Byrne, B. M., 2006: Structural Equation Modeling with EQS. Basic Concepts, Applications and Programming (2nd edition). Mahwah: Erlbaum, S. 3-17.

1

Structural Equation Models: The Basics

Structural equation modeling (SEM) is a statistical methodology that takes a confirmatory (i.e., hypothesis-testing) approach to the analysis of a structural theory bearing on some phenomenon. Typically, this theory represents "causal" processes that generate observations on multiple variables (Bentler, 1988). The term *structural equation modeling* conveys two important aspects of the procedure: (a) that the causal processes under study are represented by a series of structural (i.e., regression) equations, and (b) that these structural relations can be modeled pictorially to enable a clearer conceptualization of the theory under study. The hypothesized model can then be tested statistically in a simultaneous analysis of the entire system of variables to determine the extent to which it is consistent with the data. If goodness-of-fit is adequate, the model argues for the plausibility of postulated relations among variables; if it is inadequate, the tenability of such relations is rejected.

Several aspects of SEM set it apart from the older generation of multivariate procedures. First, as noted previously, it takes a confirmatory rather than an exploratory approach to the data analysis (although aspects of the latter can be addressed). Furthermore, by demanding that the pattern of intervariable relations be specified a priori, SEM lends itself well to the analysis of data for inferential purposes. By contrast, most other multivariate procedures are essentially descriptive by nature (e.g., exploratory factor analysis [EFA]) so that hypothesis-testing is difficult if not impossible. Second, whereas traditional multivariate procedures are incapable of either assessing or correcting for measurement error, SEM provides explicit estimates of these error variance parameters. Indeed, alternative methods (e.g., those rooted in regression or the general linear model) assume that error in the explanatory (i.e., independent) variables vanishes. Thus, applying those methods when there is error in the explanatory variables is tantamount to ignoring error that may lead, ultimately, to serious inaccuracies—especially when the errors are sizeable. Such mistakes are avoided when corresponding SEM analyses (in simple terms) are used (T. Raykov, personal communication, March 30, 2000). Third, whereas data analyses using the former methods are based on observed measurements only, those using SEM procedures can incorporate both unobserved (i.e., latent) and observed variables. Finally, there are no widely and easily applied alternative methods for modeling multivariate relations or for estimating point and/or interval indirect effects; these important features are available using SEM methodology.

Given these highly desirable characteristics, SEM has become a popular methodology for nonexperimental research in which methods for testing theories are not well developed and ethical considerations make experimental design unfeasible (Bentler, 1980). SEM can be utilized effectively to address numerous research problems involving nonexperimental research; in this book, I illustrate the most common applications (e.g., chaps. 3, 4, 6, 7, and 8) as well as some that are less frequently found in the substantive literature (e.g., chaps. 5, 9, 10, 11, 12, and 13). In each chapter, I walk you through the entire analytic process, from specification of the hypothesized model for the input file to interpretation of findings in the output file. Throughout the book, we work with EQS 6.1, the most recent version of the EQS program (Bentler, 2005; Bentler & Wu, 2002). Before showing you how to use EQS in testing SEM models, however, it is essential that I first review key concepts associated with the methodology. We turn now to their brief explanation.

BASIC CONCEPTS

Latent Versus Observed Variables

In the behavioral sciences, researchers are often interested in studying theoretical constructs that cannot be observed directly. These abstract phenomena are termed *latent variables*, or *factors*. Examples of latent variables in psychology are self-concept and motivation; in sociology, powerlessness and anomie; in education, verbal ability and teacher expectancy; and in economics, capitalism and social class.

Because latent variables are not observed directly, it follows that they cannot be measured directly. Thus, the researcher must operationally define the latent variable of interest in terms of behavior believed to represent it. As such, the unobserved variable is linked to one that is observable, thereby making its measurement possible. Assessment of the behavior, then, constitutes the direct measurement of an

observed variable albeit the indirect measurement of an unobserved variable (i.e., the underlying construct). The term *behavior* is used here in the very broadest sense to include scores on a particular measuring instrument. Thus, observation may include, for example, self-report responses to an attitudinal scale, scores on an achievement test, *in vivo* observation scores representing some physical task or activity, coded responses to interview questions, and the like. These measured scores (i.e., measurements) are termed *observed* or *manifest variables*; within the context of SEM methodology, they serve as indicators of the underlying construct that they are presumed to represent. Given this necessary bridging process between observed variables and unobserved latent variables, it should now be clear why methodologists urge researchers to be circumspect in their selection of assessment measures. Whereas the choice of psychometrically sound instruments bears importantly on the credibility of all study findings, such selection becomes even more critical when the observed measure is presumed to represent an underlying construct.[1]

Exogenous Versus Endogenous Latent Variables

In working with SEM models, it is helpful to distinguish between latent variables that are exogenous and those that are endogenous. *Exogenous latent variables* are synonymous with independent variables; they "cause" fluctuations in the values of other latent variables in the model. Changes in the values of exogenous variables are not explained by the model; rather, they are considered to be influenced by other factors external to it. Background variables, such as gender, age, and socioeconomic status, are examples of such external factors. *Endogenous latent variables* are synonymous with dependent variables and, as such, are influenced by the exogenous variables in the model, either directly or indirectly. Fluctuation in the values of endogenous variables is said to be explained by the model because all latent variables that influence them are included in the model specification. In the interest of both simplicity and consistency throughout the remainder of the book, the terms *independent* and *dependent* rather than exogenous and endogenous variables, respectively, are used in the various SEM models described.

The Factor Analytic Model

The oldest and best known statistical procedure for investigating relations between sets of observed and latent variables is that of factor analysis. In using this approach to data analyses, the researcher examines the covariation among a set of observed variables to gather information on their underlying latent constructs (i.e., factors). There are two basic types of factor analyses: exploratory factor analysis (EFA) and

[1]Throughout the remainder of the book, the terms *latent*, *unobserved*, and *unmeasured* variables are used synonymously to represent a hypothetical construct or factor; the terms *observed*, *manifest*, and *measured* variables also are used interchangeably.

confirmatory factor analysis (CFA). We turn now to a brief description of each. However, for a more extensive discussion of each, see Byrne (2005a; 2005b).

EFA is designed for the situation in which links between the observed and latent variables are unknown or uncertain. The analysis thus proceeds in an exploratory mode to determine how and to what extent the observed variables are linked to their underlying factors. Typically, the researcher wishes to identify the minimal number of factors that underlie (or account for) covariation among the observed variables. For example, suppose a researcher develops a new instrument designed to measure five facets of physical self-concept (i.e., Health, Sport Competence, Physical Appearance, Coordination, and Body Strength). Following the formulation of questionnaire items designed to measure these five latent constructs, the researcher would then conduct an EFA to determine the extent to which the item measurements (the observed variables) were related to the five latent constructs. In factor analysis, these relations are represented by *factor loadings*. The researcher would hope that items designed to measure health, for example, exhibited high loadings on that factor albeit low or negligible loadings on the other four factors. This factor analytic approach is considered exploratory in the sense that the researcher has no prior knowledge that the items do indeed measure the intended factors. (For texts that discuss EFA, see Comrey, 1992; Gorsuch, 1983; McDonald, 1985; and Mulaik, 1972. For informative articles on EFA, see Fabrigar, Wegener, MacCallum, & Strahan, 1999; MacCallum, Widaman, Zhang, & Hong, 1999; Preacher & MacCallum, 2003; and Wood, Tataryn, & Gorsuch, 1996.)

In contrast to EFA, CFA is appropriately used when the researcher has some knowledge of the underlying latent variable structure. Based on knowledge of the theory, empirical research, or both, the researcher postulates relations between the observed measures and the underlying factors a priori and then tests this hypothesized structure statistically. For instance, based on the previous example, the researcher would argue for the loading of items designed to measure sport-competence self-concept on that specific factor and not on the health, physical appearance, coordination, or body-strength self-concept dimensions. Accordingly, a priori specification of the CFA model would allow all sport-competence self-concept items to be free to load on that factor but restricted to have zero loadings on the remaining factors. The model would then be evaluated by statistical means to determine the adequacy of its goodness-of-fit to the sample data. (For more detailed discussions of CFA, see, e.g., Bollen, 1989a; Long, 1983a; and Raykov & Marcoulides, 2000.)

In summary, the factor analytic model (i.e., EFA or CFA) focuses solely on how and the extent to which the observed variables are linked to their underlying latent factors. More specifically, it is concerned with the extent to which the observed variables are generated by the underlying latent constructs and, thus, strength of the regression paths from the factors to the observed variables (i.e., the factor loadings) are of primary interest. Although interfactor relations are also of interest, any regression structure among them is not considered in the factor analytic model.

Because the CFA model focuses solely on the link between factors and their measured variables, within the framework of SEM, it represents what is termed a *measurement model*.

The Full Latent Variable Model

In contrast to the factor analytic model, the full latent variable model allows for the specification of regression structure among the latent variables. That is, the researcher can hypothesize the impact of one latent construct on another in the modeling of causal direction. This model is termed *full* (or *complete*) because it comprises both a measurement model and a structural model: the measurement model depicts the links between the latent variables and their observed measures (i.e., the CFA model), and the structural model depicts the links among the latent variables themselves.

A full latent variable model that specifies direction of cause from one direction only is termed a *recursive model*; one that allows for reciprocal or feedback effects is termed a *nonrecursive model*. Only applications of recursive models are considered in this book.

General Purpose and Process of Statistical Modeling

Statistical models provide an efficient and convenient way to describe the latent structure underlying a set of observed variables. Expressed either diagrammatically, or mathematically via a set of equations, such models explain how the observed and latent variables are related to one another.

Typically, researchers postulate a statistical model based on their knowledge of the related theory, on empirical research in the area of study, or on some combination of both. Once the model is specified, the researcher then tests its plausibility based on sample data that comprise all observed variables in the model. The primary task in this model-testing procedure is to determine the goodness-of-fit between the hypothesized model and the sample data. As such, the researcher imposes the structure of the hypothesized model on the sample data and then tests how well the observed data fit this restricted structure. Because it is highly unlikely that a perfect fit will exist between the observed data and the hypothesized model, there is necessarily a differential between the two; this differential is termed the *residual*. The model-fitting process can therefore be summarized as follows:

$$\text{Data} = \text{Model} + \text{Residual}$$

where

Data represent score measurements related to the observed variables as derived from persons comprising the sample.

Model represents the hypothesized structure linking the observed variables to the latent variables and, in some models, linking particular latent variables to one another.

Residual represents the discrepancy between the hypothesized model and the observed data.

In summarizing the general strategic framework for testing structural equation models, Jöreskog (1993) distinguished among three scenarios that he termed *strictly confirmatory*, *alternative models*, and *model-generating*. In the strictly confirmatory scenario, the researcher postulates a single model based on theory, collects the appropriate data, and tests the fit of the hypothesized model to the sample data. From the results of this test, the researcher either rejects or fails to reject the model; no further modifications to the model are made. In the alternative models scenario, the researcher proposes several alternative (i.e., competing) models, all of which are grounded in theory. Following analysis of a single set of empirical data, the researcher selects one model as most appropriate in representing the sample data. Finally, the model generating scenario represents the case in which the researcher, having postulated and rejected a theoretically derived model on the basis of its poor fit to the sample data, proceeds in an exploratory (rather than confirmatory) manner to modify and reestimate the model. The primary focus in this instance is to locate the source of misfit in the model and to determine a model that better describes the sample data. Jöreskog (1993) notes that although respecification may be either theory- or data-driven, the ultimate objective is to find a model that is both substantively meaningful and statistically well fitting. He further posits that despite the fact that "a model is tested in each round, the whole approach is model generating, rather than model testing" (Jöreskog, 1993, p. 295).

Of course, even a cursory review of the empirical literature clearly shows the model generating scenario to be the most common of the three—and for good reason. Given the many costs associated with the collection of data, it would be a rare researcher indeed who could afford to terminate his or her research on the basis of a rejected hypothesized model! As a consequence, the strictly confirmatory scenario is not commonly found in practice. Although the alternative models approach to modeling has also been a relatively uncommon practice, at least two important papers on the topic (i.e., MacCallum, Roznowski, & Necowitz, 1992; and MacCallum, Wegener, Uchino, & Fabrigar, 1993) recently precipitated more activity with respect to this analytic strategy.

Statistical theory related to these model-fitting processes is found in (a) texts devoted to the topic of SEM (e.g., Bollen, 1989a; Kaplan, 2000; Kline, 1998; Loehlin, 1992; Long, 1983b; Maruyama, 1998; Raykov & Marcoulides, 2000; and Schumacker & Lomax, 1996); (b) edited books devoted to the topic (e.g., Bollen & Long, 1993; Hoyle, 1995a; Marcoulides & Moustaki, 2002; and Marcoulides & Schumacker, 1996); and (c) methodologically oriented journals (e.g., *British Journal of Mathematical and Statistical Psychology*, *Journal of Educational and Behavioral Statistics*, *Multivariate Behavioral Research*, *Psychological Methods*, *Psychometrika*, *Sociological Methodology*, *Sociological Methods & Research*, and *Structural Equation Modeling: A Multidisciplinary Journal*).

THE GENERAL STRUCTURAL EQUATION MODEL

Symbol Notation

Structural equation models are schematically portrayed using particular configurations of four geometric symbols: a circle (or ellipse), a square (or rectangle), a single-headed arrow, and a double-headed arrow. By convention, circles (or ellipses, ⬭) represent unobserved latent factors; squares (or rectangles, ▭) represent observed variables; single-headed arrows (→) represent the impact of one variable on another; and double-headed arrows (↔) represent covariances or correlations between pairs of variables. In building a model of a particular structure under study, researchers use these symbols within the framework of four basic configurations, each of which represents an important component in the analytic process. These configurations are briefly described as follows:

⬭→▭ • Path coefficient for regression of an observed variable onto an unobserved latent variable (or factor).
⬭→⬭ • Path coefficient for regression of one factor onto another factor.
→▭ • Measurement error associated with an observed variable.
↘⬭ • Residual error in the prediction of an unobserved factor.

The Path Diagram

Schematic representations of models are termed *path diagrams* because they provide a visual portrayal of relations assumed to hold among the variables under study. Essentially, as discussed later, a path diagram depicting a particular SEM model is actually the graphical equivalent of its mathematical representation whereby a set of equations relates dependent variables to their explanatory variables. As a means of illustrating how the above four symbol configurations may represent a particular causal process let me now walk you through the simple model shown in Fig. 1.1.

In reviewing this model, we see that there are two *unobserved latent factors*—math self-concept (MSC) and math achievement (MATH)—and five *observed variables*—three are considered to measure MSC (SDQ5; SDQ11; SDQ18) and two to measure MATH (MATHGR; MATHACH). These five observed variables function as indicators of their respective underlying latent factors.

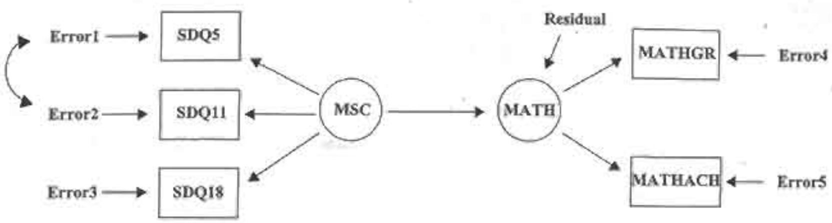

FIG. 1.1. A general structural equation model.

Associated with each observed variable is an error term (Error1–Error5) and, with the factor being predicted (MATH), a residual term (*Residual*). There is an important distinction between these two error terms. Error associated with observed variables represents *measurement error* which reflects on their adequacy in measuring the related underlying factors (MSC; MATH). Measurement error derives from two sources: *random measurement error* (in the psychometric sense) and *error uniqueness*, a term used to describe error variance arising from some characteristic considered specific (or unique) to a particular indicator variable. Such error often represents nonrandom measurement error. Residual terms represent error in the prediction of dependent factors from independent factors. For example, the residual term shown in Fig. 1.1 represents error in the prediction of MATH (the dependent factor) from MSC (the independent factor).

In essence, both measurement and residual error terms represent unobserved variables. Thus, it would seem perfectly reasonable that, consistent with the representation of factors, they too could be modeled as ellipse (or circle) enclosures. In fact, this is the modeling approach implemented in at least one SEM program: AMOS (Arbuckle, 2003).

In addition to symbols that represent variables, certain others are used in path diagrams to denote hypothesized processes involving the entire system of variables. In particular, one-way arrows represent structural regression coefficients, thus indicating the impact of one variable on another. In Fig. 1.1, for example, the unidirectional arrow pointing toward the dependent factor, MATH, implies that the independent factor MSC "causes" math achievement (MATH).[2] Likewise, the three unidirectional arrows leading from MSC to each of the three observed variables (SDQ5, SDQ11, SDQ18) and those leading from MATH to each of its indicators (MATHGR and MATHACH) suggest that these score values are each

[2]In this book, a "cause" is a direct effect of a variable on another within the context of a complete model. Its magnitude and direction are given by the partial regression coefficient. If the complete model contains all relevant influences on a given dependent variable, its causal precursors are correctly specified. In practice, however, models may omit key predictors and may be misspecified so that it may be inadequate as a "causal model" in the philosophical sense.

influenced by their respective underlying factors. As such, these path coefficients represent the magnitude of expected change in the observed variables for every change in the related latent variable (or factor). It is important to note that, typically, these observed variables represent subscale scores (see, e.g., chaps. 11 and 12), item scores (see, e.g., chaps. 4, 5, 7, 9, 10, and 13), item pairs (see, e.g., chap. 3), and/or carefully selected item bundles (see, e.g., chaps. 6 and 8).

The one-way arrows pointing from the five error terms indicate the impact of measurement error (random and unique) on the observed variables and, from the residual, the impact of error in the prediction of MATH. Finally, as noted previously, curved two-way arrows represent covariances or correlations between pairs of variables. Thus, the bidirectional arrow linking Error1 and Error2, as shown in Fig. 1.1, implies that measurement error associated with SDQ5 correlates with that associated with SDQ11.

Structural Equations

As noted at the beginning of this chapter, in addition to lending themselves to pictorial description via a schematic presentation of the causal processes under study, structural equation models can also be represented by a series of regression (i.e., structural) equations. Because (a) regression equations represent the influence of one or more variables on another, and (b) this influence, conventionally in SEM, is symbolized by a single-headed arrow pointing from the variable of influence to the variable of interest, we can think of each equation as summarizing the impact of all relevant variables in the model (observed and unobserved) on one specific variable (observed or unobserved). Thus, one relatively simple approach to formulating these equations is to note each variable that has one or more arrows pointing toward it and then record the summation of all such influences for each of these dependent variables.

To illustrate this translation of regression processes into structural equations, let's turn again to Fig. 1.1. We can see that there are six variables with arrows pointing toward them; five represent observed variables (SDQ5, SDQ11, SDQ18; MATHGR, MATHACH) and one represents an unobserved variable (or factor; MATH). Thus, we know that the regression functions symbolized in the model shown in Fig. 1.1 can be summarized in terms of six separate equation-like representations of linear dependencies, as follows:

MATH = MSC + Residual
SDQ5 = MSC + Error1
SDQ11 = MSC + Error2
SDQ18 = MSC + Error3
MATHGR = MATH + Error4
MATHACH = MATH + Error5

Nonvisible Components of a Model

Although in principle there is a one-to-one correspondence between the schematic presentation of a model and its translation into a set of structural equations, neither one of these model representations tells the whole story. Some parameters critical to the estimation of the model are not explicitly shown and thus may not be obvious to the novice structural-equation modeler. For example, in both the path diagram and the previous equations, there is no indication that the variances of the independent variables are parameters in the model; indeed, such parameters are essential to all structural equation models.

Likewise, it is equally important to notice the specified nonexistence of certain parameters in a model. For example, in Fig. 1.1, there is no curved arrow between Error4 and Error5, which suggests the lack of covariance between the error terms associated with the observed variables MATHGR and MATHACH. Similarly, there is no hypothesized covariance between MSC and the residual. Absence of this path addresses the common and most often necessary assumption that the predictor (or independent) variable is in no way associated with any error arising from the prediction of the criterion (or dependent) variable.

Basic Composition

The general SEM model can be decomposed into two submodels: a measurement model and a structural model. The measurement model defines relations between the observed and unobserved variables. In other words, it provides the link between scores on a measuring instrument (i.e., the observed indicator variables) and the underlying constructs they are designed to measure (i.e., the unobserved latent variables). The measurement model, then, represents the CFA model described earlier in that it specifies the pattern by which each measure loads on a particular factor. In contrast, the structural model defines relations among the unobserved variables. Accordingly, it specifies the manner by which particular latent variables directly or indirectly influence (i.e., "cause") changes in the values of certain other latent variables in the model.

For didactic purposes in clarifying this important aspect of SEM composition, let's now examine Fig. 1.2, in which the same model presented in Fig. 1.1 has been demarcated into measurement and structural components.

Considered separately, the elements modeled within each rectangle in Fig. 1.2 represent two CFA models. The enclosure of the two factors within the ellipse represents a full latent variable model and thus would not be of interest in CFA research. The CFA model to the left of the diagram represents a one-factor model (MSC) measured by three observed variables (SDQ5 – SDQ18), whereas the CFA model on the right represents a one-factor model (MATH) measured by two observed variables (MATHGR–MATHACH). In both cases, the regression of the observed variables on each factor and the variances of both the factor and the

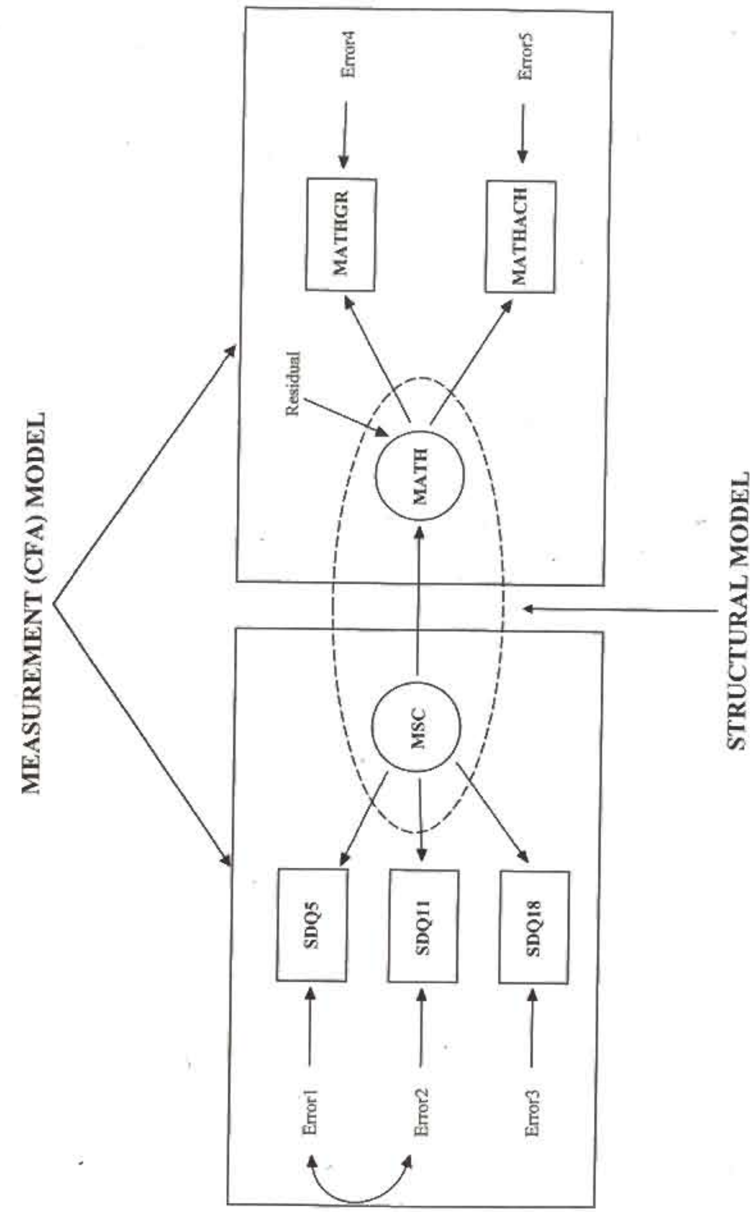

FIG. 1.2. A general structural equation model demarcated into measurement and structural components.

errors of measurement are of primary interest; the error covariance would be of interest only in analyses related to the CFA model bearing on MSC.

The Formulation of Covariance and Mean Structures

Covariance
der Koeffizienten?
Auspartialisieren orthogonal.

The core parameters in structural equation models that focus on the analysis of covariance structures are the regression coefficients and the variances and covariances of the independent variables. When the focus extends to the analysis of mean structures, the means and intercepts also become central parameters in the model. However, given that sample data comprise observed scores only, there needs to be an internal mechanism whereby the data are transposed into parameters of the model. This task is accomplished via a mathematical model representing the entire system of variables. Such representation systems can and do vary with each SEM computer program; the mechanism used by the EQS program is discussed in the next chapter.

As with any form of communication, one must first understand the language before being able to understand the message conveyed; so it is in comprehending the specification of SEM models. Now that you are familiar with the basic concepts underlying SEM, let's take a second look at the models just presented—albeit this time with the specifications recast within the framework and lexicon of the EQS program.

THE GENERAL EQS STRUCTURAL EQUATION MODEL

EQS Notation

EQS regards all variables as falling into one of two categories: measured (observed) variables or unmeasured (unobserved) variables. All measured variables are designated as V's and constitute the actual data of a study. All other variables are hypothetical and represent the structural network of the phenomenon under investigation.

Although conceptually unnecessary, it makes sense in practice to differentiate among the unmeasured variables: (a) the latent construct itself (regarded generally as a factor in EQS), designated as F; (b) a residual associated with the measurement of each observed variable (V), designated as E; and (c) a residual associated with the prediction of each factor, designated as D. Residual terms are often referred to as "disturbances," which is the terminology used in the EQS program. For consistency with EQS, the term *disturbance* is used in lieu of *residual* throughout the remainder of this book. Finally, for simplicity, all E's and D's are numbered to correspond with the V's and F's with which they are associated, respectively.

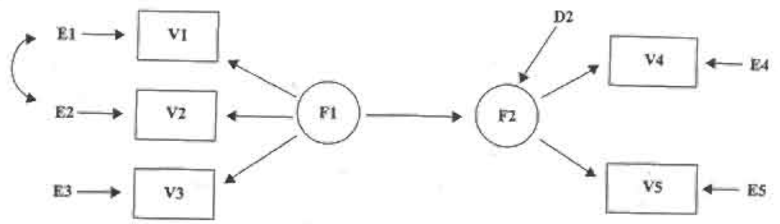

FIG. 1.3. A general EQS structural equation model.

The EQS Path Diagram

To comprehend more fully how this labeling system works, let's turn to Fig. 1.3, which you will readily note is a replication of Fig. 1.1 albeit recast as an EQS–specified model. Given the detailed description of this path diagram presented earlier, no further explanation is provided here. Nonetheless, comments bearing on the number accompanying each labeled variable in the model are noteworthy. First, EQS automatically numbers each observed variable in accordance with its data-entry placement. As such, the first variable in the data set would be designated V1, the second V2, and so on. Second, the numeric value associated with each error (E) and disturbance (D) term is consistent with its related observed (V) and unobserved (F) variables, respectively. Thus, although there is only one disturbance term in this model, it is labeled D2 rather than D1.

The Bentler–Weeks Representation System

As discussed earlier, given that sample data constitute observed scores only, every SEM program requires some means of transposing these scores into parameters of the model. This is accomplished via a mathematical model representing the entire system of variables. In EQS, the mathematical model derives from the work of Bentler and Weeks (1979, 1980). (For a comparative review of the representation systems for EQS and LISREL; see, e.g., Bentler, 1988.)

The thrust of the Bentler–Weeks model is that all variables in a model can be categorized as either *dependent* or *independent* variables. Any variable that has a unidirectional arrow aimed at it represents a dependent variable; if there is no unidirectional arrow aimed at it, a variable is considered independent. As is customary, dependent variables are explained in terms of other variables in the model, whereas independent variables serve as the explanatory variables. Not so customary, however, is the Bentler–Weeks conceptualization of that which constitutes a dependent or independent variable. Indeed, their interpretation of this concept is much broader than is typical. According to Bentler and Weeks, any variable that is not a dependent variable is automatically considered an independent variable, regardless of whether it is an observed score, an unobserved factor, or a disturbance term. For example, in Fig. 1.3, the dependent variables are V_1, V_2, V_3, V_4, V_5, and F_2; the independent variables are E_1, E_2, E_3, E_4, E_5, F_1, and D_2.

A dependent variable, then, is any variable that can be expressed as a structural regression function of other variables. Thus, for every dependent variable, this regression function can be summarized in the form of an equation. As with the equations based on the variables shown in Fig. 1.1, each regression function modeled in Fig. 1.3 can be translated into equations specific to EQS. As will become evident in the applications illustrated later, these equations essentially serve to define the model for the program. The related equations are as follows:

$$
\begin{aligned}
F_2 &= F_1 + D_2 \\
V_1 &= F_1 + E_1 \\
V_2 &= F_1 + E_2 \\
V_3 &= F_1 + E_3 \\
V_4 &= F_2 + E_4 \\
V_5 &= F_2 + E_5
\end{aligned}
$$

It is clear that the one-way arrows linking the factors to the observed variables and Factor 1 to Factor 2 represent regression coefficients. However, explanation regarding the linkage of disturbance terms to their associated variables via one-way arrows may be somewhat less obvious. Although these arrows also symbolize regression coefficients, their paths are implicit in the prediction of one variable from another; they are considered to be known and are therefore fixed to 1.0. For example, in the language of simple regression, the prediction of V_1 from F_1 can be written as $V_1 = b_{11}F_1 + E_1$, where b_{11} represents the unknown beta weight associated with the predictor F_1 and E_1 represents error in this prediction. Note that there is no beta weight associated with the error term, thereby indicating that it is not to be estimated. By implication then, the beta weight for E_1 is considered known and fixed arbitrarily to 1.0.[3] Similarly, the prediction of F_2 from F_1 can be written as $F_2 = b_{12}F_1 + D_2$, where D_2 represents error in the prediction—albeit this prediction involves the prediction of one factor from another—whereas the former prediction equation involved the prediction of an observed variable from a factor (hence, the distinction between the terms E and D).

Finally, an important corollary of the Bentler–Weeks model is that the variances of dependent variables or their covariances with other variables are never parameters of the model; rather, they remain to be explained by those parameters.

[3] If, on the other hand, it were of interest to estimate the regression path, this can certainly be done. However, in this case, the variance of E_1 would need to be fixed to 1.0 because both the regression coefficient and the variance cannot be estimated simultaneously. This fact is linked to the concept of model identification, which is addressed in chapter 2.

In contrast, the variances and covariances of independent variables are important parameters that need to be estimated.

In this chapter, I have presented you with a few of the basic concepts associated with SEM and with a general overview of model specification within the framework of EQS. In chapter 2, however, I provide you with substantially greater detail regarding the specification of models within the context of the manual, interactive, and graphical choices available in the EQS program. Throughout the chapter, I show you how to use the DIAGRAMMER feature in building models, review many of the drop-down menus, and detail specified and illustrated components of three basic SEM models. As you work your way through the applications included in this book, you will become increasingly more confident in both your understanding of SEM and in using the EQS program—and I know that you will absolutely love what the much-improved DIAGRAMMER feature can do! So, let's move on to chapter 2 and a more comprehensive look at SEM modeling with EQS 6.1.