated to those of Y_2, and vice versa. That is, the disturbances are independent, which is apparent in the figure by the *absence* of the symbol for an unanalyzed association ($\curvearrowright$) between D_1 and D_2. This specification also represents the hypothesis that the observed covariation between that pair of endogenous variables, Y_1 and Y_2, can be entirely explained by other measured variables in the model.

Types of Structural Models

There are two kinds of structural models. **Recursive models** are the most straightforward and have two basic features: their disturbances are uncorrelated, and all causal effects are unidirectional. **Nonrecursive models** have feedback loops or may have correlated disturbances. Consider the path models in Figure 5.3. The model of Figure 5.3(a) is recursive because its disturbances are independent and no observed variable is represented as both a cause and effect of another variable, directly or indirectly. For example, X_1, X_2, and Y_1 are specified as direct or indirect causes of Y_2, but Y_2 has no effect back onto one of its presumed causes. All of the models in Figure 5.2 are recursive, too. In contrast, the model of Figure 5.3(b) has a **direct feedback loop** in which Y_1 and Y_2 are specified as both causes and effects of each other ($Y_1 \rightleftarrows Y_2$). Each of these two variables is measured only once and also simultaneously. That is, direct feedback loops are estimated with data from a cross-sectional design, not a longitudinal design. **Indirect feedback loops** involve three or more variables, such as

$$Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y_1$$

Any model with an indirect feedback loop is automatically nonrecursive, too.

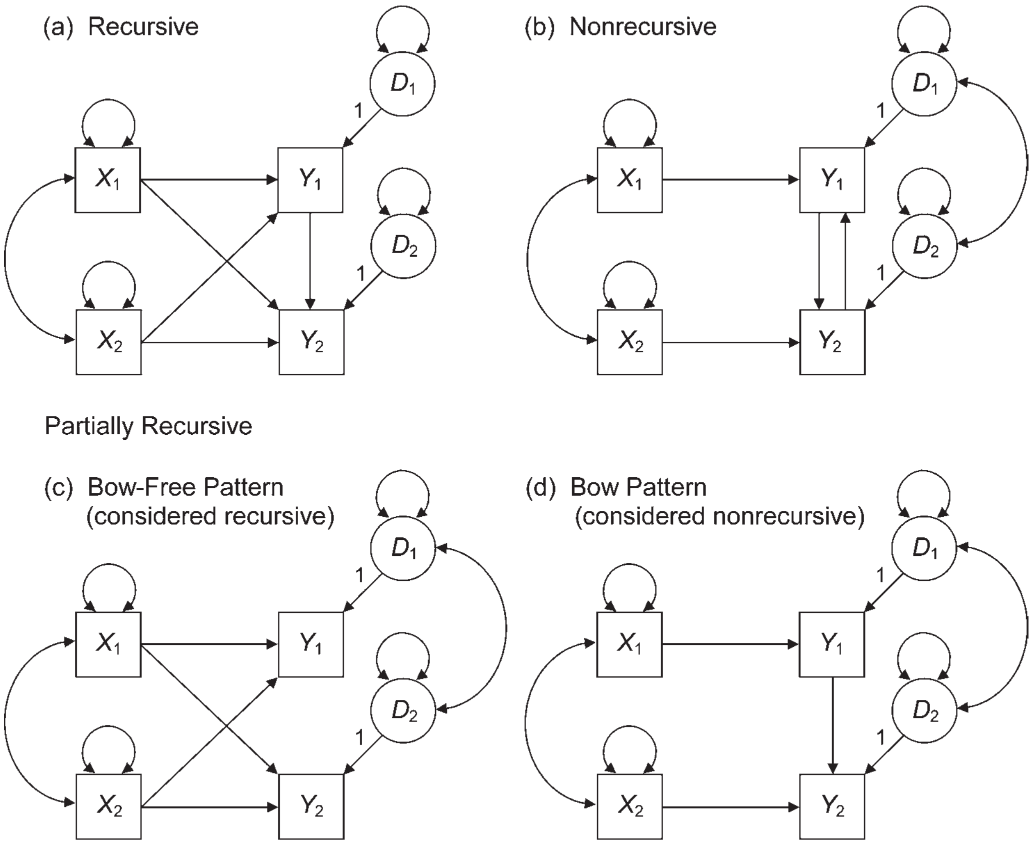

FIGURE 5.3. Examples of recursive and nonrecursive path models.

The model of Figure 5.3(b) also has a **disturbance covariance** (for unstandardized variables) or a **disturbance correlation** (for standardized variables). The term *disturbance correlation* is used from this point on regardless of whether or not the variables are standardized. A disturbance correlation, such as $D_1 \leftrightarrow D_2$, reflects the assumption that the corresponding endogenous variables (Y_1, Y_2) share at least one common omitted cause. Unlike unanalyzed associations between measured exogenous variables (e.g., $X_1 \leftrightarrow X_2$), the inclusion of disturbance correlations in the model is not routine. Why this is true is explained momentarily.

There is another type of path model, one that has unidirectional effects and correlated disturbances; two examples of this type are presented in Figures 5.3(c) and 5.3(d). Unfortunately, the classification of such models is not consistent. Some authors call these models nonrecursive, whereas others use the term **partially recursive**. But more important than the label for these models is the distinction made in the figure: Partially recursive models with a **bow-free pattern** of disturbance correlations can be treated in the analysis just like recursive models. A bow-free pattern means that correlated disturbances are restricted to pairs of endogenous variables *without* direct effects between them (see Figure 5.3(c)). In contrast, partially recursive models with a **bow pattern** of disturbance correlations must be treated in the analysis as nonrecursive models. A bow pattern means that a disturbance correlation occurs *with* a direct effect between that pair of endogenous variables (see Figure 5.3(d)) (Brito & Pearl, 2003). All ensuing references to recursive and nonrecursive models include, respectively, partially recursive models without and with direct effects among the endogenous variables.

Implications of the distinction between recursive and nonrecursive structural models are considered next. The assumptions of recursive models that all causal effects are unidirectional and that the disturbances are independent simplify the statistical demands for their analysis. For example, in the past MR was used to estimate path coefficients and disturbance variances in recursive path models. Today we use SEM computer tools to estimate recursive path models and all other kinds of models, too. The occurrence of a technical problem in the analysis is less likely for recursive models. It is also true that recursive structural models are identified, given that the necessary requirements for identification are satisfied (Chapter 6). The same assumptions of recursive models that ease the analytical burden are also restrictive. For example, causal effects that are not unidirectional, such as in a feedback loop, or disturbances that are correlated in a bow pattern cannot be represented in a recursive model.

The kinds of effects just mentioned can be represented in nonrecursive models, but such models require additional assumptions. Kaplan, Harik, and Hotchkiss (2001) remind us that data from a cross-sectional design give only a "snapshot" of an ongoing dynamic process. Therefore, the estimation of reciprocal effects in a feedback loop with cross-sectional data requires the assumption of **equilibrium**. This means that any changes in the system underlying a presumed feedback relation have already manifested their effects and that the system is in a steady state. That is, the values of the estimates of the direct effects that make up the feedback loop do not depend on the particular time point of data collection. Heise (1975) described equilibrium this way: it means that a dynamic system has completed its cycles of response to a set of inputs and that the inputs do not vary over time. That is, the causal process has basically dampened out and is not just beginning (Kenny, 1979). It is important to realize that there is generally no statistical way to directly evaluate whether the equilibrium assumption is tenable when the data are cross-sectional; that is, it must be argued substantively. Kaplan et al. (2001) note that rarely is this assumption explicitly acknowledged in the literature on applications of SEM where feedback effects are estimated with cross-sectional data. This is unfortunate because the results of computer simulation studies by Kaplan et al. (2001) indicate that violation of the equilibrium assumption can lead to severely biased estimates of the direct effects in feedback loops. Another assumption in the estimation of reciprocal effects in feedback loops with cross-sectional data is that of **stationarity**, the requirement that the causal structure does not change over time. Both assumptions just described, that of equilibrium and stationarity, are very demanding (i.e., probably unrealistic).

A feedback loop between Y_1 and Y_2 is represented in Figure 5.4(a) without disturbances or other variables. Another way to estimate reciprocal effects requires a longitudinal design where Y_1 and Y_2 are each measured at ≥ 2 different points in time. For example, the symbols Y_{11} and Y_{21} in the **panel model** shown in Figure 5.4(b) without disturbances or other variables represent, respectively, Y_1 and Y_2 at the first measurement occasion. Likewise, the symbols Y_{12} and Y_{22} represent the same two variables at the second measurement. Presumed reciprocal causation is represented in Figure 5.4(b) by the **cross-lag direct effects** between Y_1 and Y_2 measured at different times, such as $Y_{11} \rightarrow Y_{22}$ and $Y_{21} \rightarrow Y_{12}$. A panel model may be recursive or nonrecursive depending on its pattern of disturbance correlations.

Panel models for longitudinal data offer potential advantages over models with feedback loops for cross-sectional data. One is the explicit representation of a finite causal lag that corresponds to the measurement occasions. In this sense, the measurement occasions in a design where all variables are concurrently measured are *always* incorrect, if we assume that causal effects require a finite amount of time. However, the analysis of a panel model is no panacea for estimating reciprocal causality. For example, it can be difficult to specify measurement occasions that match actual causal lags. Panel designs are *not* generally useful for resolving effect priority between reciprocally related variables—for example, does Y_1 cause Y_2 or vice versa?—unless some restrictive assumptions are met, including that of stationarity. Maruyama (1998) reminds us that the requirement that there are no omitted causes correlated with those in the model is even more critical for panel models because of repeated sampling over time. The complexity of panel models can increase rapidly as more variables are added to the model (Cole & Maxwell, 2003). See Frees (2004) for more information about the analysis of panel data in longitudinal designs.

For many researchers, the estimation of reciprocal causation between variables measured simultaneously is the only viable alternative to a longitudinal design. Given all the restrictive assumptions for estimating such effects in a cross-sectional design, however, it is critical not to be too cavalier in the specification of feedback loops. One example is when different directionalities are each supported by two different theories (e.g., $Y_1 \rightarrow Y_2$ according to theory 1, $Y_2 \rightarrow Y_1$ according to theory 2). As mentioned, it can happen that two models with different directionality specifications among the same variables can fit the same data equally well. An even clearer example is when you haven't really thought through the directionality question. In this case, the specification of $Y_1 \rightleftarrows Y_2$ may be a smokescreen that covers up the basic uncertainty.

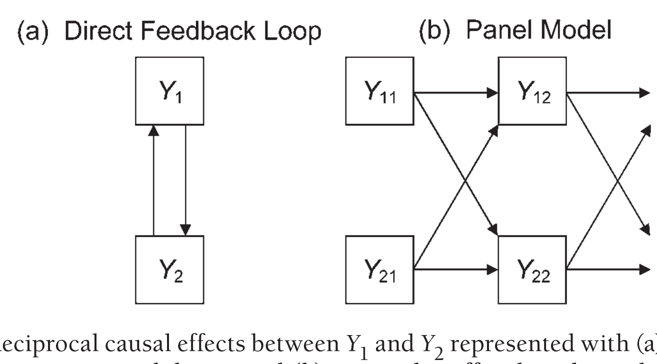

FIGURE 5.4. Reciprocal causal effects between Y_1 and Y_2 represented with (a) a direct feedback loop based on a cross-sectional design and (b) a cross-lag effect based on a longitudinal design (panel model) shown without disturbances or other variables.

Recall that the presence of a disturbance correlation reflects the assumption that the corresponding endogenous variables share at least one common unmeasured cause. The disturbances of variables involved in feedback loops are often specified as correlated. This specification often makes sense because if variables are presumed to mutually cause each other, then it seems plausible to expect that they may have shared omitted causes. In fact, the presence of disturbance correlations in particular patterns in nonrecursive models helps to determine their identification status (Chapter 6). In recursive models, disturbance correlations can be specified only between endogenous variables with no direct effect between them (e.g., Figure 5.3(c)). The addition of each disturbance correlation to the model "costs" one degree of freedom and thus makes the model more complicated. If there are substantive reasons for specifying disturbance correlations, then it is probably better to estimate the model with these terms than without them. This is because the constraint that a disturbance correlation is zero when there are common causes tends to redistribute this association toward the exogenous end of the model, which can result in biased estimates of direct effects. In general, disturbances should be specified as correlated if there are theoretical bases for doing so; otherwise, be wary of making the model overly complex by adding parameters without a clear reason.

Another complication of nonrecursive models is that of identification. There are some straightforward ways that a researcher can determine whether some, but not all, types of nonrecursive models are identified. These procedures are described in Chapter 6, but it is worthwhile to make this point now: adding exogenous variables is one way to remedy an identification problem of a nonrecursive model. However, this typically can only be done *before* the data are collected. *Thus it is critical to evaluate whether a nonrecursive model is identified right after it is specified and before the study is conducted.*

Before we continue, let's apply the rules for counting observations, parameters, and degrees of freedom to the recursive model in Figure 5.3(a). Because there are $v = 4$ observed variables in this model, the number of observations is 4(5)/2 = 10 (Rule 5.2). It is assumed that the constants (1) in the figure, such as that for the path $D_1 \rightarrow Y_1$, are fixed parameters that scale the disturbances. Applying Rule 5.1 for counting free parameters gives us the results that are summarized in Table 5.1. Because the number of observations and free parameters for this model are equal (10), the model degrees of freedom are zero ($df_M = 0$). Exercise 3 for this chapter asks you to count the number of parameters and df_M for the other path models in Figure 5.3.

PA Research Example

Presented in Figure 5.5 is a recursive path model of presumed causes and effects of positive teacher–pupil interactions analyzed in a sample of 109 high school teachers and 946 students by Sava (2002).[5] This model reflects the hypothesis that both the level of

[5]I renamed some of the variables in Figure 5.5 in order to clarify the meaning of low versus high scores in the Sava (2002) data set.

TABLE 5.1. Number and Types of Free Parameters for the Recursive Path Model of Figure 5.3(a)

		Endogenous variables		
Model	Direct effects on endogenous variables	Variances	Covariances	Total
Figure 5.3(a)	$X_1 \rightarrow Y_1$ $X_2 \rightarrow Y_1$ $X_1 \rightarrow Y_2$ $X_2 \rightarrow Y_2$ $Y_1 \rightarrow Y_2$	X_1, X_2 D_1, D_2	X_1 X_2	10

school support for teachers (e.g., resource availability) and a coercive view of student discipline that emphasizes a custodial approach to education affect teacher burnout. All three variables just mentioned are expected to affect the level of positive teacher–pupil interactions. In turn, better student–teacher interactions should lead to better school experience and general somatic status (e.g., less worry about school) on the part of students. Note in Figure 5.5 the absence of direct effects from school support, coercive control, and burnout to the two endogenous variables in the far right side of the model, school experience and somatic status. Instead, the model depicts the hypothesis of "pure" mediation through positive teacher–pupil interactions.

The article by Sava (2002) is a model in that it offers a clear account of specification and a detailed description of all measures, including internal consistency score reliabilities. Sava (2002) reported the data matrices analyzed (covariance, correlation) and used an appropriate method to analyze a correlation matrix without standard deviations. This author also reported all parameter estimates, both unstandardized and stan-

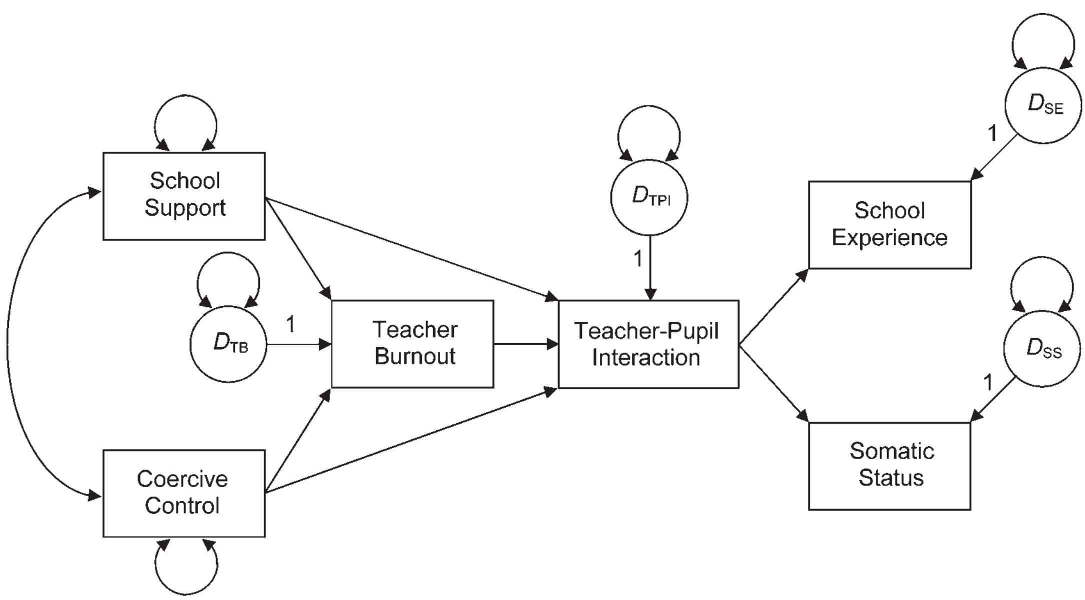

FIGURE 5.5. A path model of causes and effects of positive teacher–pupil interactions.

dardized, with the appropriate standard errors. However, Sava (2002) did not consider equivalent path models. Detailed analysis of the path model in Figure 5.5 is discussed in Chapter 7.

CFA MODELS

Issues in the specification of CFA models are considered next.

Standard CFA Models

The technique of CFA analyzes a priori measurement models in which both the number of factors and their correspondence with the indicators are explicitly specified. Presented in Figure 5.6 is an example of a **standard CFA model**—the type most often tested in the literature—with two factors and six indicators. This model represents the hypothesis that (1) indicators X_1–X_3 measure factor A, (2) X_4–X_6 measure factor B, and (3) the factors covary. Each indicator has a measurement error term, such as E_1 for indicator X_1. Standard CFA models have the following characteristics:

1. Each indicator is a continuous variable represented as having two causes—a single factor that the indicator is supposed to measure and all other unique sources of influence (omitted causes) represented by the error term.
2. The measurement errors are independent of each other and of the factors.
3. All associations between the factors are unanalyzed (the factors are assumed to covary).

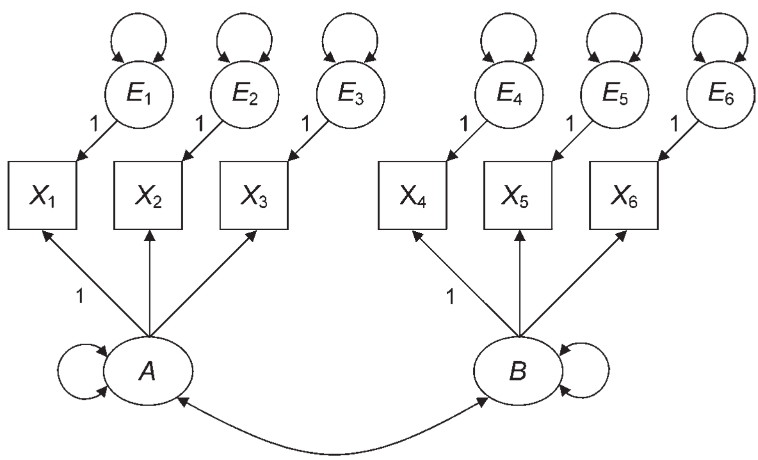

FIGURE 5.6. A standard confirmatory factor analysis model.

The lines with single arrowheads that point from a factor to an indicator, such as $A \rightarrow X_1$ in Figure 5.6, represent the presumed causal effect of the factor on the observed scores. Statistical estimates of these direct effects are called **factor loadings** or **pattern coefficients**, and they are generally interpreted as regression coefficients that may be in unstandardized or standardized form. Indicators assumed to be caused by underlying factors are referred to as **effect indicators** or **reflective indicators**. In this sense, indicators in standard CFA models are endogenous, and the factors are exogenous variables that are free to vary and covary. This also describes **reflective measurement**. The numeral (1) that appears in the figure next to the paths from the factors to one of their indicators (e.g., $B \rightarrow X_4$) are scaling constants that assign a metric to each factor, which allows the computer to estimate factor variances and covariances. The logic behind this specification and another option to scale factors is discussed in the next chapter, but scaling the factors is required for identification.

Each measurement error term in Figure 5.6 represents **unique variance**, a factor-analytic term for indicator variance not explained by the factors. Like disturbances in path models, measurement errors are proxy variables for all sources of residual variation that are not explained by the model. That is, they are unmeasured exogenous variables, so the symbol appears next to each of the error terms in the figure. The measurement errors in Figure 5.6 are specified as independent, which is apparent by the absence of the symbol for an unanalyzed association () that connects pairs of measurement error terms. This specification assumes that all omitted causes of each indicator are unrelated to those for all other indicators in the model. It is also assumed that the measurement errors are independent of the factors.

Two types of unique variance are represented by measurement errors: random error (score unreliability) and all sources of systematic variance not due to the factors. Examples of the latter type include systematic effects due to a particular measurement method or the particular stimuli that make up a task. When it is said that SEM takes account of measurement error, it is the error terms in measurement models to which this statement refers. The paths in the figure that point to the indicators from the measurement errors represent the direct effect of all unmeasured sources of unique variance on the indicators. The constants (1) that appear in the figure next to paths from measurement errors to indicators (e.g., $E_1 \rightarrow X_1$) represent the assignment of a scale to each term.

The representation in standard CFA models that each indicator has two causes, such as

$$A \rightarrow X_1 \leftarrow E_1$$

in Figure 5.6, is consistent with the view in classical measurement theory that observed scores (X) are comprised of two components: a true score (T) that reflects the construct of interest and a random error component (E) that is normally distributed with a mean of zero across all cases, or

$$X = T + E \quad (5.2)$$

The rationale that underlies the specification of reflective measurement in a standard CFA model comes from the **domain sampling model** (Nunnally & Bernstein, 1994, chap. 6). In this view of measurement, effect indicators X_1–X_3 in Figure 5.6 should as a set be internally consistent. This means that their intercorrelations should be positive and at least moderately high in magnitude (e.g., > .50). The same should also hold for indicators X_4–X_6 in the figure. Also, correlations among indicators of the same factor should be greater than cross-factor correlations. The patterns of indicator intercorrelations just described correspond to, respectively, convergent validity and discriminant validity in construct measurement. The domain sampling model also assumes that equally reliable effect indicators of the same construct are interchangeable (Bollen & Lenox, 1991). This means that the indicators can be substituted for each other without appreciably affecting construct measurement.

Sometimes the items of a particular indicator are negatively worded compared with other indicators of the same factor. Consequently, scores on that indicator will be negatively correlated with those from the other indicators, which is problematic from a domain sampling perspective. Suppose that a life satisfaction factor has three indicators. High scores on two indicators indicate greater contentment, but the third indicator is scaled to reflect degree of unhappiness, which implies negative correlations with scores from the other two indicators. In this case, the researcher could use **reverse scoring** or **reverse coding**, which reflects or reverses the scores on the negatively worded indicated indicator. One way to reflect the scores is to multiply them by –1.0 and then add a constant to the reflected scores so that the minimum score is at least 1.0 (Chapter 3). In this example, high scores on the unhappiness indicator are reflected to become low happiness scores, and vice versa. Now intercorrelations among all three indicators of the life satisfaction factor in this example should be positive.

It makes no sense to specify a factor with effect indicators that do not measure something in common. For example, suppose that the variables gender, ethnicity, and education are specified as effect indicators of a factor named "background" or some similar term. There are two problems here. First, gender and ethnicity are unrelated in representative samples, so one could not claim that these variables somehow measure a common domain.[6] Second, none of these indicators, such as a person's gender, is in any way "caused" by the some underlying "background" factor.

A common question about CFA concerns a minimum number of indicators per factor. In general, the absolute minimum for CFA models with two or more factors is two indicators per factor, which is required for identification. However, CFA models—and SR models, too—with factors that have only two indicators are more prone to problems in the analysis, especially in small samples. Also, it may be difficult to estimate measure-

[6] L. Wothke, personal communication, November 25, 2003.

ment error correlation for factors with only two indicators, which can result in a specification error. Kenny's (1979) rule of thumb about the number of indicators is apropos: "Two *might* be fine, three is better, four is best, and anything more is gravy" (p. 143; emphasis in original.)

Dimensionality of Measurement

The specifications that (1) each indicator loads on a single factor and (2) the error terms are independent describe **unidimensional measurement**. The first specification just mentioned describes **restricted factor models**. If any indicator loads on ≥ 2 factors or if its error term is assumed to covary with that of another indicator, then **multidimensional measurement** is specified. For example, adding the direct effect $B \rightarrow X_1$ to the model of Figure 5.6 would specify multidimensional measurement. There is controversy about allowing indicators to load on multiple factors. On the one hand, some indicators may actually measure more than one domain. An engineering aptitude test with text and diagrams, for instance, may measure both verbal and visual-spatial reasoning. On the other hand, unidimensional models offer more precise tests of the convergent and discriminant validity. For example, if every indicator in Figure 5.6 were allowed to load on both factors, an exploratory factor analysis (EFA) model that allows correlated factors (an oblique rotation) would be specified. It is **unrestricted factor models** that are estimated in EFA. (Other differences between CFA and EFA are outlined below.)

The specification of correlated measurement errors is a second way to represent multidimensional measurement. An error correlation reflects the assumption that the two corresponding indicators share something in common that is not explicitly represented in the model. Because error correlations are unanalyzed associations between latent exogenous variables (e.g., $E_1 \leftrightarrow E_2$), what this "something" may be is unknown as far as the model is concerned. Error term correlations may be specified as a way to test hypotheses about shared sources of variability over and beyond the factors. For example, the specification of error correlations for repeated measures variables represents the hypothesis of **autocorrelated errors**. The same specification can also reflect the hypothesis of a common method effect. In contrast, the absence of a measurement error correlation between a pair of indicators reflects the assumption that their observed correlation can be explained by their underlying factors. This refers to the **local independence assumption** that the indicators are independent, given the (correctly specified) latent variable model.[7]

The specification of multidimensional measurement makes a CFA model more complex compared with a standard (unidimensional) model. There are also implications for identification. Briefly, straightforward ways can be used to determine whether a standard CFA model is identified, but this may not be true for nonstandard models

[7]W. Wothke, personal communication, November 24, 2003.

(Chapter 6). It is important to evaluate whether nonstandard CFA models are identified when they are specified and before the data are collected. This is because one way to respecify a nonidentified CFA model is to add indicators, which increases the number of observations available to estimate effects.

Other Characteristics of CFA

The results of a CFA include estimates of factor variances and covariances, loadings of the indicators on their respective factors, and the amount of measurement error for each indicator. If the researcher's model is reasonably correct, then one should see the following pattern of results: (1) all indicators specified to measure a common factor have relatively high standardized factor loadings on that factor (e.g., > .70); and (2) estimated correlations between the factors are not excessively high (e.g., < .90 in absolute value). The first result indicates convergent validity; the second, discriminant validity. For example, if the estimated correlation between factors *A* and *B* in Figure 5.6 is .95, then the six indicators can hardly be said to measure two distinct constructs. If the results of a CFA do not support the researcher's a priori hypotheses, the measurement model can be respecified in the context of model generation (Chapter 1).

Hierarchical confirmatory factor analysis models depict at least one construct as a second-order factor that is not directly measured by any indicator. This exogenous second-order factor is also presumed to have direct effects on the first-order factors, which have indicators. These first-order factors are endogenous and thus do not have unanalyzed associations with each other. Instead, their common direct cause, the second-order factor, is presumed to explain the covariances among the first-order factors. Hierarchical models of intelligence, in which a general ability factor (*g*) is presumed to underlie more specific ability factors (verbal, visual-spatial, etc.), are examples of theoretical models that have been tested with hierarchical CFA. This special type of CFA model is discussed in Chapter 9.

Contrast with EFA

A standard statistical technique for evaluating measurement models is EFA. Originally developed by psychologists to test theories of intelligence, EFA is not generally considered a member of the SEM family. The term *EFA* refers to a class of procedures that include centroid, principal components, and principal (common) factor analysis methods that differ in their statistical criteria used to derive factors. This technique does not require a priori hypotheses about factor–indicator correspondence or even the number of factors. For example, all indicators are allowed to load on every factor; that is, EFA tests unrestricted factor models. There are ways to conduct EFA in a more confirmatory mode, such as instructing the computer to extract a certain number of factors based on theory. But the point is that EFA does not require specific hypotheses in order to apply it.

Another difference between CFA and EFA is that unrestricted factor models are not generally identified. That is, there is no single, unique set of parameter estimates for a given EFA model. This is because an EFA solution can be rotated an infinite number of ways. Among rotation options in EFA—varimax, quartimin, and promax to name just a few—researchers try to select one that clarifies factor interpretation. A parsimonious explanation in EFA corresponds to a solution that exhibits **simple structure** where each factor explains as much variance as possible in nonoverlapping sets of indicators (Kaplan, 2009). There is no need for rotation in CFA because factor models estimated in this technique are identified. Factors are allowed to covary in CFA, but the specification of correlated factors is not required in EFA (it is optional).

Cause Indicators and Formative Measurement

The assumption that indicators are caused by underlying factors is not always appropriate. Some indicators are viewed as **cause indicators** or **formative indicators** that affect a factor instead of the reverse. Consider this example by Bollen and Lennox (1991): The variables income, education, and occupation are used to measure socioeconomic status (SES). In a standard CFA model, these variables would be specified as effect indicators that are caused by an underlying SES factor (and by measurement errors). But we usually think of SES as the *outcome* of these variables (and others), not vice versa. For example, a change in any one of these indicators, such as a salary increase, may affect SES. From the perspective of **formative measurement**, SES is a *composite* that is caused by its indicators. Chapter 10 deals with formative measurement models.

CFA Research Example

Presented in Figure 5.7 is a standard CFA measurement model for the Mental Processing scale of the first edition Kaufman Assessment Battery for Children (KABC-I) (Kaufman & Kaufman, 1983), an individually administered cognitive ability test for children 2½ to 12½ years old. The test's authors claimed that the eight subtests represented in the figure measure two factors, sequential processing and simultaneous processing. The three tasks believed to reflect sequential processing all require the correct recall of auditory stimuli (Word Order, Number Recall) or visual stimuli (Hand Movements) in a particular order. The other five tasks represented in the figure are supposed to measure more holistic, less order-dependent reasoning, or simultaneous processing. Each of these tasks requires that the child grasp a "gestalt" but with somewhat different formats and stimuli.

The results of several CFA analyses of the KABC-I conducted in the 1980–1990s generally supported the two-factor model presented in Figure 5.7 (e.g., Cameron et al., 1997). However, other results have indicated that some subtests, such as Hand Movements, may measure both factors and that some of the measurement errors may covary (e.g., Keith, 1985). Detailed analysis of the model in Figure 5.7 with data for 10-year-olds from the KABC-I's normative sample is described in Chapter 9.

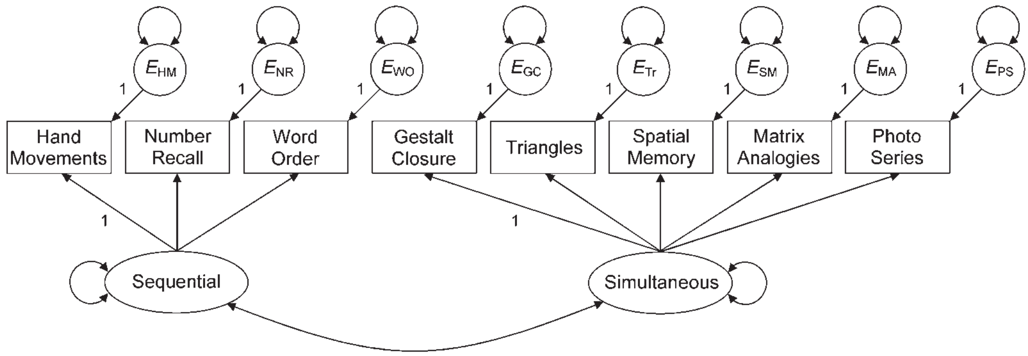

FIGURE 5.7. A confirmatory factor analysis model of the first-edition Kaufman Assessment Battery for Children.

STRUCTURAL REGRESSION MODELS

The most general kind of core structural equation model is an SR model, also called a full LISREL model. This term reflects the fact that LISREL was one of the first computer programs to analyze SR models, but any contemporary SEM computer tool can do so now. An SR model is the synthesis of a structural model and a measurement model. As in PA, the specification of an SR model allows tests of hypotheses about direct and indirect causal effects. Unlike path models, though, these effects can involve latent variables because an SR model also incorporates a measurement component that represents observed variables as indicators of underlying factors, just as in CFA. The capability to test hypotheses about both structural and measurement relations within a single model affords much flexibility.

Presented in Figure 5.8(a) is a structural model with observed variables—a path model—that features single-indicator measurement. The observed exogenous variable of this model, X_1, is assumed to be measured without error, an assumption usually violated in practice. This assumption is not required for the endogenous variables of this model, but measurement error in Y_1 or Y_3 is manifested in their disturbances. The model of Figure 5.8(b) is an SR model with both structural and measurement components. Its measurement model has the same three observed variables represented in the path model, X_1, Y_1, and Y_3. Unlike the path model, each of these three indicators in the SR model is specified as one of a pair for an underlying factor.[8] Consequently, (1) all the observed variables in Figure 5.8(b) have measurement error terms, and (2) effects for the endogenous latent variables, such as direct effects (e.g., $A \rightarrow B$) and disturbance variances (for D_B and D_C) are all estimated controlling for measurement error in the observed variables.

[8]I saved space in Figure 5.8 by showing only two indicators per factor, but remember that it is generally better to have at least three indicators per factor.

This SR model of Figure 5.8(b) also has a structural component that depicts the same basic pattern of direct and indirect causal effects as the path model but among latent variables ($A \rightarrow B \rightarrow C$) instead of observed variables. The structural model of Figure 5.8(b) is recursive, but it is also generally possible to specify an SR model with a nonrecursive structural model. Each latent endogenous variable in the structural model of Figure 5.8(b) has a disturbance (D_B, D_C). Unlike path models, the disturbances of SR models reflect only omitted causes and not also measurement error. For the same reason, path coefficients of the direct effects $A \rightarrow B$ and $B \rightarrow C$ in Figure 5.8(b) are corrected for measurement error, but those for the paths $X_1 \rightarrow Y_1$ and $Y_1 \rightarrow Y_3$ in Figure 5.8(a) are not.

The model of Figure 5.8(b) could be described as a **fully latent SR model** because every variable in its structural model is latent. Although this characteristic is desirable because it implies multiple-indicator measurement, it is also possible to represent in SR models an observed variable that is a single indicator of a construct. This reflects the reality that sometimes there is just a single measure of a some construct of interest. Such models could be called **partially latent SR models** because at least one variable in their structural model is a single indicator. However, unless measurement error of a single indicator is taken into account, partially latent SR models have the same limitations as path models outlined earlier. A way to address this problem for single indicators is described in Chapter 10.

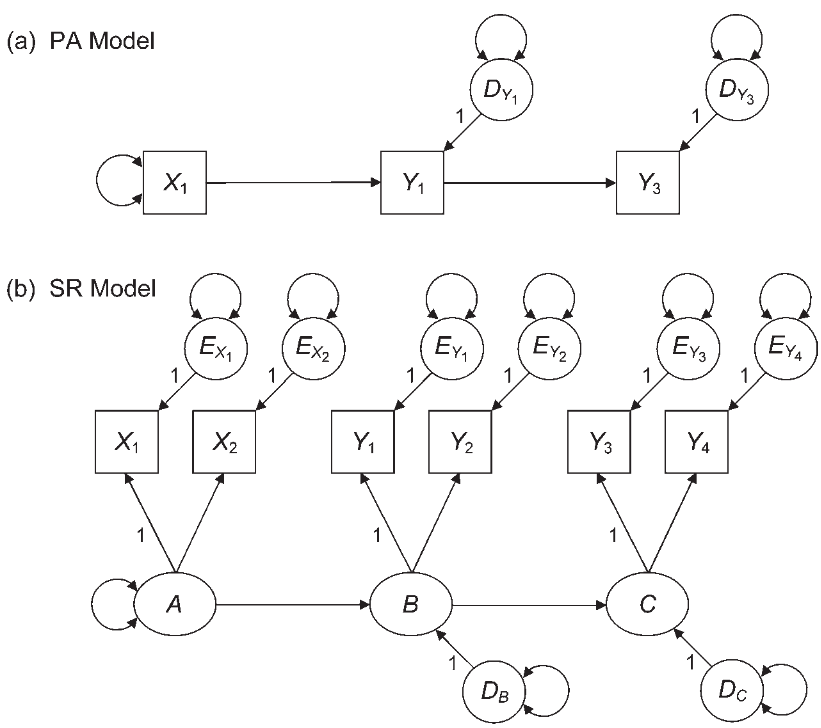

FIGURE 5.8. Examples of a path analysis model (a) and a structural regression model (b).

SR Model Research Example

Within a sample of 263 full-time university employees, Houghton and Jinkerson (2007) administered multiple measures of four constructs, including constructive (opportunity-oriented) thinking, dysfunctional (obstacle-oriented) thinking, subjective well-being (sense of psychological wellness), and job satisfaction. Based on their review of theory and empirical results in this area, Houghton and Jinkerson (2007) specified the four-factor fully latent SR model presented in Figure 5.9. The structural part of this model represents the hypotheses that (1) dysfunctional thinking and subjective well-being each have direct effects on job satisfaction; (2) constructive thinking has a direct effect on dysfunctional thinking; (3) the effect of constructive thinking on subjective well-being is mediated by dysfunctional thinking; and (4) the effects of constructive thinking on job satisfaction are mediated by the other two factors.

The measurement part of the SR model in Figure 5.9 features three indicators per factor. Briefly, indicators of (1) constructive thinking include measures of belief evaluation, positive self-talk, and positive visual imagery; (2) dysfunctional thinking includes two scales regarding worry about performance evaluations and a third scale about need for approval; (3) subjective well-being include ratings about general happiness and two positive mood rating scales; and (4) job satisfaction include three scales that reflect one's work experience as positively engaging.

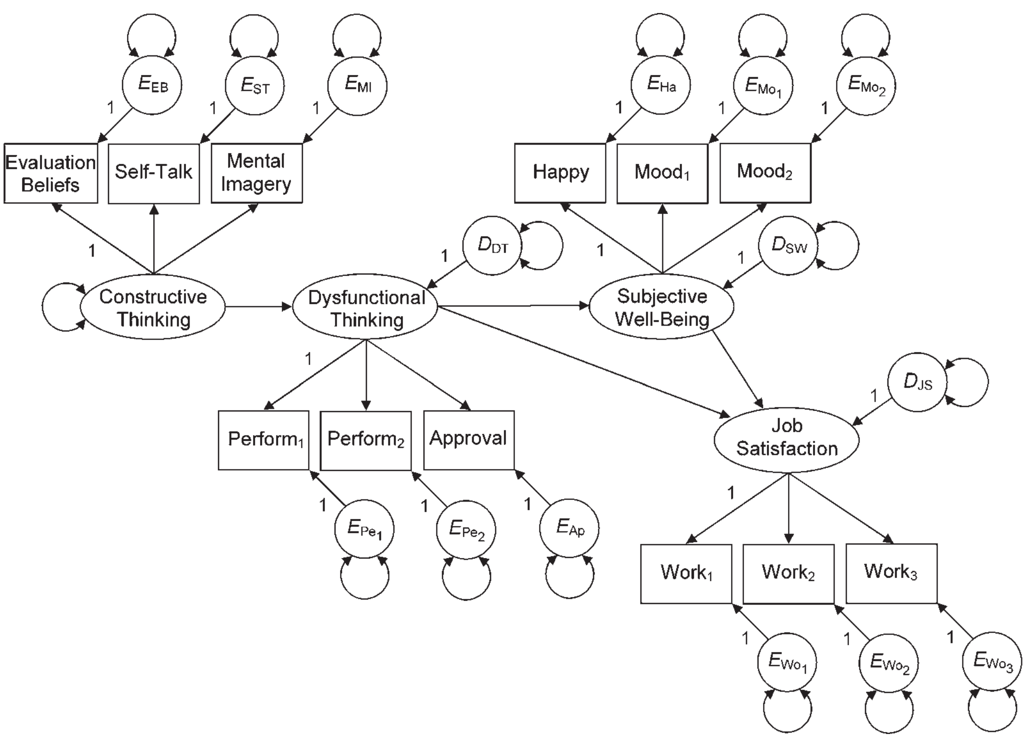

FIGURE 5.9. A structural regression model of factors of job satisfaction.

The article by Houghton and Jinkerson (2007) is exemplary in that the authors describe the theoretical rationale for each and every direct effect among the four factors in the structural model, provide detailed descriptions of all indicators including internal consistency score reliabilities, report the correlations and standard deviations for the covariance data matrix they analyzed, and test alternative models. However, Houghton and Jinkerson (2007) did not report unstandardized parameter estimates, nor did they consider equivalent versions of their final model. The detailed analysis of this SR model is described in Chapter 10.

EXPLORATORY SEM

Recall that Mplus has capabilities for exploratory structural equation modeling (ESEM) (Chapter 4). In ESEM, some parts of the measurement model are unrestricted instead of restricted. That is, the analysis incorporates features of both EFA and SEM. This type of analysis may be suitable when the researcher has weaker hypotheses about multiple-indicator measurement of some constructs than is ordinarily represented in CFA or SR models. Consider the ESEM model presented in Figure 5.10, which is also described in the Mplus 6 manual (Muthén & Muthén, 1998–2010, p. 90). The measurement model for factors *A* and *B* in the figure is an unrestricted EFA model where the indicators are allowed to load on every factor. In Mplus, the factor solution for this part of the model will be rotated according to the method specified by the user. Factors *A* and *B* are scaled by fixing their variances to 1.0, which standardizes them. In contrast, the measurement model for factors *C* and *F* in the figure is restricted where each indicator loads on a single factor. There is a structural model in Figure 5.10, too, and it features direct or indirect effects from the exogenous factors *A* and *B* onto the endogenous factors *C* and *F*. See Asparouhov and Muthén (2009) for more information about ESEM.

SUMMARY

Considered in this chapter were the specification of core SEM models and the types of research questions that can be addressed in their analysis. Path analysis allows researchers to specify and test structural models that reflect a priori assumptions about spurious associations and direct or indirect effects among observed variables. Measurement models that represent hypotheses about relations between indicators and factors can be evaluated with the technique of confirmatory factor analysis. Structural regression models with both a structural component and a measurement component can also be analyzed. Rules that apply to all the kinds of models just mentioned for counting the number of observations and the number of model parameters were also considered. The counting rules just mentioned are also relevant for checking whether a structural equation model is identified, which is the topic of the next chapter.

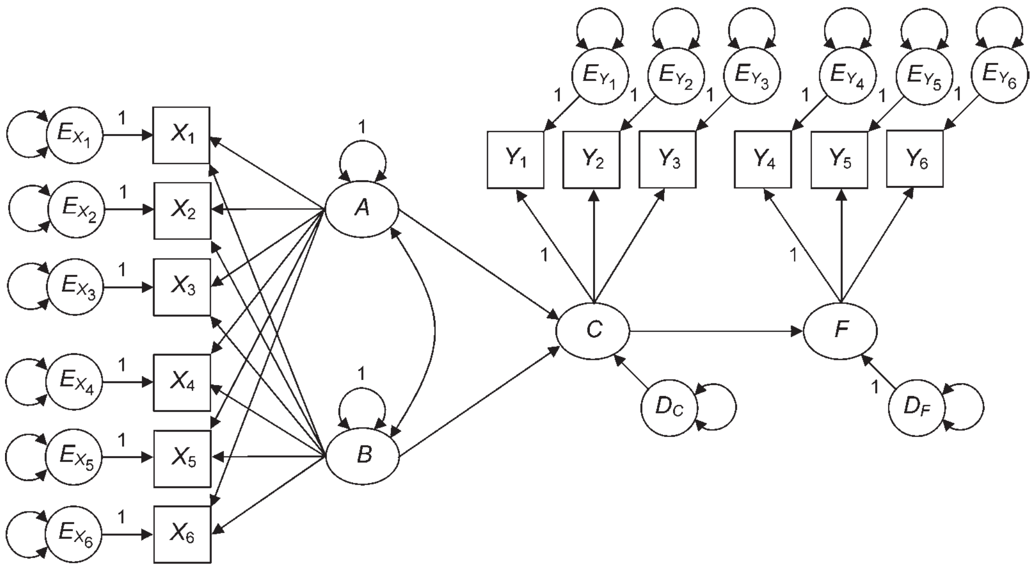

FIGURE 5.10. An exploratory structural equation model.

RECOMMENDED READINGS

MacCallum and Austin (2000) and Shah and Goldstein (2006) describe various types of shortcomings in articles published in psychology, education, and business journals in which results of SEM analyses are reported. Holbert and Stephenson (2002) survey the use of SEM in communication and note some of the same problems. All three articles should provide you with a good sense of common specification pitfalls to avoid.

Holbert, R. L., & Stephenson, M. T. (2002). Structural equation modeling in the communication sciences, 1995–2000. *Human Communication Research, 28*, 531–551.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201–236.

Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management, 24*, 148–169.

EXERCISES

1. What is the "explanation" of Figure 5.3a about why scores on Y_1 and Y_2 are correlated?

2. Does the CFA model of Figure 5.6 have a structural component?

3. Count the number of free parameters for the path models of Figures 5.3(b)–5.3(d).

4. Calculate the model degrees of freedom for (a) Figure 5.5, (b) Figure 5.7, and (c) Figure 5.9.

5. How are covariates represented in structural models?

6. Respond to this question: "I am uncertain about the direction of causality between Y_1 and Y_2. In SEM, why can't I just specify two different models, one with $Y_1 \rightarrow Y_2$ and the other with $Y_2 \rightarrow Y_1$, fit both models to the same data, and then pick the model with the best fit?"

7. What is the difference between a measurement error (E) and a disturbance (D)?

8. Specify a path model where the effects of a substantive exogenous variable X_1 on the outcome variable Y_2 are entirely mediated through variable Y_1. Also represent in the model the covariate X_2 (e.g., level of education in years).

9. What is the role of sample size in SEM?