1

Introduction

USES OF CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis (CFA) is a type of structural equation modeling (SEM) 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*. A fundamental feature of CFA is its hypothesis-driven nature. Unlike its counterpart, exploratory factor analysis (EFA), CFA requires the researcher to prespecify all aspects of the model. Thus the researcher must have a firm a priori sense, based on past evidence and theory, of the number of factors that exist in the data, of which indicators are related to which factors, and so forth. In addition to its greater emphasis on theory and hypothesis testing, the CFA framework provides many other analytic possibilities that are not available in EFA (e.g., evaluation of method effects, examination of the stability or invariance of the factor model over time or informants). Moreover, for the reasons discussed below, CFA should be conducted prior to the specification of a structural equation model.

CFA has become one of the most commonly used statistical procedures in applied research. This is because CFA is well equipped to address the types of questions that researchers often ask. Some of the most common uses of CFA are discussed below.

Psychometric Evaluation of Test Instruments

CFA is almost always used during the process of scale development to examine the latent structure of a test instrument (e.g., a questionnaire). In this context, CFA is used to verify 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. 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 may be scored by using subscales; that is, the number of factors is indicative of the number of subscales, and the pattern of item–factor relationships (which items load on which factors) indicates how the subscales should be scored. Depending on other results and extensions of the analysis, CFA may support the use of total scores (composite of all items) in addition to subscale scores (composites of subsets of items). For example, the viability of a single total score might be indicated when the relationships among the latent dimensions (subscales) of a test can be accounted for by one higher-order factor, and when the test items are meaningfully related to the higher-order factor (see the discussion of higher-order CFA and bifactor modeling in Chapter 8). CFA is an important analytic tool for other aspects of psychometric evaluation. It can be used to estimate the *scale reliability* of test instruments in a manner that avoids the problems of traditional methods (e.g., Cronbach's alpha; see Chapter 8). Given recent advances in the analysis of categorical data (e.g., binary true–false test items), CFA now offers a comparable analytic framework to item response theory (IRT). In fact, in some ways, CFA provides more analytic flexibility than the traditional IRT model (see Chapter 9).

Construct Validation

Akin to a factor in CFA, a *construct* is a theoretical concept. In clinical psychology and psychiatry, for example, the mental disorders (e.g., major depression, schizophrenia) are constructs manifested by various clusters of symptoms that are reported by the patient or observed by others. In sociology, juvenile delinquency may be construed as a multidimensional construct defined by various forms of misconduct (e.g., property crimes, interpersonal violence, participation in drugs, academic misconduct). CFA is an indispensable analytic tool for construct validation in the social and behavioral sciences. 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; for example, 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; for example, behaviors purported to be manifestations of different types of delinquency 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 (MTMM) matrices (see Chapter 6). A fundamental strength of CFA approaches to construct validation is that the resulting estimates of convergent and discriminant validity are adjusted for measurement error and an error theory (see the "Method Effects" section, below). Thus CFA provides a stronger analytic framework than traditional 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).

Method Effects

Often some of the covariation of observed measures is due to sources other than the substantive latent variables. For instance, consider the situation where four measures of employee morale have been collected. Two indicators are the employees' self-reports (e.g., questionnaires); the other two are obtained from supervisors (e.g., behavioral observations). It may be presumed that the four measures are intercorrelated, because each is a manifest indicator of the underlying construct of morale. However, it is also likely that the employee self-report measures are more highly correlated with each other than with the supervisor measures, and vice versa. This additional covariation is not due to the underlying construct of morale, but reflects shared method variance. A *method effect* exists when additional covariation among indicators is introduced by the measurement approach. Method effects can also occur within a single assessment modality. For example, method effects are usually present in questionnaires that contain some combination of positively and negatively worded items (e.g., see Chapters 3 and 6). Unfortunately, traditional EFA is incapable of estimating method effects. In fact, the use of EFA when method effects exist in the data can produce misleading results—that is, can yield additional factors that are not substantively meaningful, but instead stem from artifacts of measurement. In CFA, however, method effects can be specified as part of the error theory of the measurement model. The advantages of estimating method effects within CFA include the ability to (1) specify measurement models that are more conceptually viable; (2) determine the amount of method variance in each indicator; and (3) obtain better estimates of the relationships of the indicators to the factors, and the relationships among latent variables (see Chapters 5 and 6).

Measurement Invariance Evaluation

Another key strength of CFA is its ability to determine how well measurement models generalize across groups of individuals or across time. *Measurement invariance* evaluation is an important aspect of test development. If a test is intended to be administered in a heterogeneous population, it should be established that its measurement properties are equivalent in subgroups of the population (e.g., gender, race). A test is said to be biased when some of its items do not measure the underlying construct comparably across groups. *Test bias* can be serious, such as in situations where a given score on a cognitive ability or job aptitude test does not represent the same true level of ability/aptitude in male and female respondents. Stated another way, the test is biased against women if, for a given level of true intelligence, men tend to score several IQ units higher on the test than women. These questions can be addressed in CFA by multiple-groups solutions and "multiple indicators, multiple causes" (MIMIC) models (see Chapter 7). For instance, in a multiple-groups CFA solution, the measurement model is estimated simultaneously in various subgroups (e.g., men and women). Other restrictions are placed on the multiple-groups solution to determine the equivalence of the measurement model across groups; for instance, if the factor loadings are equivalent, the magnitude of the relationships between the test items and the underlying construct (e.g., cognitive ability) are the same in men and women. Multiple-groups CFA solutions are also used to examine longitudinal measurement invariance. This is a very important aspect of latent variable analyses of repeated measures designs. In the absence of such evaluation, it cannot be determined whether temporal change in a construct is due to true change or to changes in the structure or measurement of the construct over time. Multiple-groups analysis can be applied to any type of CFA or SEM model. For example, these procedures can be incorporated into the analysis of MTMM data to examine the generalizability of construct validity across groups.

WHY A BOOK ON CFA?

It also seems appropriate to begin this volume by addressing this question: "Is there really a need for a book devoted solely to the topic of CFA?" On my bookshelf sit dozens of books on the subject of SEM. Why not go to one of these SEM books to learn about CFA? Given that CFA is a form of SEM, virtually all of these books provide some introduction to CFA. However, this coverage typically consists of a chapter at best. As this book will attest, CFA is a very broad and complex topic, and extant SEM books only scratch the surface. This is unfortunate, because in applied SEM research, most of the work deals with measurement models (CFA). Indeed, many applied research questions are addressed by using CFA as the primary analytic procedure (e.g., psychometric evaluation of test instruments, construct validation). Another large proportion of SEM studies focus on structural regression models—that is, the manner in which latent variables are interrelated. Although CFA is not the ultimate analysis in such studies, a viable measurement model (CFA) must be established prior to evaluating the structural (e.g., regressive) relationships among the latent variables of interest. When poor model fit is encountered in such studies, it is more likely that it will stem from misspecifications in the measurement portion of the model (i.e., the manner in which observed variables are related to factors) than from the structural component that specifies the interrelationships of the factors. 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 latent variables). Existing SEM resources do not provide sufficient details on the sources of ill fit in CFA measurement models or on how such models can be diagnosed and respecified. Moreover, advanced applications of CFA are rarely discussed in general SEM books (e.g., CFA with categorical indicators, scale reliability evaluation, MIMIC models, formative indicators, multilevel measurement models, Bayesian CFA).

Given the popularity of CFA, this book was written to provide an in-depth treatment of the concepts, procedures, pitfalls, and extensions of this methodology. Although the overriding objective of the book is to provide critical information on applied CFA that has not received adequate coverage in the past, it is important to note that the topics pertain to SEM in general (e.g., sample size/power analysis, missing data, non-normal or categorical data, formative indicators, multilevel modeling, Bayesian analysis). Thus it is hoped that this book will also provide a useful resource to researchers using any form of SEM.

COVERAGE OF THE BOOK

The first five chapters of this book present the fundamental concepts and procedures of CFA. Chapter 2 introduces the reader to the concepts and terminology of the common factor model. The common factor model is introduced in the context of EFA. This book is not intended to be a comprehensive treatment of the principles and practice of EFA. However, an overview of the concepts and operations of EFA is provided in Chapter 2 for several reasons: (1) Most of the concepts and terminology of EFA generalize to CFA; (2) this overview fosters the discussion of the similarities and differences of EFA and CFA in later chapters (e.g., Chapter 3); and (3) in programmatic research, an EFA study is typically conducted prior to a CFA study to develop and refine measurement models that are reasonable for CFA (thus the applied CFA researcher must also be knowledgeable about EFA). An introduction to CFA is provided in Chapter 3. After providing a detailed comparison of EFA and CFA, this chapter presents the various parameters, unique terminology, and fundamental equations of CFA models. Many other important concepts are introduced in this chapter that are essential to the practice of CFA and that must be understood in order to proceed to subsequent chapters; these include model identification, model estimation (e.g., maximum likelihood or ML), and goodness of model fit. Chapter 4 illustrates and extends these concepts, using a complete example of a CFA measurement model. In this chapter, the reader will learn how to program and interpret basic CFA models, using several of the most popular latent variable software packages (LISREL, Mplus, EQS, SAS/CALIS). The procedures for evaluating the acceptability of the CFA model are discussed. In context of this presentation, the reader is introduced to other important concepts, such as model misspecification and Heywood cases. Chapter 4 concludes with a section on the material that should be included in the report of a CFA study. Chapter 5 covers the important topics of model respecification and model comparison. It deals with the problem of poor-fitting CFA models and the various ways a CFA model may be misspecified. This chapter also presents EFA within the CFA framework and exploratory SEM—methods of developing more viable CFA measurement models on the basis of EFA findings. The concepts of nested models, equivalent models, and method effects are also discussed.

The second portion of the book focuses on more advanced or specialized topics and issues in CFA. Chapter 6 discusses how CFA can be conducted to analyze MTMM data in the validation of social or behavioral constructs. Although the concepts of method effects, convergent validity, and discriminant validity are introduced in earlier chapters (e.g., Chapter 5), these issues are discussed extensively in context of MTMM models in Chapter 6. Chapter 7 discusses CFA models that contain various combinations of equality constraints (e.g., estimation of a CFA model with the constraint of holding two or more parameters to equal the same value), multiple groups (e.g., simultaneous CFA in separate groups of males and females), and mean structures (CFAs that entail the estimation of the intercepts of indicators and factors). These models are discussed and illustrated in context of the analysis of measurement invariance; that is, is the measurement model equivalent in different groups or within the same group across time? Two different approaches to evaluating CFA models in multiple groups are presented in detail: multiple-groups solutions and MIMIC models. This chapter also illustrates another method of scaling of latent variables, called *effects coding*, that can foster the interpretation of the unstandardized solution in single-group and multiple-groups analyses.

Chapter 8 presents four other types of CFA models: higher-order CFA, bifactor analysis, CFA approaches to scale reliability estimation, and CFA with formative indicators. Higher-order factor analysis is conducted in situations where the researcher can posit a more parsimonious conceptual account for the interrelationships of the factors in the initial CFA model. Bifactor modeling is another form of hierarchical analysis, but unlike higher-order factor analysis, the overarching dimension (or dimensions) exerts direct effects on the indicators. The section on scale reliability evaluation shows that the unstandardized parameter estimates of a CFA solution can be used to obtain point estimates and confidence intervals of the reliability of test instruments (i.e., reliability estimate = the proportion of the total observed variance in a test score that reflects true score variance). This approach has important advantages over traditional estimates of internal consistency (Cronbach's alpha). Models with formative indicators contain observed measures that "cause" the latent construct. In the typical CFA, indicators are defined as linear functions of the latent variable, plus error; that is, indicators are considered to be the effects of the underlying construct. In some situations, however, it may be more plausible to view the indicators as causing a latent variable; for example, socioeconomic status is a concept determined by one's income, education level, and job status (not the other way around). Although formative indicators pose special modeling challenges, Chapter 8 shows how such models can be handled in CFA.

The next two chapters consider issues that must often be dealt with in applied CFA research, but are rarely discussed in extant SEM sourcebooks. Chapter 9 addresses data set complications such as how to accommodate missing data, and how to conduct CFA when the distributions of continuous indicators are non-normal. Various methods of handling each issue are discussed and illustrated (e.g., missing data: multiple imputation, direct ML; non-normal data: alternative statistical estimators, bootstrapping, item parceling). Chapter 9 also includes a detailed treatment of CFA with categorical outcomes (e.g., tests with binary items such as true–false scales). In addition to illustrating the estimation and interpretation of such models, this section of the chapter demonstrates the parallels and extensions of CFA to traditional IRT analysis. The section also contains a detailed discussion and illustration of measurement invariance evaluation with categorical outcomes. Chapter 10 deals with the often overlooked topic of determining the sample size necessary to achieve sufficient statistical power and precision of the parameter estimates in a CFA study. Two different approaches to this issue are presented (Satorra–Saris method, Monte Carlo method).

The final chapter of this book (Chapter 11) presents and illustrates two relatively new modeling possibilities involving CFA: Bayesian analysis and multilevel factor models. Among the many advantages of Bayesian CFA is the ability to specify approximate zeroes for parameters that are fixed to zero in traditional CFA (e.g., cross-loadings). This allows for the testing of more reasonable measurement models that may be more closely aligned with substantive theory. Multilevel factor models should be conducted in data sets that violate the assumption of independence of observations (e.g., clustered data, such as in a study of families where more than one member per family has contributed data). In addition to properly adjusting for the dependency in the data, multilevel models allow the researcher to study the structure and nature of relationships at each level of the data.

OTHER CONSIDERATIONS

This book was written with applied researchers and graduate students in mind. It is intended to be a user-friendly guide to conducting CFA with real data sets. To achieve this goal, conceptual and practical aspects of CFA are emphasized, and quantitative aspects are kept to a minimum (or separated from the main text; e.g., Chapter 3). Formulas are not avoided altogether, but are provided in instances where they are expected to foster the reader's conceptual understanding of the principles and operations of CFA. Although this book does not require a high level of statistical acumen, a basic understanding of correlation/multiple regression will facilitate the reader's passage through the occasional more technically oriented section.

It is important that a book of this nature not be tied to a specific latent variable software program. For this reason, most of the examples provided in this book are accompanied with input syntax from each of the most widely used software programs (LISREL, Mplus, EQS, SAS/CALIS). Although Amos Basic was included in the first edition of this book, Amos is not included in the current edition because the vast majority of Amos users rely on the graphical interface for model specification (Amos Basic syntax can still be found on the book's companion website, *www.guilford.com/brown3-materials*). Several comments about the examples are in order. First, readers will note that many of the syntax examples are first discussed in context of the LISREL program. This is not intended to imply a preference for LISREL over other software programs. Rather, this is more reflective of the historical fact that the etiological roots of SEM are strongly tied to LISREL. For instance, the widely accepted symbols of the parameters and computational formulas of a CFA model stem from LISREL notation (e.g., λ = factor loading). The illustrations of LISREL matrix programming allow interested readers to understand computational aspects of CFA more quickly (e.g., by using the provided formula, how the model-implied covariance of two indicators can be calculated on the basis of the

CFA model's parameter estimates). Knowledge of this notation is also useful to readers who are interested in developing a deeper quantitative understanding of CFA and SEM in more technical sources (e.g., Bollen, 1989). On the other hand, the output from the Mplus program is relied on heavily in various examples in the book. Again, a preference for Mplus should not be inferred. This reflects the fact that the results of CFA are provided more succinctly by Mplus than by other programs (concise output = concise tables in this book), as well as the fact that some analytic features are only available in Mplus.

Another potential pitfall of including computer syntax examples is the high likelihood that soon after a book is published, another version of the software will be released. When this book was written, the following versions of the software programs were current: LISREL 9.1, Mplus 7.11, EQS 6.2, and SAS/CALIS 9.4. For all examples, the associated computer syntax was revised to be up to date with these versions of these software programs. New releases typically introduce new features to the software, but do not alter the overall programming framework. In terms of this book, the most probable consequence of new software releases is that some claims about the (in)capabilities of the programs will become outdated. However, the syntax examples should be upwardly compatible (i.e., fully functional) with any subsequent software releases (e.g., although LISREL 7 does not contain many of the features of LISREL 9.1, syntax written in this version is fully operational in subsequent LISREL releases).

Especially in earlier chapters of this book, the computer syntax examples contain few, if any, programming shortcuts. Again, this is done to foster the reader's understanding of CFA model specification. This is another reason why LISREL is often used in the programming examples: in LISREL matrix-based programming, the user must specify every aspect of the CFA model. Thus CFA model specification is more clearly conveyed in LISREL than in some programs where these specifications occur "behind the scenes" (e.g., Mplus contains a series of defaults that automatically specify marker indicators, free and fixed factor loadings, factor variances and covariances, etc., in a standard CFA model). On a related note, many latent variable software programs (e.g., Amos, LISREL, EQS) now contain graphical interfaces that allow the user to specify the CFA model by constructing a path diagram with a set of drawing tools. Indeed, graphical interfaces are an increasingly popular method of model programming, particularly with researchers new to CFA and SEM. The primary reason why graphical input is not discussed in this book is that it does not lend itself well to the written page. Yet there are other reasons why syntax programming can be more advantageous. For instance, it is often quicker to generate a CFA solution from a syntax program than from constructing a path diagram in a drawing editor. In addition, many of the advanced features of model specification are more easily invoked through syntax. Users who understand the logic of syntax programming (either matrix- or equation-based syntax operations) are able to move from one latent variable software package to another much more quickly and easily than users who are adept only in the use of a graphical interface of a given software program.

In attempt to make the illustrations more provocative to the applied researcher, most of the examples in this book are loosely based on findings or test instruments in the extant literature. The examples are drawn from a variety of domains within the social and behavioral sciences—clinical psychology, personality psychology, social psychology, industrial/organizational psychology, and sociology. In some instances, the examples use actual research data, but in many cases the data have been artificially generated strictly for the purposes of illustrating a given concept. Regardless of the origin of the data, the examples should not be used to draw substantive conclusions about the research domain or test instrument in question.

Many of the examples in this book use a variance–covariance matrix as input data (specifically, the correlations and standard deviations of the indicators are inputted, from which the program generates the sample variance–covariance matrix). This was done to allow interested readers to replicate examples directly from the information provided in the figures and tables of the book. Although matrix input is used as a convenience feature in this book, it is not necessarily the best method of reading data into an analysis. All leading latent variable programs are capable of reading raw data as text files, and many can read data saved in other software formats (e.g., SPSS .sav files, Microsoft Excel files). There are several advantages of using raw data as input. First, it is more convenient, because the user does not need to compute the input matrices prior to conducting the latent variable analysis. Second, the input data are more precise when the software program computes the input matrices from raw data (user-generated matrices usually contain rounding error). Third, there are some situations where raw data must be analyzed—for instance, models that have missing, non-normal, or categorical data. Some sections of this book (e.g., Chapter 9) illustrate how raw data are read into the analysis. The interested reader can download the files used in these and other examples from the book's companion website (*www.guilford.com/brown3-materials*).

SUMMARY

This chapter has provided a general overview of the nature and purposes of CFA, including some of the fundamental differences between EFA and CFA. The ideas introduced in this chapter provide the background for a more detailed discussion of the nature of the common factor model and EFA, the subject of Chapter 2. This book is intended to be a user-friendly guide to conducting CFA in real data sets, aimed at students and applied researchers who do not have an extensive background in quantitative methods. Accordingly, practical and conceptual aspects of CFA are emphasized over mathematics and formulas. In addition, most of the chapters are centered on data-based examples drawn from various realms of the social and behavioral sciences. The overriding rationale of these examples is discussed (e.g., use of software programs, method of data input) to set the stage for their use in subsequent chapters.