Confirmatory Factor Analysis

Timothy A. Brown and Michael T. Moore

Boston University

Correspondence concerning this chapter should be addressed to Timothy A. Brown, Center for Anxiety & Related Disorders, Department of Psychology, Boston University, 648 Beacon Street, 6th floor, Boston, MA 02215-2013.

22.1 Introduction

Confirmatory factor analysis (CFA) is a type of structural equation modeling that deals specifically with measurement models; that is, the relationships between observed measures or *indicators* (e.g., test items, test scores, behavioral observation ratings) and latent variables or *factors*. The goal of latent variable measurement models (i.e., factor analysis) is to establish the number and nature of factors that account for the variation and covariation among a set of indicators. A factor is an unobservable variable that influences more than one observed measure and which accounts for the correlations among these observed measures. In other words, the observed measures are intercorrelated because they share a common cause (i.e., they are influenced by the same underlying construct); if the latent construct was partialled out, the intercorrelations among the observed measures would be zero. Thus, a measurement model such as CFA provides a more parsimonious understanding of the covariation among a set of indicators because the number of factors is less than the number of measured variables.

These concepts originate from the *common factor model* (Thurstone, 1947) which states that each indicator in a set of observed measures is a linear function of one or more common factors and one unique factor. Factor analysis partitions the variance of each indicator (derived from the sample correlation or covariance matrix) into two parts: (1) *common variance*, or the variance accounted for by the latent variable(s), which is estimated on the basis of variance shared with other indicators in the analysis; and (2) *unique variance*, which is a combination of reliable variance specific to the indicator (i.e., systematic latent variables that influence only one indicator) and random error variance (i.e., measurement error or unreliability in the indicator). There are two main types of analyses based on the common factor model: *exploratory factor analysis* (EFA) and *CFA* (Jöreskog, 1969, 1971). EFA and CFA both aim to reproduce the observed relationships among a group of indicators with a smaller set of latent variables. However, EFA and CFA differ fundamentally by the number and nature of *a priori* specifications and restrictions made on the latent variable measurement model. EFA is a data-driven approach such that no specifications are made in regard to the number of common factors (initially) or the pattern of relationships between the common factors and the indicators (i.e., the *factor loadings*). Rather, the researcher employs EFA as an exploratory or descriptive data technique to determine the appropriate number of common factors, and to ascertain which measured variables are reasonable indicators of the various latent dimensions (e.g., by the size and differential magnitude of the factor loadings). In CFA, the researcher specifies the number of factors and the pattern of indicator-factor loadings in advance as well as other parameters such as those bearing on the independence or covariance of the factors and indicator unique variances. The pre-specified factor solution is evaluated in terms of how well it reproduces the sample covariance matrix of the measured variables. Unlike EFA, CFA requires a strong empirical or conceptual foundation to guide the specification and evaluation of the factor model. Accordingly, EFA is often used early in the process of scale development and construct validation, whereas CFA is used in the later phases when the underlying structure has been established on prior empirical and theoretical grounds. Other differences between EFA and CFA are discussed throughout this chapter (for a more detailed discussion, see Brown, 2006).

22.2 Purposes of CFA

CFA can be used for a variety of purposes, such as psychometric evaluation, the detection of method effects, construct validation, and the evaluation of measurement invariance. Nowadays, CFA is almost always used in the process of scale development to examine the latent structure of a test instrument. CFA verifies the number of underlying dimensions of the instrument (factors) and the pattern of item-factor relationships (factor loadings). CFA also assists in the determination of how a test should be scored. For instance, when the latent structure is multifactorial (i.e., two or more factors), the pattern of factor loadings supported by CFA will designate how a test might be scored using subscales; i.e., the number of factors are indicative of the number of subscales, the pattern of item-factor relationships (which items load on which factors) indicate how the subscales should be scored. CFA is an important analytic tool for other aspects of psychometric evaluation such as the estimation of scale reliability (e.g., Raykov, 2001).

Unlike EFA, the nature of relationships among the indicator unique variances can be modeled in CFA. Because of the nature of the identification restrictions in EFA, factor models must be specified under the assumption that measurement error is random. In contrast, correlated measurement error can be modeled in a CFA solution provided that this specification is substantively justified and that other identification requirements are met. When measurement error is specified to be random (i.e., the indicator unique variances are uncorrelated), the assumption is that the observed relationship between any two indicators loading on the same factor is due entirely to the shared influence of the latent variable; i.e., if the factor was partialled out, the correlation of the indicators would be zero. The specification of correlated indicator uniquenesses assumes that, whereas indicators are related in part because of the shared influence of the latent variable, some of their covariation is due to sources other than the common factor. In CFA, the specification of correlated errors may be justified on the basis of *method effects* that reflect additional indicator covariation that resulted from common assessment methods (e.g., observer ratings, questionnaires), reversed or similarly worded test items, or differential susceptibility to other influences such as response set, demand characteristics, acquiescence, reading difficulty, or social desirability. The inability to specify correlated errors is a significant limitation of EFA because the source of covariation among indicators that is not due to the substantive latent variables may be manifested in the EFA solution as additional factors (e.g., “methods” factors stemming from the assessment of a unidimensional trait with a questionnaire comprised of both positively and negatively worded items; cf. Brown, 2003; Marsh, 1996).

CFA is an indispensable analytic tool for construct validation. The results of CFA can provide compelling evidence of the *convergent* and *discriminant validity* of theoretical constructs. Convergent validity is indicated by evidence that different indicators of theoretically similar or overlapping constructs are strongly interrelated; e.g., symptoms purported to be manifestations of a single mental disorder load on the same factor. Discriminant validity is indicated by results showing that indicators of theoretically distinct constructs are not highly intercorrelated; e.g., psychiatric symptoms thought to be features of different types of disorders load on separate factors, and the factors are not so highly correlated as to indicate that a broader construct has been erroneously separated into two or more factors. One of the most elegant uses of CFA in construct validation is the analysis of multitrait-multimethod matrices (cf. Campbell & Fiske, 1959; Kenny & Kashy, 1992; aided by the fact that indicator error variances can be estimated in CFA to model method effects). A fundamental strength of CFA approaches to construct validation is the resulting estimates of convergent and discriminant validity are adjusted for measurement error and an error theory. Thus, CFA is superior to traditional analytic methods that do not account for measurement error (e.g., ordinary least squares approaches such as correlation/multiple regression assume that variables in the analysis are free of measurement error).

In addition, CFA offers a very strong analytic framework for evaluating the equivalence of measurement models across distinct groups (e.g., demographic groups such as sexes, races, or cultures). This is accomplished by either multiple-group solutions (i.e., simultaneous CFAs in two or more groups) or MIMIC models (i.e., the factors and indicators are regressed onto observed covariates representing group membership). These capabilities permit a variety of important analytic opportunities in applied research, such as the evaluation of whether a scale's measurement properties are invariant across population subgroups; e.g., are the factors, factor loadings, item intercepts, etc., that define the latent structure of a questionnaire equivalent in males and females? Indeed, *measurement invariance* is an important aspect of scale development, as this endeavor determines whether a testing instrument is appropriate for use in various groups (e.g., does the score issued by the test instrument reflect the same level of the underlying characteristic or ability in males and females?). Another chapter in this book is devoted to this topic (Millsap & Olivera-Aguilar, in press).

CFA should be employed as a precursor to structural equation models (SEM) that specify structural relationships (e.g., regressions) among the latent variables. SEM models consist of two major components: (1) the *measurement model*, which specifies the number of factors, how the various indicators are related to the factors, and the relationships among indicator errors (i.e., a CFA model); and (2) the *structural model*, which specifies how the various factors are related to one another (e.g., direct or indirect effects, no relationship). When poor model fit is encountered in SEM studies, it is more likely that this will be due to misspecifications in the measurement portion of the model than from the structural component. This is because there are usually more things that can go wrong in the measurement model than in the structural model (e.g., problems in the selection of observed measures, misspecified factor loadings, additional sources of covariation among observed measures that cannot be accounted for by the specified factors). Thus, although CFA is not the central analysis in SEM studies, an acceptable measurement model should be established before estimating and interpreting the structural relationships among latent variables.

22.3 CFA Model Parameters

All CFA models contain the parameters of factor loadings, unique variances, and factor variances. Factor loadings are the regression slopes for predicting the indicators from the factor. Unique variance is variance in the indicator that is not accounted for by the factors. Unique variance is typically presumed to be measurement error and is thus usually referred to as such (other synonymous terms include “error variance” and “indicator unreliability”). A factor variance expresses the sample variability or dispersion of the factor; that is, the extent to which sample participants’ relative standing on the latent dimension is similar or different. If substantively justified, a CFA can include error covariances (also referred to as “correlated uniquenesses,” “correlated residuals,” or “correlated errors”) which designate that two indicators covary for reasons other than the shared influence of the latent factor (e.g., method effects). When the CFA solution has two or more factors, a factor covariance is almost always specified to estimate the relationship between the latent dimensions.

The latent variables may be either exogenous or endogenous. An *exogenous variable* is a variable that is not caused by other variables in the solution. Conversely, an *endogenous variable* is caused by one or more variables in the model (i.e., other variables in the solution exert direct effects on the variable). Thus, exogenous variables can be viewed as synonymous to

X, independent, or predictor (causal) variables. Similarly, endogenous variables are equivalent to Y, dependent, or criterion (outcome) variables. CFAs are typically considered to be exogenous (latent X) variable solutions because the latent variables are specified to be freely intercorrelated without directional relationships among them. However, when the CFA analysis includes covariates (i.e., predictors of the factors or indicators as in MIMIC models; see Brown, 2006) or higher-order factors (e.g., superordinate dimensions specified to account for the correlations among the CFA factors), the factors are endogenous (latent Y).

A data-based example is now introduced to illustrate some of these concepts. In this example, a 10-item questionnaire of the symptoms of obsessive-compulsive disorder (OCD) has been administered to a sample of 400 outpatients with anxiety and mood disorders. Participants responded to each item on a 0-8 scale. A two-factor model is anticipated; that is, the first five items (X1-X5) are conceptualized as indicators of the latent construct of Obsessions, and the remaining five items (X6-X10) are conjectured to be features of the underlying dimension of Compulsions. A method effect (i.e., error covariance) was anticipated for items X9 and X10 because these items were reverse-worded (i.e., unlike the other eight items, these items were phrased in the nonsymptomatic direction). The posited CFA measurement model is presented in Figure 1. Per conventional path diagram notation, the latent variables are depicted by circles and indicators by squares or rectangles. Figure 1 also provides the Greek symbols used to notate the various parameters in this CFA model. Factor loadings are symbolized by lambdas (λ), which are subscripted by two numbers to denote the order of the factors and indicators, respectively (e.g., $\lambda_{2,1}$ = the second indicator, X2, loads on the first factor, Obsessions). The unidirectional arrows ($\rightarrow$) from the factors to the indicators depict direct effects (regressions) of the latent dimensions onto the observed measures; the specific regression coefficients are the lambdas (λ). Thetas (Θ) represent matrices of indicator error variances and covariances; theta delta (Θ_δ) in the case of indicators of latent X variables, theta epsilon (Θ_ε) for indicators of latent Y variables. For notational ease, the symbols δ and ε are often used in place of θ_δ and θ_ε, respectively, in reference to elements of Θ_δ and Θ_ε (as is done in Figure 1). Although unidirectional arrows connect the thetas to the observed measures, these arrows are not regressive paths; i.e., Θ_δ and Θ_ε are symmetric variance-covariance matrices consisting of error variances on the diagonal, and error covariances, if any, in the off-diagonal. In fact, some notational systems do not use directional arrows in the depiction of error variances to avoid this potential source of confusion (one notational variation is to symbolize error variances with ovals, because like latent variables, measurement errors are not observed). Factor variances and covariances are notated by phi (ϕ) and psi (ψ) in latent X and latent Y models, respectively. Curved, bidirectional arrows are used to symbolize covariances (correlations) among factors (e.g., $\phi_{2,1}$ reflects the covariance between Obsessions and Compulsions).

In practice, CFA is often confined to the analysis of variance-covariance structures. In this case, the aforementioned parameters (factor loadings, error variances and covariances, factor variances and covariances) are estimated to reproduce the input variance-covariance matrix. The analysis of covariance structures is based on the assumption that indicators are measured as deviations from their means (i.e., all indicator means equal zero). However, the CFA model can be expanded to include the analysis of mean structures, in which case the solution also strives to reproduce the observed sample means of the indicators (which are included along with the sample variances and covariances as input data). The analysis of mean structures is the topic of the chapter by Green and Thompson (in press) in this volume.

22.4 CFA Model Specification

Types of parameters. There are three types of parameters that can be specified in a CFA model: free, fixed, or constrained. A *free parameter* is unknown, and the researcher allows the analysis to find its optimal value that, in tandem with other model estimates, minimizes the differences between the observed and predicted variance-covariance matrices. A *fixed parameter* is pre-specified by the researcher to be a specific value, most commonly either 1.0 (e.g., in the case of marker indicators or factor variances to define the metric of a latent variable; see next section) or 0 (e.g., the absence of indicator cross-loadings or error covariances). A third type of estimate is a *constrained parameter*. As with a free parameter, a constrained parameter is unknown. However, the parameter is not free to be any value, but rather the specification places restrictions on the values it may assume. The most common type of constrained parameter is an *equality constraint*, in which some of the parameters in the CFA solutions are restricted to be equal in value. For instance, equality constraints are used in the evaluation of measurement invariance to determine whether the measurement parameters of the CFA model (e.g., factor loadings) are equivalent in subgroups of the sample (see Millsap & Olivera-Aguilar, in press).

The output of CFA can render parameter estimates in three different metrics: completely standardized, unstandardized, and partially standardized. In the case of a *completely standardized* estimate, both the latent variable and indicator are in standardized metrics (i.e., $M = 0$, $SD = 1$). For instance, if an indicator is specified to load on only one factor (as is true for each item in the Figure 1 measurement model), the completely standardized factor loading can be interpreted as the correlation between the indicator and the factor (although strictly speaking, this estimate is a standardized regressive path reflecting the degree of standardized score change in the indicator given a standardized unit increase in the factor). In many popular software programs (e.g., SPSS), EFA is exclusively a completely standardized analysis (i.e., a correlation matrix is used as input and all results are provided in the completely standardized metric). However, unlike EFA, the results of CFA also include an *unstandardized solution* (parameter estimates expressed in the original metrics of the latent variables and indicators), and possibly a *partially standardized solution* (relationships where either the indicators or latent variable is standardized and the other is unstandardized). Although completely standardized and unstandardized estimates are usually the primary focus in applied CFA research, partially standardized estimates can also be informative in some contexts. For example, when a latent variable is regressed onto a dummy code (e.g., a binary variable indicating whether the participant is male or female), the resulting partially standardized path is more substantively meaningful than a completely standardized estimate (e.g., when the latent variable is standardized and the dummy code is unstandardized, the partially standardized path estimate reflects the extent to which males and females differ on the latent variable in *SD* units; cf.

Cohen's *d* index of effect size; Cohen, 1988).

Model identification. To estimate a CFA solution, the measurement model must be *identified*. A model is identified if, on the basis of known information (i.e., the variances and covariances in the sample input matrix), a unique set of estimates for each parameter in the model can be obtained (e.g., factor loadings, factor covariances, etc.). The two primary aspects of CFA model identification are scaling the latent variables and statistical identification.

Latent variables have no inherent metrics and thus their units of measurement must be defined by the researcher. In CFA, this is accomplished in one of two ways. The most widely used method is the *marker indicator* approach whereby the unstandardized factor loading of one observed measure per factor is fixed to a value of 1.0. As will be illustrated shortly, this specification serves the function of passing the metric of the marker indicator along to the latent variable. In the second method, the variance of the latent variable is fixed to a value of 1.0. Although most CFA results are identical to the marker indicator approach when the factor variance is fixed to 1.0, this method does not produce an unstandardized solution. The absence of an unstandardized solution often contraindicates the use of this approach (e.g., when the researcher is interested in evaluating the measurement invariance of a test instrument; see Millsap & Olivera-Aguilar, in press).

Statistical identification refers to the concept that a CFA solution can be estimated only if the number of freely estimated parameters (e.g., factor loadings, uniquenesses, factor correlations) does not exceed the number of pieces of information in the input matrix (e.g., number of sample variances and covariances). A model is *over-identified* when the number of knowns (i.e., individual elements of the input matrix) exceeds the number of unknowns (i.e., the freely estimated parameters of the CFA solution). The difference in the number of knowns and the number of unknowns constitutes the model's *degrees of freedom* (*df*). Over-identified solutions have positive *df*. For over-identified models, goodness of fit evaluation can be implemented to determine how well the CFA solution was able to reproduce the relationships among indicators observed in the sample data. If the number of knowns equals the number of unknowns, the model has zero *df* and is said to be *just-identified*. Although just-identified models can be estimated, goodness of fit evaluation does not apply because these solutions perfectly reproduce the input variance-covariance matrix. When the number of freely estimated parameters exceeds the number of pieces of information in the input matrix (e.g., when too many factors are specified for the number of indicators in the sample data), *df* are negative and the model is *under-identified*. Under-identified models cannot be estimated because the solution cannot arrive at a unique set of parameter estimates.

In some cases, the researcher may encounter an *empirically under-identified* solution. In these solutions the measurement model is statistically just- or over-identified, but there are aspects of the input data or the model specification that prevent the analysis from arriving at a valid set of parameter estimates (i.e., the estimation will not reach a final solution, or the final solution will include one or more parameter estimates that have out-of-range values such as a negative indicator error variance). Although the various causes and remedies for empirically under-identified solutions is beyond the scope of this chapter (see Brown, 2006, and Wothke, 1993, for further discussion), a basic example would be the situation where the observed measure selected to be the marker indicator is in fact uncorrelated with all other indicators of the latent variable (thus, the metric of the latent variable would be unidentified).

Example. The sample data for the measurement model of the OCD questionnaire are provided in Table 1. Specifically, Table 1 presents the sample standard deviations (*SD*) and correlations (*r*) for the 10 questionnaire items. These data will be read into the latent variable software program and converted into variances and covariances which will be used by the analysis as the input matrix (e.g., $\mathrm{VAR}_{X1} = SD_{X1}^2$; $\mathrm{COV}_{X1,X2} = r_{X1,X2}SD_{X1}SD_{X2}$). Generally, it is preferable to use a raw data file as input for CFA (e.g., to avoid rounding error and to adjust for missing or non-normal data, if needed). In this example, however, the sample *SD*s and *r*s are presented to foster the illustration of concepts covered later in this chapter. Also, the data in Table 1 can be readily used as input if the reader is interested in replicating the analyses presented in this chapter.

The measurement model in Figure 1 is over-identified with *df* =33. The model *df* indicates that there are 33 more elements in the input matrix than there are freely estimated parameters in the two-factor CFA model. Specifically, there are 55 variances and covariances in the input matrix (cf. Table 1) and 22 freely estimated parameters in the CFA model; i.e., 8 factor loadings (the factor loadings of X1 and X6 are not included in the tally because they are fixed to 1.0 to serve as marker indicators), 2 factor variances, 1 factor covariance, 10 error variances, and 1 error covariance (see Figure 1). With the exception of X9 and X10, all error covariances are fixed to zero (no curved, double-headed arrows connecting the unique variances of items X1 through X8) which assumes measurement error in these indicators is random.

A noteworthy aspect of the model specification depicted in Figure 1 is that each indicator loads on only one factor—this parameterization does not include *cross-loadings* where an indicator is predicted by more than one factor (i.e., all cross-loadings are fixed to zero). This is another key difference between EFA and CFA. In traditional EFA, the factor loading matrix is said to be *saturated* because all possible relationships (factor loadings) between the indicators and factors are freely estimated. Thus, for EFA models with two or more factors, a mathematical transformation referred to as *rotation* is conducted to foster the interpretability of the solution by by maximizing (primary) factor loadings close to 1.0 and minimizing cross-loadings close to 0.0. Rotation does not apply to CFA because most or all indicator cross-loadings are typically fixed to zero. Consequently, CFA models are usually more parsimonious than EFA solutions because while primary loadings and factor correlations are freely estimated, no other relationships are specified between the indicators and factors.[1]

22.5 CFA Model Estimation

The objective of CFA is to obtain estimates for each parameter of the measurement model (i.e., factor loadings, factor variances and covariances, indicator error variances and possibly error covariances) that produce a predicted variance-covariance matrix (also referred to as the model-implied variance-covariance matrix) that resembles the sample variance-covariance matrix as closely as possible. For instance, in over-identified models (such as the model in

Figure 1), perfect fit is rarely achieved. Thus, the goal of the analysis is to find a set of parameter estimates (e.g., factor loadings, factor correlations) that yield a predicted variance-covariance matrix that best reproduces the input variance-covariance matrix.

In the example shown in Figure 1, the following three equations provide the model-implied covariances of the 10 indicators in this measurement model. Given the absence of additional restrictions in this model (e.g., equality constraints on the factor loadings), the variances of the indicators are just-identified (i.e., guaranteed to be perfectly reproduced by the CFA solution; see "CFA Model Evaluation" section). For two indicators that load on the same factor (and do not load on any other factors), the model-implied covariance is the product of their factor loadings and the factor variance. For instance, the predicted covariance of X2 and X3 is:

$$\mathrm{COV}(X2, X3) = \lambda_{2,1}\phi_{1,1}\lambda_{3,1} \qquad \text{(Eq. 22.1)}$$

If the two indicators load on different factors (but do not cross-load on other factors), the model-implied covariance is the product of their factor loadings and the factor covariance; e.g., for X2 and X7:

$$\mathrm{COV}(X2, X7) = \lambda_{2,1}\phi_{2,1}\lambda_{7,2} \qquad \text{(Eq. 22.2)}$$

It is worth mentioning here that for a given indicator set and factor model (e.g., a two-factor solution), factor correlation estimates are usually larger in CFA than in EFA with oblique rotation (unlike orthogonal rotations, oblique rotations in EFA allow the factors to be intercorrelated). This stems from how the factor loading matrix is parameterized in CFA and EFA. Unlike EFA where the factor loading matrix is saturated, in CFA most if not all cross-loadings are fixed to zero. Thus, the model-implied correlation of indicators loading on separate factors in CFA is reproduced solely by the primary loadings and the factor correlation (cf. Eq. 22.2). Compared to oblique EFA (where the model-implied correlation of indicators with primary loadings on separate factors can be estimated in part by the indicator cross-loadings), in CFA there is more burden on the factor correlation to reproduce the correlation between indicators specified to load on different factors because there are no cross-loadings to assist in this model-implied estimate (i.e., in the mathematical process to arrive at CFA parameter estimates that best reproduce the sample matrix, the magnitude of the factor correlation estimate may be increased somewhat to better account for the relationships of indicators that load on separate factors).

Finally, if the model specification includes an indicator error covariance, this estimate must be added to Eqs. 22.1 or 22.2 to yield the model-implied covariance for the indicators. For the X9 and X10 indicators in Figure 1, the equation would be:

$$\text{COV}(X9, X10) = \lambda_{9,2}\phi_{2,2}\lambda_{10,2} + \delta_{10,9} \qquad \text{(Eq. 22.3)}$$

The equations will be illustrated in subsequent sections of this chapter using the parameter estimates from the OCD questionnaire example.

The estimation process in CFA (and SEM, in general) entails a *fitting function*, a mathematical operation to minimize the difference between the sample and model-implied variance-covariance matrices. By far, the fitting function most widely used in applied CFA and SEM research is *maximum likelihood* (ML). ML is the default statistical estimator in most latent variable software programs. The underlying principle of ML estimation is to find the model parameter estimates that maximize the probability of observing the available data if the data were collected from the same population again. In other words, ML aims to find the parameter values that make the observed data most likely (or conversely, maximize the likelihood of the parameters given the data). Finding the parameter estimates for an over-identified CFA model is an *iterative* procedure. That is, the computer program (such as Mplus, LISREL, EQS, or Amos) begins with an initial set of parameter estimates (referred to as *starting values* or *initial estimates*, which can be automatically generated by the software or specified by the user), and repeatedly refines these estimates in effort to minimize the difference between the sample and model-implied variance-covariance matrices. Each refinement of the parameter estimates is an *iteration*. The program conducts internal checks to evaluate its progress in obtaining the best set of parameter estimates. *Convergence* is reached when the program arrives at a set of parameter estimates that cannot be improved upon to further reduce the difference between the input and predicted matrices.

It is important to note that ML is only one of several methods that can be used to estimate CFA models. ML has several requirements that render it an unsuitable estimator in some circumstances. Some key assumptions of ML are: (1) the sample size is large (asymptotic); (2) the indicators of the factors have been measured on continuous scales (i.e., approximate interval-level data); and (3) the distribution of the indicators is multivariate normal. Although the actual parameter estimates (e.g., factor loadings) may not be affected, non-normality in ML analysis can result in biased standard errors (and hence faulty significance tests) and goodness of fit statistics. If non-normality is extreme (e.g., marked floor effects, as would occur if the majority of the sample responded to items using the lowest response choice—e.g., 0 on the 0-8 scale of the OCD questionnaire if the symptoms were infrequently endorsed by participants), then ML will produce incorrect parameter estimates (i.e., the assumption of a linear model is invalid). Thus, in the case of non-normal, continuous indicators, it is better to use a different estimator, such as ML with robust standard errors and χ^2 (e.g., Bentler, 1995). These robust estimators provide the same parameter estimates as ML, but both the goodness of fit statistics (e.g., χ^2) and standard errors of the parameter estimates are corrected for non-normality in large samples. If one or more of the factor indicators is categorical (or non-normality is extreme), normal theory ML should not be used. In this instance, estimators such as a form of weighted least squares (e.g., WLSMV; Muthén, du Toit, & Spisic, 1997), and unweighted least squares (ULS) are more appropriate. Weighted least squares estimators can also be used for non-normal, continuous data, although robust ML is often preferred given its ability to outperform weighted least squares in small and medium-sized samples (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992). For more details, the reader is referred to the chapter by Lei and Wu (in press) in this book.

22.5 CFA Model Evaluation

Using the information in Table 1 as input, the two-factor model presented in Figure 1was fit to the data using the Mplus software program (version 6.0; Muthén & Muthén, 1998-2010). The Mplus syntax and selected output are presented in Table 2. As shown in the Mplus syntax, both the indicator correlation matrix and standard deviations (TYPE = STD CORRELATION;)

were input because this CFA analyzed a variance-covariance matrix. The CFA model specification occurs under the "MODEL:" portion of the Mplus syntax. For instance, the line "OBS BY X1-X5" specifies that a latent variable to be named "OBS" (Obsessions) is measured by indicators X1 through X5. The Mplus programming language contains several defaults that are commonly implemented aspects of model specification (but nonetheless can be easily overridden by additional programming). For instance, Mplus automatically sets the first indicator after the "BY" keyword as the marker indicator (e.g., X1) and freely estimates the factor loadings for the remaining indicators in the list (X2 through X5). By default, all error variances (uniquenesses) are freely estimated and all error covariances and indicator cross-loadings are fixed to zero; the factor variances and covariances are also freely estimated by default. These and other convenience features in Mplus are very appealing to the experienced CFA researcher. However, novice users should become fully aware of these system defaults to ensure their models are specified as intended. In addition to specifying that X5 through X10 are indicators of the second factor (COM; Compulsions), the Mplus MODEL: syntax includes the line "X9 WITH X10." In Mplus, WITH is the keyword for "correlated with"; in this case, this allows the X9 and X10 uniquenesses to covary (based on the expectation that a method effect exists between these items because these are the only two items of the OCD questionnaire that are reverse-worded). Thus, this statement overrides the Mplus default of fixing error covariances to zero.

There are three major aspects of the results that should be examined to evaluate the acceptability of the CFA model. They are: (1) overall goodness of fit; (2) the presence or absence of localized areas of strain in the solution (i.e., specific points of ill-fit); and (3) the interpretability, size, and statistical significance of the model's parameter estimates. As discussed earlier, goodness of fit pertains to how well the parameter estimates of the CFA solution (i.e., factor loadings, factor correlations, error covariances) are able to reproduce the relationships that were observed in the sample data. For example, as seen in Table 2 (under the heading, "STDYX Standardization"), the completely standardized factor loadings for X1 and X2 are .760 and .688, respectively. Using Eq. 22.1, the model-implied correlation of these indicators is the product of their factor loading estimates; i.e., .760(1)(.688) = .523 (the factor variance = 1 in the completely standardized solution). Goodness of fit addresses the extent to which these model-implied relationships are equivalent to the relationships seen in the sample data (e.g., as shown in Table 1, the sample correlation of X1 and X2 was .516, so the model-implied estimate differed by only .007 standardized units).

There are a variety of goodness of fit statistics that provide a global descriptive summary of the ability of the model to reproduce the input covariance matrix. The classic goodness of fit index is χ^2. In this example, the model $\chi^2 = 46.16$, $df = 33$, $p = .06$. The critical value of the χ^2 distribution ($\alpha = .05$, $df = 33$) is 47.40. Because the model χ^2 (46.16) does not exceed this critical value (computer programs provide the exact probability value, e.g., $p = .06$) the null hypothesis that the sample and model-implied variance-covariance matrices do not differ is retained. On the other hand, a statistically significant χ^2 would lead to rejection of the null hypothesis, meaning that the model estimates do not sufficiently reproduce the sample variances and covariances (i.e., the model does not fit the data well).

Although χ^2 is steeped in the traditions of ML and SEM, it is rarely used in applied research as a sole index of model fit. There are salient drawbacks of this statistic including the fact that it is highly sensitive to sample size (i.e., solutions involving large samples would be routinely rejected on the basis of χ^2 even when differences between the sample and model-implied matrices are negligible). Nevertheless, χ^2 is used for other purposes such as nested model comparisons (discussed later in this chapter) and the calculation of other goodness of fit indices. While χ^2 is routinely reported in CFA research, other fit indices are usually relied on more heavily in the evaluation of model fit.

Indeed, in addition to χ^2, the most widely accepted global goodness of fit indices are the standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), Tucker-Lewis index (TLI), and the comparative fit index (CFI). In practice, it is suggested that each of these fit indices be reported and considered because they provide different information about model fit (i.e., absolute fit, fit adjusting for model parsimony, fit relative to a null model; see Brown, 2006, for further details). Considered together, these indices provide a more conservative and reliable evaluation of the fit of the model. In one of the more comprehensive and widely cited evaluations of cutoff criteria, the findings of simulation studies conducted by Hu and Bentler (1999) suggest the following guidelines for acceptable model fit: (a) SRMR values are close to .08 or below; (b) RMSEA values are close to .06 or below; and (c) CFI and TLI values are close to .95 or greater. For the two-factor solution in this example, each of these guidelines was consistent with acceptable overall fit; SRMR = .035, RMSEA = 0.032, TLI = 0.99, CFI = .99 (provided by Mplus but not shown in Table 2). However, it should be noted that this topic is debated by methodologists. For instance, some researchers assert that these guidelines are far too conservative for many types of models (e.g., measurement models comprised of many indicators and several factors where the majority of cross-loadings and error covariances are fixed to zero; cf. Marsh, Hau, & Wen, 2004).

The second aspect of model evaluation is to determine whether there are specific areas of ill-fit in the solution. A limitation of goodness of fit statistics (e.g., SRMR, RMSEA, CFI) is that they provide a *global*, descriptive indication of the ability of the model to reproduce the observed relationships among the indicators in the input matrix. However, in some instances, overall goodness of fit indices suggest acceptable fit despite the fact that some relationships among indicators in the sample data have not been reproduced adequately (or alternatively, some model-implied relationships may markedly exceed the associations seen in the data). This outcome is more apt to occur in complex models (e.g., models that entail an input matrix consisting of a large set of indicators) where the sample matrix is reproduced reasonably well on the whole, and the presence of a few poorly reproduced relationships have less impact on the global summary of model fit. On the other hand, overall goodness of fit indices may indicate a model poorly reproduced the sample matrix. However, these indices do not provide information on the reasons why the model fit the data poorly (e.g., misspecification of indicator-factor relationships, failure to model salient error covariances).

Two statistics that are frequently used to identify specific areas of misfit in a CFA solution are *standardized residuals* and *modification indices*. A residual reflects the difference between the observed sample value and model-implied estimate for each indicator variance and covariance (e.g., the deviation between the sample covariance and the model-implied covariance of indicators X1 and X2). When these residuals are standardized, they are analogous to standard scores in a sampling distribution and can be interpreted like *z* scores. Stated another way, these values can be conceptually considered as the number of standard deviations that the residuals differ from the zero-value residuals that would be associated with a perfectly fitting model. For instance, a standardized residual at a value of 1.96 or higher would indicate that there exists significant additional covariance between a pair of indicators that was not reproduced by the model's parameter estimates. Modification indices can be computed for each fixed parameter (e.g., parameters that are fixed to zero such as indicator cross-loadings and error covariances) and each constrained parameter in the model (e.g., parameter estimates that are constrained to be same the value). The modification index reflects an approximation of how much the overall model χ^2 will decrease if the fixed or constrained parameter is freely estimated. Because the modification index can be conceptualized as a χ^2 statistic with 1 *df*, indices of 3.84 or greater (i.e., the critical value of χ^2 at $p < .05$, $df = 1$) suggest that the overall fit of the model could be significantly improved if the fixed or constrained parameter was freely estimated. For instance, when the two-factor model is specified without the X9-X10 error covariance, the model $\chi^2(34) = 109.50$, $p < .001$, and the modification index for this parameter is 65.20 (not shown in Table 2). This suggests that the model χ^2 is expected to decrease by roughly 65.20 units if the error covariance of these two indicators is freely estimated. As can be seen, this is an approximation because the model χ^2 actually decreased 63.34 units (109.50 – 46.16) when this error covariance is included. Because modification indices are also sensitive to sample size, software programs provide expected parameter change (EPC) values for each modification index. As the name implies, EPC values are an estimate of how much the parameter is expected to change in a positive or negative direction if it were freely estimated in a subsequent analysis. In the current example, the completely standardized EPC for the X9-X10 correlated error was .480. Like the modification index, this is an approximation (as seen in Table 2, the estimate for the correlated error of X9 and X10 was .420). Although standardized residuals and modification indices provide specific information for how the fit of the model can be improved, such revisions should only be pursued if they can be justified on empirical or conceptual grounds (e.g., MacCallum, Roznowski, & Necowitz, 1992). Atheoretical specification searches (i.e., revising the model solely on the basis of large standardized residuals or modification indices) will often result in further model misspecification and over-fitting (e.g., inclusion of unnecessary parameter estimates due to chance associations in the sample data).

The final major aspect of CFA model evaluation pertains to the interpretability, strength, and statistical significance of the parameter estimates. The parameter estimates (e.g., factor loadings and factor correlations) should only be interpreted in context of a good-fitting solution. If the model did not provide a good fit to the data, the parameter estimates are likely biased (incorrect). For example, without the error covariance in the model, the factor loading estimates for X9 and X10 are considerably larger than the factor loadings shown in Table 2 because the solution must strive to reproduce the observed relationship between these indicators solely through the factor loadings. In context of a good-fitting model, the parameter estimates should first be evaluated to ensure they make statistical and substantive sense. The parameter estimates should not take on out-of-range values (often referred to as *Heywood cases*) such as a negative indicator error variance. These results may be indicative of model specification error or problems with the sample or model-implied matrices (e.g., a non-positive definite matrix, small *N*). Thus, the model and sample data must be viewed with caution to rule out more serious causes of these outcomes (again, see Wothke, 1993, and Brown, 2006, for further discussion). From a substantive standpoint, the parameters should be of a magnitude and direction that is in accord with conceptual or empirical reasoning (e.g., each indicator should be strongly and significantly related to its respective factor, the size and direction of the factor correlations should be consistent with expectation). Small or statistically nonsignificant estimates may be indicative of unnecessary parameters (e.g., a nonsalient error covariance or indicator cross-loading). In addition, such estimates may highlight indicators that are not good measures of the factors (i.e., a small and nonsignificant primary loading may suggest that the indicator should be removed from the measurement model). On the other hand, extremely large parameter estimates may be substantively problematic. For example, if in a multifactorial solution the factor correlations approach 1.0, there is strong evidence to question whether the latent variables represent distinct constructs (i.e., they have poor discriminant validity). If two factors are highly overlapping, the model could be re-specified by collapsing the dimensions into a single factor. If the fit of the re-specified model is acceptable, it is usually favored because of its better parsimony.

Selected results for the two-factor solution are presented in Table 2. The unstandardized and completely standardized estimates can be found under the headings “MODEL RESULTS” and “STANDARDIZED MODEL RESULTS,” respectively (partially standardized estimates have been deleted from the Mplus output). Starting with the completely standardized solution, the factor loadings can be interpreted along the lines of standardized regression coefficients in multiple regression. For instance, the factor loading estimate for X1 is .760 which would be interpreted as indicating that a standardized unit increase in the Obsessions factor is associated with an .76 standardized score increase in X1. However, because X1 loads on only one factor, this estimate can also be interpreted as the correlation between X1 and the Obsessions latent variable. Accordingly, squaring the factor loading provides the proportion of variance in the indicator that is explained by the factor; e.g., 58% of the variance in X1 is accounted for by Obsessions ($.76^2 = .58$). In the factor analysis literature, these estimates are referred to as a *communalities* (also provided in the Mplus output in Table 2 under the R-SQUARE heading).

The completely standardized estimates under the “Residual Variances” heading (see Table 2) represent the proportion of variance in the indicators that has not been explained by the latent variables (i.e., unique variance). For example, these results indicate that 42% of the variance in X1 was not accounted for by the Obsessions factor. Note that the analyst could readily hand calculate these estimates by subtracting the indicator communality from one; e.g., $\delta_1 = 1 - \lambda_{1,1}^2 = 1 - .76^2 = .42$. Recall it was previously stated that the indicator variances are just-identified in this model (are not potential source of poor fit in this solution). Accordingly, it can be seen that the sum of the indicator communality (λ^2) and the residual variance (δ) will always be 1.0 (e.g., X1: .578 + .422). Finally, the completely standardized results also provide the correlation of the Obsessions and Compulsions factors ($\phi_{2,1} = .394$) and the correlated error of the X9 and X10 indicators ($\delta_{10,9} = .418$). Using these estimates, the model-implied correlations for the 10 indicators can be computed using Eqs. 22.1 through 22.3. For instance, using Eq 22.1, the model-implied correlation for the X2 and X3 indicators is: .688(1).705 = .485. Inserting the completely standardized estimates into Eq. 22.2 yields the following model-implied correlation between X2 and X7: .688(.394).754 = .204. Per Eq. 22.3, the correlation of X9 and X10 that is predicted by the model's parameter estimates is: [.630(.559)] + .268 = .620.[2] Comparison of the sample correlations in Table 1 indicates that these relationships were well approximated by the solution's parameter estimates. This is further reflected by satisfactory global goodness of fit statistics as well as standardized residuals and modification indexes indicating no salient areas of strain in the solution.

The first portion of the Mplus results shown in Table 2 is the unstandardized factor solution (under the heading "MODEL RESULTS"). In addition to each unstandardized estimate (under "Estimate" heading), Mplus provides the standard error of the estimate ("S.E."), the test ratio which can be interpreted as a *z* statistic ("Est./S.E."; i.e., values greater than 1.96 are significant at $\alpha = .05$, two-tailed), and the exact two-sided probability value. In more recent versions of this software program, Mplus also provides standard errors and test statistics for completely (and partially) standardized estimates (also shown in Table 2). The standard errors and significance tests for the unstandardized factor loadings for X1 and X6 are unavailable because these variables were used as marker indicators for the Obsessions and Compulsions factors, respectively (i.e., their unstandardized loadings were fixed to 1.0). The variances for the Obsessions and Compulsions latent variables are 2.49 and 2.32, respectively. These estimates can be calculated using the sample variances of the marker indicators multiplied by their respective communalities. As noted earlier, the communality for X1 was .578 indicating that 57.8% of the variance in this indicator was explained by the Obsessions factor. Thus, 57.8% of the sample variance of X1 ($SD^2 = 2.078^2 = 4.318$; cf. Table 1) is passed along as variance of the Obsessions latent variable; 4.318(.578) = 2.49. As in the completely standardized solution, the factor loadings are regression coefficients expressing the direct effects of the latent variables on the indicators, but in the unstandardized metric; e.g., a unit increase in Obsessions is associated with a .586 increase in X2. The COM WITH OBS estimate (0.948) is the factor covariance of Obsessions and Compulsions, and the X9 WITH X10 estimate (1.249) is the error covariance of these indicators. The residual variances are the indicator uniquenesses or errors (i.e., variance in the indicators that was not explained by the Obsessions and Compulsions latent variables); e.g., $\delta_1 = 4.318(.422) = 1.82$ (again, .422 is the proportion of variance in X1 not explained by Obsessions, see completely standardized residual variances in Table 2).

22.6 CFA Model Respecification

In applied research, a CFA model will often have to be revised. The most common reason for respecification is to improve the fit of the model because at least one of the three aforementioned criteria for model acceptability has not been satisfied (i.e., inadequate global goodness of fit, large standardized residuals or modification indices denoting salient localized areas of ill-fit, the parameter estimates are not uniformly interpretable). On occasion, respecification is conducted to improve the parsimony and interpretability of the model. In this scenario, respecification does not improve the fit of the solution; in fact, it may worsen fit to some degree. For example, the parameter estimates in an initial CFA may indicate the factors have poor discriminant validity; that is, two factors are correlated so highly that the notion they represent distinct constructs must be rejected. The model might be re-specified collapsing the redundant factors; i.e., the indicators that loaded on separate, overlapping factors are respecified to load on one factor. Although this respecification could foster the parsimony and interpretability of the measurement model, it will lead to some decrement in model fit (e.g., χ^2)

relative to the more complex initial solution.

The next sections of this chapter will review the three primary ways a CFA model may be misspecified: (1) the selection of indicators and patterning of indicator-factor relationships, (2) the measurement error theory (e.g., uncorrelated *vs* correlated measurement errors); and (3) the number of factors (too few or too many). As will be discussed further in the ensuing sections, modification indices and standardized residuals are often useful for determining the particular sources of strain in the solution when the model contains minor misspecifications. However, it cannot be over-emphasized that revisions to the model should only be made if they can be strongly justified by empirical evidence or theory. This underscores the importance of having sound knowledge of the sample data and the measurement model before proceeding with a CFA.

In many situations, the success of a model modification can be verified by the χ^2 difference test (χ^2_{diff}). This test can be used when the original and respecified models are nested. A *nested model* contains a subset of the free parameters of another model (the parent model). Consider the two-factor model of Obsessions and Compulsions where the model was evaluated with and without the error covariance for the X9 and X10 indicators. The latter model is nested under the former model because it contains one less freely estimated parameter (i.e., the error covariance is fixed to zero instead of being freely estimated). A nested model will have more *df*s than the parent model, in this example the *df* difference was 1, reflecting the presence/absence of one freely estimated error covariance. If a model is nested under a parent model, the simple difference in the model χ^2s produced by the solutions is distributed as χ^2 in many circumstances (e.g., adjustments must be made if a fitting function other than ML is used to estimate the model; cf. Brown, 2006). The χ^2_{diff} test in this example would be calculated as follows:

	df	χ^2
Model without error covariance	34	109.50
Model with error covariance	33	46.16
χ^2 difference (χ^2_{diff})	1	63.34

Because the models differ by a single degree of freedom, the critical value for the χ^2_{diff} test is

3.84 (α = .05, df = 1). Because the χ^2_{diff} test value exceeds 3.84 (63.34), it would be concluded that the two-factor model with the error covariance provides a significantly better fit to the data than the two-factor model without the error covariance. It is important to note that this two-factor model fit the data well. Use of the χ^2_{diff} test to compare models is not justified when neither solution provides an acceptable fit to the data.

Moreover, the χ^2_{diff} test should not be used if the models are not nested. For instance, two models could not be compared to each other with the χ^2_{diff} test if the input matrix is changed (e.g., an indicator has been dropped from the revised model) or the model has been structurally revised (e.g., two factors have been collapsed into a single latent variable). If the competing models are not nested, in many cases they can be compared using information criterion fit indices such as the Akaike information criterion (AIC) and the expected cross-validation index (ECVI). Both indices take into account model fit (as reflected by χ^2) and model complexity/ parsimony (as reflected by the number of freely estimated parameters). The ECVI also incorporates sample size—specifically, a more severe penalty function for fitting a non-parsimonious model in a smaller sample. Generally, models that produce the smallest AIC and ECVI values are favored over alternative specifications.

Selection of indicators and specification of factor loadings. A common source of CFA model misspecification is the incorrect designation of the relationships between indicators and the factors. This can occur in the following manners (assuming the correct number of factors was specified): (a) the indicator was specified to load on a factor, but actually has no salient relationship to any factor; (b) the indicator was specified to load on one factor, but actually has salient loadings on two or more factors; (c) the indicator was specified to load on the wrong factor. Depending on the problem, the remedy will be either to re-specify the pattern of relationships between the indicator and the factors, or eliminate the indicator from the model.

If an indicator does not load on any factor, this misspecification will be readily detected by results showing the indicator has a nonsignificant or nonsalient loading on the conjectured factor, as well as modification indices suggesting that the fit of the model could not be improved by allowing the indicator to load on a different factor. This conclusion would be further supported by inspection of standardized residuals and sample correlations which point to the fact that the indicator is weakly related or unrelated to other indicators in the model. Although the proper remedial action is to eliminate the problematic indicator, the overall fit of the model will usually not improve appreciably (because the initial solution does not have difficulty reproducing covariance associated with the problematic indicator).

The other two types of indicator misspecifications (failure to specify the correct primary loading or salient cross-loadings) will usually be diagnosed by large modification indexes suggesting that model will be significantly improved if the correct loading was freely estimated (moreover, if the indicator was specified to load on the wrong factor, the factor loading estimate may be small or statistically nonsignificant). However, it is important to emphasize that specification searches based on modification indices are most likely to be successful when the model contains only minor misspecifications. If the initial model is grossly misspecified (e.g., incorrect number of factors, many misspecified factor loadings), specification searches are unlikely to lead the researcher to the correct measurement model (MacCallum, 1986). In addition, even if only one indicator-factor relationship has been misspecified, this can have a serious deleterious impact on overall model fit and fit diagnostic statistics (standardized residuals, modification indices) in small measurement models or when a given factor has been measured by only a few indicators. Although an acceptable model might be obtained using the original set of indicators (after correct specification of the indicator-factor relationships), it is often the case that a better solution will be attained by dropping bad indicators from the model. For example, an indicator may be associated with several large modification indices and standardized residuals, reflecting that it is rather nonspecific (i.e., evidences similar relationships to all latent variables in the model). Dropping this indicator will eliminate multiple strains in the solution.

Measurement error theory. Another common source of CFA model misspecification pertains to the relationships among the indicator residual variances. When no error covariances are specified, the researcher is asserting that all of the covariation among indicators is due to the latent variables (i.e., all measurement error is random). Indicator error covariances are specified on the basis that some of the covariance of the indicators not explained by the latent variables is due to an exogenous common cause. For instance, in measurement models involving multiple-item questionnaires, salient error covariances may arise from items that are very similarly worded, reverse-worded, or differentially prone to social desirability (e.g., see Figure 1). In construct validation studies of multitrait-multimethod (MTMM) matrices, an error theory must be specified to account for method covariance arising from disparate assessment modalities (e.g., self-report, behavioral observation, interview rating; cf. the correlated uniqueness approach to MTMM, Brown, 2006).

Unnecessary error covariances can be readily detected by results indicating their statistical or clinical nonsignificance. The next step would be to refit the model with the error covariance fixed to zero, and verify that the respecification does not result in a significant decrease in model fit. The χ^2_{diff} test can be used in this situation. The more common difficulty is the failure to include salient error covariances in the model. The omission of these parameters is typically manifested by large standardized residuals, modification indices, and EPC values.

Because of the large sample sizes typically involved in CFA, researchers will often encounter "borderline" modification indices (e.g., larger than 3.84, but not of particularly strong magnitude) that suggest that the fit of the model could be improved if error covariances were added to the model. As with all parameter specifications in CFA, error covariances must be supported by a substantive rationale and should not be freely estimated simply to improve model fit. The magnitude of EPC values should also contribute to the decision about whether to free these parameters. The researcher should resist the temptation to use borderline modification indices to over-fit the model. These trivial additional estimates usually have minimal impact on the key parameters of the CFA solution (e.g., factor loadings) and are apt to be highly unstable (i.e., reflect sampling error rather than an important relationship). In addition, it is important to be consistent in the decision rules used to specify correlated errors; i.e., if there is a plausible reason for correlating the errors of two indicators, then all pairs of indicators for which this reasoning applies should also be specified with correlated errors (e.g., if it is believed that method effects exist for questionnaire items that are reverse-worded, error covariances should be freely estimated for all such indicators, not just a subset of them).

Number of factors. This final source of misspecification should be the least frequently encountered by applied CFA researchers. If the incorrect number of latent variables has been specified, it is likely that the researcher has moved into the CFA framework prematurely. CFA hinges on a strong conceptual and empirical basis. Thus, in addition to strong conceptual justification, the CFA model is usually supported by prior exploratory analyses (i.e., EFA) that have established the appropriate number of factors, and the correct pattern of indicator-factor relationships. Accordingly, gross misspecifications of this nature should be unlikely if the proper groundwork for CFA has been established.

As was discussed earlier, over-factoring is often reflected by excessively high factor correlations. The remedial action would to be collapse factors or perhaps eliminate a redundant factor altogether (e.g., if an objective is to make a test instrument as brief as possible). An under-factored CFA solution will fail to adequately reproduce the observed relationships among the indicators. Generally speaking, if indicators are incorrectly specified to load on the same factor (but belong on separate factors), standardized residuals will reveal that the solution's parameter estimates markedly underestimated the observed relationships among these indicators. It is noteworthy that if a one-factor model has been specified (but the "true" model consists of two factors), modification indices will only appear in sections of the results that pertain to indicator measurement errors (i.e., modification indices will not appear as possible cross-loadings because a single factor has been specified). The fact that modification indices appear in this fashion might lead the novice CFA researcher to conclude that indicator error covariances are required. However, modification indices can point to problems with the model that are not the real source of ill-fit (e.g., a salient modification index may suggest a revision that makes no conceptual sense). This again underscores the importance of having a clear substantive basis (both conceptual and empirical) to guide the model specification.

22.7 Summary

This chapter provided an overview of the purposes and methods of applied CFA. CFA is in fact the foundation of structural equation modeling because all latent variable analyses rely on a sound measurement model. The fundamental principles and procedures discussed in this chapter are indispensible to a variety of common research applications such as psychometric evaluation, construct validation, data reduction, and identification of bias in measurement. More advanced applications of CFA offer a host of other modeling opportunities such as the analysis of categorical outcomes (as a complement or alternative to IRT), measurement invariance (across time or across population subgroups), higher-order factor analysis (e.g., second-order factor analysis, bifactor analysis), and the analysis of mean structures (see Brown, 2006, for detailed discussion of these applications). Moreover, recent developments in statistical software packages have fostered the integration of CFA with other analytic traditions such as latent class modeling (e.g., factor mixture models; Lubke & Muthén, 2005) and hierarchical linear modeling (e.g., multilevel factor analysis; Muthén & Asparouhov, 2010). Based on its pertinence to a wide range of empirical endeavors, CFA will continue to be one of the most frequently used statistical methods in the applied social and behavioral sciences.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397-438.

Bentler, P.M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Brown, T.A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy*, *41*, 1411-1426.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

Green, S.B., & Thompson, M.S. (in press). Analyzing means in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York: Guilford Press.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351-362.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109-133.

Kenny, D.A., & Kashy, D.A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165-172.

Lei, P.W., & Wu, Q. (in press). Estimation in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York: Guilford Press.

Lubke, G.H., & Muthén, B.O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21-39.

MacCallum, R.C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107-120.

MacCallum, R.C., Roznowski, M., & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.

Marsh, H.W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810-819.

Marsh, H.W., Hau, K.T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*, 439-447.

Millsap, R.E., & Olivera-Aguilar, M. (in press). Investigating measurement invariance using confirmatory factor analysis. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York: Guilford Press.

Muthén, B., & Asparouhov, T. (2010). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J.K. Roberts (Eds.), *Handbook of advanced multilevel analysis*. New York: Taylor and Francis.

Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished technical report, University of California, Los Angeles.

Muthén, L.K., & Muthén, B.O. (1998-2010). *Mplus user's guide* (6th ed). Los Angeles: Author.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, *54*, 315-323.

Rosellini, A.J., & Brown, T.A. (in press). The NEO Five Factor Inventory: Psychometric properties and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment*.

Thurstone, L.L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

Wothke, W.A. (1993). Nonpositive definite matrices in structural modeling. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage.

Footnotes

[1]An exception to the distinctions between EFA and CFA discussed in this chapter is a new method called "exploratory structural equation modeling" (ESEM; Asparouhov & Muthén, 2009), available only in the Mplus software program (beginning with version 5.21). ESEM integrates EFA and CFA measurement models within the same solution. That is, within a given measurement model, some factors can be specified per the conventions of CFA (i.e., zero cross-loadings) whereas other factors can be specified as an EFA (i.e., rotation of a full factor loading matrix). Unlike traditional EFA, the EFA measurement model in ESEM provides the same information as CFA such as multiple indices of goodness of fit, standard errors for all rotated parameters, and modification indices (i.e., highlighting possible correlated residuals among indicators). Moreover, most of the modeling possibilities of CFA are available in ESEM including correlated residuals, regressions of factors on covariates, regression among factors (among different EFA factor blocks or between EFA and CFA factors), multiple-group solutions, mean structure analysis, and measurement invariance examination across groups or across time. A technical description of ESEM can be found in Asparouhov and Muthén (2009); see Marsh et al. (2009, in press) and Rosellini and Brown (in press) for initial applied studies.

[2]In recent releases, Mplus began to compute standardized error covariances differently than other leading software programs (e.g., LISREL). The following formula converts a covariance into a correlation: $CORR_{1,2} = COV_{1,2}$ / $SQRT(VAR_1 * VAR_2)$. While the unstandardized error covariances ($COV_{1,2}$) are the same across software programs, Mplus computes the error correlation using the indicator residual variances (VAR_1, VAR_2) rather than the sample variances of the indicators. Consequently, the standardized error covariances reported by Mplus are larger than those derived from the same unstandardized factor solution in another software program. To obtain the correct model-implied correlation (e.g., between X9 and X10) the more traditional computational method should be used; e.g., error correlation of X9 and X10 = 1.249 / (2.666 * 1.745) = .268 (cf. error correlation = .418 in Mplus output, Table 2).

Table 1

Sample Correlations and Standard Deviations (SD) for the Ten-Item Questionnaire of Obsessive and Compulsive Symptoms (N = 400)

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.000									
X2	0.516	1.000								
X3	0.563	0.465	1.000							
X4	0.453	0.462	0.395	1.000						
X5	0.457	0.479	0.480	0.436	1.000					
X6	0.331	0.133	0.210	0.214	0.147	1.000				
X7	0.286	0.170	0.211	0.187	0.148	0.633	1.000			
X8	0.287	0.177	0.253	0.200	0.196	0.639	0.550	1.000		
X9	0.247	0.094	0.130	0.166	0.114	0.519	0.500	0.467	1.000	
X10	0.243	0.128	0.112	0.118	0.156	0.463	0.430	0.419	0.621	1.000
SD:	2.078	1.346	1.958	1.247	1.588	1.820	2.460	2.368	2.666	1.745

Table 2

Mplus Syntax and Selected Output for CFA Model of Obsessions and Compulsions

```
TITLE: TWO-FACTOR CFA OF OBSESSIONS AND COMPULSIONS
DATA:
  FILE IS CFA.DAT;
  NOBSERVATIONS = 400;
  TYPE = STD CORRELATION;
VARIABLE:
  NAMES ARE X1-X10;
ANALYSIS:
  ESTIMATOR IS ML;
MODEL:
  OBS BY X1-X5;
  COM BY X6-X10;
  X9 WITH X10;
OUTPUT: STANDARDIZED MODINDICES(4) RES;
```

```
MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 OBS      BY
    X1                 1.000      0.000    999.000    999.000
    X2                 0.586      0.047     12.479      0.000
    X3                 0.874      0.068     12.764      0.000
    X4                 0.488      0.043     11.277      0.000
    X5                 0.658      0.055     11.922      0.000

 COM      BY
    X6                 1.000      0.000    999.000    999.000
    X7                 1.215      0.079     15.355      0.000
    X8                 1.168      0.076     15.324      0.000
    X9                 1.100      0.088     12.526      0.000
    X10                0.639      0.059     10.912      0.000

 COM      WITH
    OBS                0.948      0.159      5.954      0.000

 X9       WITH
    X10                1.249      0.184      6.781      0.000

 Variances
    OBS                2.490      0.305      8.162      0.000
    COM                2.322      0.244      9.529      0.000

 Residual Variances
    X1                 1.817      0.183      9.956      0.000
    X2                 0.953      0.083     11.412      0.000
    X3                 1.924      0.173     11.127      0.000
    X4                 0.958      0.078     12.274      0.000
    X5                 1.437      0.121     11.866      0.000
    X6                 0.983      0.120      8.181      0.000
    X7                 2.607      0.242     10.785      0.000
```

```
    X8                 2.428      0.224     10.818      0.000
    X9                 4.280      0.343     12.484      0.000
    X10                2.088      0.161     12.963      0.000


STANDARDIZED MODEL RESULTS

STDYX Standardization
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 OBS      BY
    X1                 0.760      0.029     26.282      0.000
    X2                 0.688      0.033     20.838      0.000
    X3                 0.705      0.032     22.029      0.000
    X4                 0.618      0.037     16.747      0.000
    X5                 0.655      0.035     18.783      0.000

 COM      BY
    X6                 0.838      0.023     36.280      0.000
    X7                 0.754      0.028     27.322      0.000
    X8                 0.752      0.028     27.185      0.000
    X9                 0.630      0.035     17.906      0.000
    X10                0.559      0.039     14.236      0.000

 COM      WITH
    OBS                0.394      0.052      7.563      0.000

 X9       WITH
    X10                0.418      0.045      9.343      0.000

 Variances
    OBS                1.000      0.000    999.000    999.000
    COM                1.000      0.000    999.000    999.000

 Residual Variances
    X1                 0.422      0.044      9.592      0.000
    X2                 0.527      0.045     11.620      0.000
    X3                 0.503      0.045     11.147      0.000
    X4                 0.618      0.046     13.533      0.000
    X5                 0.571      0.046     12.510      0.000
    X6                 0.297      0.039      7.677      0.000
    X7                 0.432      0.042     10.381      0.000
    X8                 0.434      0.042     10.423      0.000
    X9                 0.604      0.044     13.633      0.000
    X10                0.688      0.044     15.664      0.000


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    X1                 0.578      0.044     13.141      0.000
    X2                 0.473      0.045     10.419      0.000
    X3                 0.497      0.045     11.015      0.000
    X4                 0.382      0.046      8.374      0.000
    X5                 0.429      0.046      9.392      0.000
```

X6	0.703	0.039	18.140	0.000
X7	0.568	0.042	13.661	0.000
X8	0.566	0.042	13.592	0.000
X9	0.396	0.044	8.953	0.000
X10	0.312	0.044	7.118	0.000

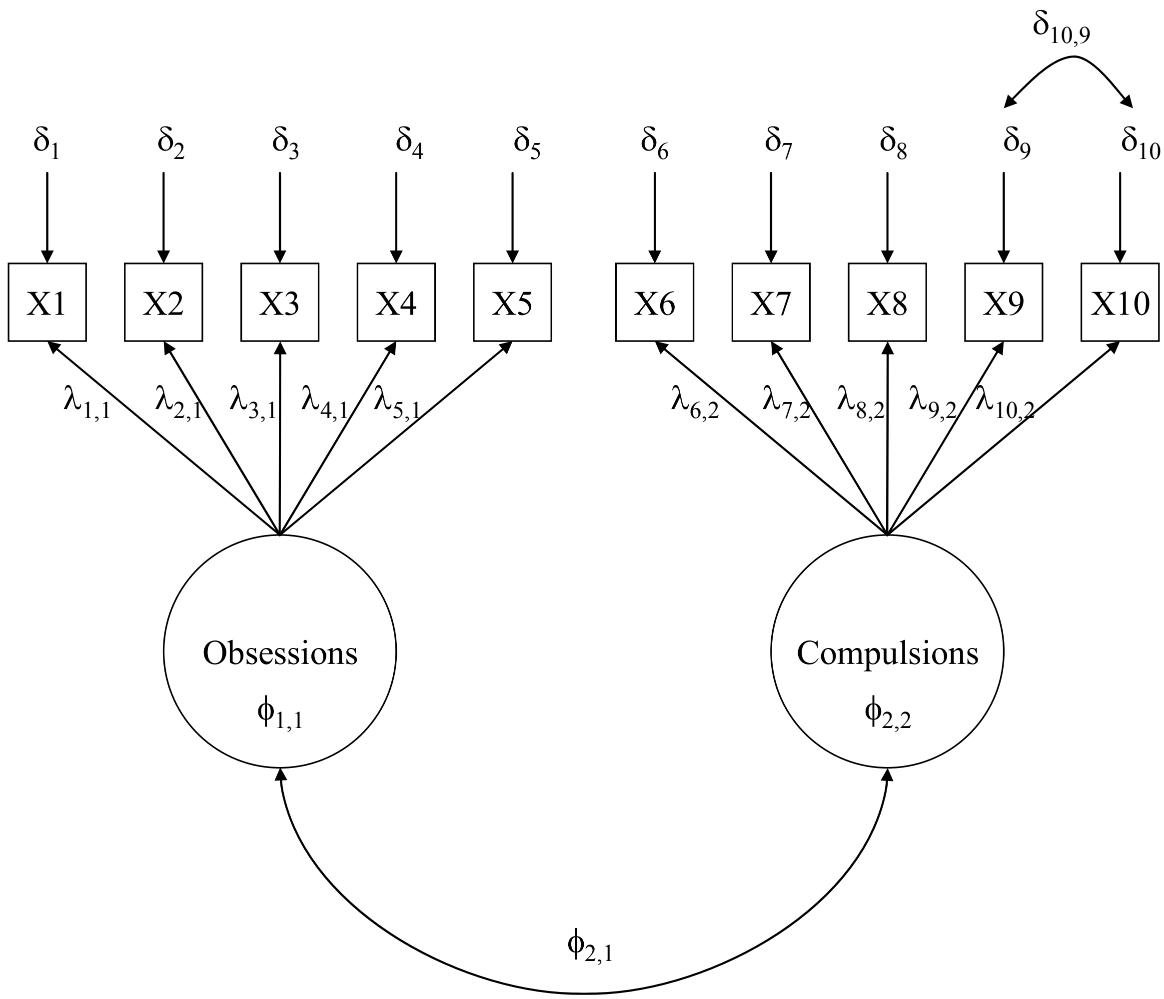

Figure 1. Two-factor measurement model of obsessions and compulsions.