THE SEQUENTIAL KAISER-MEYER-OLKIN PROCEDURE AS AN ALTERNATIVE FOR DETERMINING THE NUMBER OF FACTORS IN COMMON-FACTOR ANALYSIS: A MONTE CARLO SIMULATION

By

BRENT DALE HILL

Bachelor of Science
University of Tulsa
Tulsa, Oklahoma
1997

Master of Science
Oklahoma State University
Stillwater, Oklahoma
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2011

THE SEQUENTIAL KAISER-MEYER-OLKIN PROCEDURE AS AN ALTERNATIVE FOR DETERMINING THE NUMBER OF FACTORS IN COMMON-FACTOR ANALYSIS: A MONTE CARLO SIMULATION

Dissertation Approved:

Dr. Dale R. Fuqua

Dissertation Advisor

Dr. Katye Perry

Dr. Janice Miller

Dr. C. Robert Davis

Dr. Mark E. Payton

Dean of the Graduate College

For Mom & Dad

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The family of statistical and psychometric methods known as factor analysis has long been a reliable workhorse for the behavioral and social sciences. Broadly defined as the analysis of interdependence (Kendall & Babington-Smith, 1950), factor analysis is actually composed of numerous specialized forms. In particular, two exploratory techniques have become prominent in the literature—namely, common-factor analysis (CFA) and principal-components analysis (PCA). Unfortunately, the taxonomy for the assorted contemporary factor-analytic models and methods is somewhat arbitrary as the nomenclature and mode of categorization vary from source to source (Costello & Osborne, 2005). Even the term *factor analysis* itself is ambiguous: Some authors use it generically to refer to both CFA and PCA together, while others use the term synonymously with CFA alone. The abbreviated taxonomy presented here was adapted from those of Gorsuch (1983) and Rummel (1970) and uses the term *factor analysis* in the more general sense to encompass both CFA and PCA.

Although the original model proposed by Spearman (1904) allowed for only a simple one-factor solution, Thurstone (1947) refined the methods of factor analysis to produce a more sophisticated system that fully embraced a multidimensional framework. This multiple-factor model naturally raised the question regarding the correct number of factors to retain in applied research. In exploratory factor analysis the correct number of factors is not known a priori, which forces the researcher to decide upon the number of factors to extract (Gorsuch, 1983). The question regarding the proper number of factors to extract is one of the most important issues for a

researcher to consider when using factor analysis (Cattell & Vogelmann, 1977; Crawford, 1975; Hakstain & Muller, 1973). This is a critical decision, in part, since the number of factors has a direct influence on the subsequent parameter estimates and the interpretation of the solution (Lambert, Wildt, & Durand, 1990). Of course, the theoretical importance of the number of factors is also crucial.

Before electronic computers, CFA and PCA were not frequently used techniques due to the overwhelming number of computations involved (Harman, 1976; Harris, 1975). In the early 1960s when computers first became accessible to the general academic community, statistical programs had to be fairly uncomplicated due to the limitations in processing resources and speed. Since PCA is not as complex and computationally demanding as CFA, the former became popular as a computationally expedient approximation for the latter (Gorsuch, 1990). This practice has carried over to the present as most PCAs found in the literature are actually approximations of CFAs (Gorsuch, 2003). With modern computers, however, limitations of processing resources are no longer a consideration. So, there is no reason to rely on a convenient approximation rather than using the most theoretically appropriate model (McArdle, 1990). Humphreys and Ilgen (1969) succinctly summarize this point: "Factor analysis should not be confounded with component analysis" (p. 572). Specifically, CFA is more effective as a method to detect and delineate latent traits, whereas PCA is best used to provide a parsimonious and empirical description of observed data (Floyd & Widaman, 1995; Hong, Mitchell, & Harshman, 2006). Proper theoretical alignment also applies to the various auxiliary methods, such as dimensionality tests. Thus, any method for determining the number of factors to retain in CFA should be based upon the common-factor model (Hakstain & Muller, 1973; Humphreys & Ilgen, 1969).

Problem Statement

That the evolution of factor analysis has paralleled that of the electronic computer is no coincidence. Early methods for computing factor-analytic solutions were computationally

expedient approximations; calculating a common-factor or principal-component solution by hand was simply not feasible. A steady increase in the availability and power of computers has afforded researchers the ability to produce computationally intensive factor solutions. Ancillary methods—such as factor-rotation algorithms and number-of-factors tests—also became increasingly sophisticated and numerous. However, the antiquated Kaiser-Guttman rule remains the most popular method even though much better dimensionality indicators now exist (Nasser, Benson, & Wisenbaker, 2002; Stellefson & Hanik, 2008). This is, in part, due to the relatively large amount computational resources required to perform these newer methods (O'Connor, 2000). However, given the readily available power of modern desktop computers, this should no longer be an issue.

Unfortunately, there is also an issue involving the theoretical alignment of the two most promising dimensionality indicators. The MAP test is based explicitly on the component model (Fabrigar, Wegener, MacCallum, & Strahan, 1999), and many argue that parallel analysis is in fact aligned with PCA as well (e.g., Crawford & Koopman, 1973). Just as PCA has been used as an approximation of a common-factor solution, these number-of-factor methods based on PCA have been used as approximations for the number of factors in CFA (Velicer, Eaton, & Fava, 2000). The lack of a number-of-factors method based on the common-factor model is most likely due to the fact that CFA is computationally more intensive and complex than PCA. But now that computational power is no longer an issue, a number-of-factors technique based explicitly on the common-factor model is feasible and obtainable. Ironically, the rationale behind the MAP test is the most theoretically compatible with the common-factor model (Velicer et al., 2000), yet the test cannot be directly applied to CFA (Piccone, 2009). However, it is possible to incorporate the rationale of a seemingly unrelated procedure known as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (i.e., factorability) with the MAP test to produce an indicator of dimensionality based specifically on the common-factor model. This proposed approach applies the MAP test to a reduced correlation matrix, and the KMO test is then applied to the residual correlation matrix at

each sequential stage of the MAP test; therefore, it is referred to as the sequential KMO (SKMO) procedure.

Research Questions

It is the purpose of this study to formally introduce and evaluate the new SKMO procedure via a Monte Carlo simulation; therefore, the following research questions were constructed in order to achieve these goals. Note that these research questions are each understood to deal with data known to have a common-factor structure. Further, these research questions are to be considered under varying conditions, such as the true number of factors, sample size, variable-to-factor ratio, magnitude of the factor loadings, and degree of inter-factor correlation.

Research Question 1

What are the performance characteristics of the SKMO procedure with respect to the accuracy, bias, and precision of the predicted number of factors to retain?

Research Question 2

How does the SKMO procedure compare to well-known tests of dimensionality (viz., the proportion of variance extracted, Bartlett's chi-square test, the Cattell-Nelson-Gorsuch scree test, the Kaiser-Guttman criterion, parallel analysis, and the minimum average partial test) with respect to the accuracy, precision, and bias of the predicted number of factors to retain?

Assumptions

The general assumptions for this study are the same as those for conducting common-factor analysis (e.g., data are interval or ratio level, relationships are linear; these are discussed in detail in Chapter II). Of particular importance is the assumption that the Pearson product-moment correlation coefficient is an appropriate index of association for the variables and factors. It is also assumed that pseudo-random data as generated by a computer are an acceptable approximation of true random data.

Limitations and Restrictions

Although the component model is discussed in detail, the Monte Carlo simulation for this study deals exclusively with the common-factor model. While CFA is actually a broad category of factor-analytic methods, the present study uses principal-axes factor analysis as the lone representative of CFA. Further, the population loading matrices used in the simulation are all characterized by simple structure; specifically, each observed variable has a complexity of one. It should also be acknowledged that Monte Carlo simulations may not be truly representative of real data and could therefore lead to spurious conclusions (Tucker, Koopman, & Linn, 1969). In particular, this simulation uses random data from a standard-normal distribution even though this was not a prerequisite for the common-factor model.

Definitions of Common Terms

Common Factor: Broadly, a "substantive determining influence" (Cattell, 1966, p. 245). More specifically, a common factor is a special type of latent variable that accounts for some portion of the variances in two or more observed variables and, subsequently, their correlations (Thurstone, 1947). It is also assumed to be generating the observations (Harris, 1975).

Communality: The common variance of an observed variable. More specifically, it is the proportion of variance in an observed variable that can be linked to the set of common factors.

Correlation Coefficient: An indicator of the magnitude and orientation of the linear relationship between two variables. In the present study, all correlations are Pearson product-moment correlations.

Discrepancy: The difference between the number of factors predicted by a dimensionality procedure and the actual number of factors in the population. If a test predicts the correct number of factors, the discrepancy is zero.

Eigenanalysis: The process of computing the eigenvalues and eigenvectors for a square matrix. If the dimensions of the matrix are $p \times p$, there will be p eigenvalues and p eigenvectors.

Eigenvalue: A characteristic root of a square matrix. In factor analysis, each eigenvalue from a correlation matrix represents the sum of the variances accounted for in the set of p observed variables by a given factor.

Eigenvector: A characteristic vector of a square matrix. With the application of a linear transformation, eigenvectors are used to produce the factor loadings.

Iteration: Any recursive process where the output from one step is used as the input for the following step in order to approach a desired outcome. For an example, see *iteration*.

Latent Trait/Construct: An unobservable characteristic or theoretical process that is assumed to be a contributing source of influence on related observable phenomena.

Loading Coefficient: In its most basic form, a loading represents the correlation between a factor and an observed variable. See *Pattern Coefficient* and *Structure Coefficient*.

Monte Carlo Simulation: A class of methods which use computer algorithms to generate large numbers of random datasets that conform to specific parametric conditions. These methods can be used to determine the sampling behavior of a statistic that is otherwise implausible to ascertain analytically.

Oblique: A solution where the factors/components are allowed to be correlated. That is, the inter-factor correlations are not necessarily zero.

Observed Variable: An empirical indicator or measurement, typically designed to tap a specific latent trait. Also known as a *manifest variable* or *indicator variable*.

Orthogonal: 1.) The general condition wherein two (or more) variables are mutually independent and thus uncorrelated. 2.) A solution where the factors are uncorrelated.

Overextraction: The condition where a dimensionality test overestimates the true number of factors. In such cases, the discrepancy will be a positive value.

Pattern Coefficient: A specific type of loading. If the solution is oblique, then the pattern coefficients are the weights for the linear combinations of factors that reproduce the observed variables, analogous to regression coefficients. If the solution is orthogonal, then pattern coefficients are equivalent to structure coefficients and are typically referred to simply as loadings. See also *Structure Coefficient*.

Principal Component: A linear combination of observed variables that has been optimized to capture the maximum amount of variance in the original variables.

Proxy: A surrogate model that is used to approximate the solution from an ideal model. For example, PCA is often used to approximate CFA.

Reduced Correlation Matrix: A correlation matrix with communalities placed on the main diagonal rather than unities.

Residualized Variable: The remaining part of a variable that cannot be predicted by another variable (or set thereof). The residual portion of a variable is orthogonal to the reproduced portion.

Salient: A somewhat subjective term used to denote a sizeable (and hopefully meaningful) loading. Most sources cite a minimum absolute value of .3 to .4 as a general guideline for salience. Note that *salient* is the adjectival form of the noun *salience*.

Saturation: The absolute magnitude of a factor loading.

Structure Coefficient: A specific type of loading. If the solution is oblique, then the structure coefficients are the zero-order correlation coefficients between the factors and observed variables. If the solution is orthogonal, then pattern coefficients are equivalent to structure coefficients and are typically referred to simply as loadings. See also *Pattern Coefficient*.

Underextraction: The condition where a dimensionality test underestimates the true number of factors. In such cases, the discrepancy will be a negative value.

Unique Factor: A latent source of influence assumed to be associated exclusively with a particular observed variable. Unique factors are a part of the common-factor model.

Unreduced Correlation Matrix: A correlation matrix with unities on the main diagonal.

Variable Complexity: The number of salient factor loadings associated with an observed variable.

Notation

The following symbols are used to represent quantities found in samples. Parametric quantities will be introduced as needed per the discussion. This notation uses *i*, *j*, and *k* as index variables.

Scalars

n Sample size

p Number of observed variables

m Number of factors extracted ($m \leq p$)

z_{ik} Score on observed variable *i* ($i = 1, 2, \ldots, p$) for subject *k* ($k = 1, 2, \ldots, n$)

x_{jk} Score on common factor *j* (j = 1, 2, ..., m) for subject *k*

y_{ik} Score on unique factor *i* for subject *k*

$r_{ii'}$ Correlation coefficient for variables *i* and *i*′

$q_{ii'}$ Anti-image correlation coefficient for variables *i* and *i*′

a_{ij} Initial factor loading for variable *i* on factor *j*

b_{ij} Rotated factor loading (pattern) for variable *i* on factor *j*

c_{ij} Rotated factor loading (structure) for variable *i* on factor *j*

d_j Eigenvalue corresponding to factor *j*

h_i^2 Communality for variable *i*

u_i^2 Uniqueness for variable *i*

Vectors and Matrices

R Unreduced correlation matrix ($p \times p$)

$\mathbf{R}^*$ Reduced correlation matrix ($p \times p$)

$\mathbf{Q}$ Anti-image correlation matrix ($p \times p$)

$\mathbf{A}$ Initial common-factor loading matrix ($p \times m$)

$\mathbf{B}$ Rotated common-factor loading (pattern) matrix ($p \times m$)

$\mathbf{C}$ Rotated common-factor loading (structure) matrix ($p \times m$)

$\mathbf{D}$ Diagonal matrix of eigenvalues ($m \times m$)

$\mathbf{H}^2$ Diagonal matrix of communalities ($p \times p$)

$\mathbf{I}$ Identity matrix (square)

$\mathbf{U}$ Diagonal matrix of unique-factor loadings ($p \times p$)

$\mathbf{v}_j$ Eigenvector corresponding to factor j ($p \times 1$)

$\mathbf{V}$ Augmented matrix of eigenvectors ($p \times m$)

$\mathbf{X}$ Common factor-score matrix ($n \times m$)

$\mathbf{Y}$ Unique factor-score matrix ($n \times p$)

$\mathbf{Z}$ Observed-score matrix ($n \times p$)

CHAPTER II

REVIEW OF THE LITERATURE

As the present study focuses on various aspects of the dimensionality question in exploratory factor analysis, this chapter will provide a general overview of the major factor-analytic models and methods as well as an intensive examination of the literature dealing with well-known number-of-factors tests. The first section of this chapter reviews the historical development of factor analysis. The second section serves as a primer for the fundamental concepts and theoretical underpinnings of the most frequently used contemporary factor-analytic models, while the basic procedures for conducting a common-factor analysis are presented in the third section. These first three sections provide a basic foundation that will facilitate the detailed discussion contained in the fourth section regarding six prominent methods for determining the proper number of factors to retain—namely, the proportion of variance extracted, Bartlett's scree test, the Kaiser-Guttman criterion, Cattell's scree test, Horn's parallel analysis, and Velicer's minimum average partial test. The fourth section concludes with a summary of several comprehensive Monte Carlo studies pertaining to the number-of-factors methods.

A Brief History of the Factor-Analytic Methods

Over the past one-hundred years of its existence, factor analysis has undergone a number of evolutionary (and revolutionary) changes to become the modern suite of tools for the analysis of interdependence and latent-variable structure. Two historical developments are particularly noteworthy. First, factor analysis has grown to fully embrace a multidimensional framework; hence, the decision as to the number of factors to extract has become a critical step. Secondly, due

to the growth in efficiency and availability of computers, increasingly sophisticated computational methods have been integrated into the procedures.

The Birth of Factor Analysis

As with any great scientific idea, factor analysis has a number of philosophical and intellectual forerunners; names such as Pythagoras, Plato, Francis Bacon, and Charles Galton have been linked to its academic pedigree (Mulaik, 1987). However, English psychologist Charles Spearman is credited with the introduction of the first coherent and systematic formulation of what would become known as factor analysis (Cudeck, 2007; Thurstone, 1947). These ideas were first published in the paper, "General Intelligence, Objectively Determined and Measured" (1904), which was primarily concerned with Spearman's theory of human intelligence and ability. Factor analysis was essentially a methodological afterthought, appearing on only one page of the lengthy 91-page manuscript (Bartholomew, 2007). Despite its unceremonious presentation, this unifactor[1] model was ultimately the empirical foundation of Spearman's theory of intelligence (Cudeck, 2007). Spearman proposed that human intelligence arises from a single latent source—specifically, a hereditary trait he called the general intelligence factor, or *g*. He further theorized that a person's observed ability, as measured by some given test, was a function of *g* and the unique factor intrinsic to the particular test used. It was this general factor that was postulated to be the underlying cause of the correlations between different tests in a battery.

The value of Spearman's work is immeasurable as it represents the first true incarnation of factor analysis. In particular, there are two significant features found in the unifactor model. First, it introduced what would come to be known as a (latent) common factor that is presumably responsible for the relationship among observed variables; hence, it was the original common-factor model (Rummel, 1970; Mulaik, 2010). A complementary concept is that of the unique factor which accounts for the idiosyncratic variance in a variable—one unique factor for each particular measure. Second, this method utilized Pearson's product-moment correlation coefficients as a

means of uncovering structure between the factors and variables. This, of course, also implies an assumption that the relationships are linear (Bartholomew, 2007). The importance of these features is demonstrated by the fact that these are still characteristics found in contemporary factor-analytic methods.

A Transitional Stage

Many historical accounts of factor analysis overlook the transitional period that bridged the original unifactor model and the multifactor model. Nevertheless, this transitional approach—the bifactor model (Harman, 1976)—is important in that it was the first to acknowledge the need for multiple common factors, although in a very limited capacity. Recognition of this model elucidates the evolution from a simple one-factor framework to the complex structure of multiple common factors.

Many researchers (including Spearman himself) realized that a unifactor model was often insufficient to properly deal with the complex factor structures that frequently resulted from empirical data (Pohlmann, 2004). In 1909, Cyril Burt first questioned the validity of such a simplistic theory of intelligence. Beginning in 1916, Godfrey Thomson published a series of papers that discussed (among other things) the new concept of a group factor (MacCallum, Brown, & Cai, 2007). A group factor refers to a latent source of influence for a specific subset of observed variables—in other words, a restricted form of a common factor (Thurstone, 1947). The group factors are assumed to be uncorrelated with one another, and the variable subsets are typically mutually exclusive (i.e., there are few, if any, cross loadings). Karl Holzinger further developed these ideas in the mid-1930s (Harman, 1976). Holzinger's model retained the single common factor of the unifactor model, but it also incorporated Thomson's group factors. This Thomson-Holzinger hybrid model became known as the bifactor approach since the variance of each variable could be ascribed to two common sources—namely, the single common factor and a group factor.

The historical significance of the bifactor model is understated. Although Holzinger claimed it to be nothing more than a revision of Spearman's work, the bifactor model was, in a sense, the first multifactor model and could therefore handle more complex data (Harman, 1976). In fact, Rummel (1970) refers to it as "a significant shift in theory" (p. 330). Another contribution of this model is rather subtle if unappreciated: The bifactor approach served as an important developmental step en route to the modern multiple-factor model.

The Modern Era of Factor Analysis Begins

Although an important revision of Spearman's model, the bifactor model was essentially a stop-gap procedure to deal with the more complex factor patterns for which the unifactor model was ill-equipped. Holzinger's refusal to completely break with Spearman's theory ultimately limited the scope of the bifactor model (Harman, 1976). A new model and method that allowed for multiple common factors was the next natural step in the evolution of factor analysis. That new multifactor approach is now better known as common-factor analysis (CFA).

Louis Leon Thurstone, an American psychologist, presented the first systematic outline for a true multiple-factor model (and break from Spearman) in his paper, "Multiple Factor Analysis," published in 1931(Bock, 2007). This model was further refined over the next few years in which Thurstone produced a number of articles, monographs, and books, most notably *The Vectors of Mind* (1935). The theory and methods of multiple factors continued to mature over the next decade, culminating in Thurstone's best known book, *Multiple-Factor Analysis: A Development and Expansion of the Vectors of Mind*, published in 1947. Within this body of work he developed the common-factor model as an application of eigenanalysis[2] on a reduced correlation matrix—that is, a correlation matrix with estimated communalities rather than unities on the main diagonal (Gorsuch, 1983; Thurstone, 1947).

While the development of the common-factor model was certainly a monumental accomplishment, it was hardly Thurstone's only achievement. Jones (2007)—a colleague of

Thurstone—reported that Thurstone felt his most important contribution was the idea of rotating a factor solution to achieve simple structure. Aside from these significant accomplishments, there are two often overlooked contributions by Thurstone that are of particular historical interest. First, he can be credited with introducing the now ubiquitous matrix algebra to factor analysis (Bock, 2007). Second, he forced the question as to the proper number of factors to extract from a given set of observed correlations (Harman, 1976).

A Side Story: The Development of Principal-Components Analysis

Related to the evolution of the common-factor model was the creation of a related method known as principal-components analysis (PCA). In 1901, Karl Pearson outlined a technique for transforming a set of variables to a new set of variables (i.e., components) that would preserve the entirety of the variance in the original variables (Mulaik, 1987). In his 1933 article, "Analysis of a Complex of Statistical Variables into Principal Components," Harold Hotelling brought Pearson's original idea to fruition. These concepts were further developed in Truman L. Kelley's 1935 monograph, *Essential Traits of Mental Life* (Harman, 1976). Although superficially similar, CFA and PCA do in fact differ on a number of key issues (these differences are discussed in greater detail in a later section of this chapter). Whereas factor analysis was developed primarily within the field of psychology, PCA was the product of mathematicians and statisticians (Hakstain & Muller, 1973); hence, the two techniques were created for different purposes.[3]

Later Developments and Innovations

One of the most remarkable trends began in the early 1960s when digital computers first became accessible to the educational and psychological research community (Gorsuch, 2003). Before the era of computers, it was impractical in most cases to compute an entire common-factor or principal-component solution by hand due to the prohibitively time-consuming amount of calculations. Rather, researchers opted to use the more expedient approximation techniques such as the centroid and diagonal methods (Harris, 1975). CFA and PCA have since become the most

popular factor-analytic techniques due to the computational efficiency and precision afforded by computers (Harman, 1976).

Ancillary procedures such as factor rotation and communality estimation have subsequently become increasingly sophisticated and numerous. Of particular interest here are the many rules and tests for determining the most appropriate number of factors or components to extract. The rationales for these methods range from inductive to deductive, from simple and intuitive to complex and arcane. With sophistication comes specialization, as some of these dimensionality techniques are based upon the common-factor model, while others are applicable only to the component model.

Contemporary Factor-Analytic Methods: The Principal-Factor Solutions

Virtually all of the modern factor-analytic methods belong to a major class of techniques known as the principal-factor solutions[4] (Gorsuch, 1983). Naturally, all of the methods within this family have a number of key similarities. However, the principal-factor solutions can be categorized according to two distinct modes of classification that are of prime importance. The first type of categorization is ontological in nature as it distinguishes between those techniques that produce factors (CFA) and those that produce components (PCA; Suhr, 2005). For simplicity, the term *factor* will be used in this section to denote both factors (from CFA) and components (from PCA) unless an explicit distinction is made. There is also an epistemological mode of classification that distinguishes between exploratory and confirmatory applications of the principal-factor solutions (Suhr, 2006).

General Features of Principal-Factor Solutions

All of the constituent methods in the family of principal-factor solutions share a number of key features. The most basic characteristic is the use of a single sample of n independent observational units (i.e., subjects), each of which are measured on p observed variables (Rummel, 1970). This will result in np observations that are collectively represented by an $n \times p$ raw-data

matrix **Z**. Denoted by z_{ki}, the element found in row *k*, column *i* of the raw-data matrix represents the observed value of subject *k* on variable *i* ($k = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, p$). For simplicity, an observed variable can be denoted generically by z_i (omitting the *k* subscript) when there is no need to reference any particular subject. The linear relationship between any two of these observed variables is expressed as a zero-order Pearson product-moment correlation coefficient (Everitt, 2006). Typically, the observed variables are assumed to be standardized (i.e., zero mean and unit variance), so a matrix of correlations among these variables can be computed from the raw-data matrix by

$$\mathbf{R} = \frac{1}{n-1}\mathbf{Z}'\mathbf{Z} \tag{2.1}$$

(Thurstone, 1947). The observed correlation matrix **R** is a $p \times p$ symmetric matrix composed of the correlations of the observed variables. The element in row *i*, column *i*′ (denoted by $r_{ii'}$) is the correlation between variables *i* and *i*′.

There is also a set of *m* unobserved variables, better known as factors. Initially, there will be as many potential factors as there are observed variables. Since parsimony is a central issue in factor analysis, only a subset of these initial factors will be retained; hence, $m < p$ (Thurstone, 1947). Conceptually, each of the *n* subjects will have a score on each of the *m* factors, although it will often be the case that these values are not actually needed (Gorsuch, 1983). Like the observed variables, the factors are also assumed to be standardized. This array of *nm* values can be arranged in an $n \times m$ matrix **X**, where x_{kj} (the element found in row *k*, column *j*) represents the score for subject *k* on factor *j*. The *k* subscript can be dropped when specifying a factor without reference to a particular subject. As a matter of convenience, the factors are initially derived under the condition of mutual orthogonality—that is, factor *j* is uncorrelated with factor *j*′ for all *j*, *j*′ ($j \neq j'$; Mulaik, 2010). If warranted, this constraint can be removed in a subsequent stage of the analysis so that a correlated-factor (oblique) solution may be obtained by a variety of rotational algorithms (Thurstone, 1947).

The defining feature of the principal-factor solutions is the use of eigenanalysis on some form of an observed correlation matrix in order to extract the initial set of factors (Gorsuch, 1983). Eigenanalysis is the method of calculating the characteristic roots and vectors (otherwise known as eigenvalues and eigenvectors) of a square matrix (Fraleigh & Beauregard, 1995). This process will produce p eigenvalues (d_j) and p corresponding p-dimensional eigenvectors ($\mathbf{v}_j$). In a principal-factor solution, an eigenvalue represents the sum of the variances across all p variables that can be accounted for by the factor corresponding to that eigenvalue (Harman, 1976). The first factor is extracted so that the sum of the variances associated with it is maximized. Under the condition that it will be orthogonal to the first, the second factor is extracted in a likewise manner so that it is maximized with respect to the remaining variance in the observed variables that is not associated with the first factor. The third factor will be maximized given that it is orthogonal with the first two, and so on. Hence, the overall amount of the variance extracted from the observed variables is maximized under the imposed condition of factor orthogonality (Anderson, 2003). Note that $d_j \geq d_{j+1}$ for any pair of adjacent eigenvalues because the total variance in the system is fixed, and the factors are maximized at each step of the extraction process (Mulaik, 2010).

The relationships of the factors with the observed variables are the true focus of a factor-analytic study (Thurstone, 1947). Consequently, the centerpiece of any principal-factor solution is the set of factor loadings (Harman, 1976). A loading is defined as the zero-order Pearson product-moment correlation between an observed variable and a factor, and it is an indicator of the direction and strength of the linear relationship between that variable and factor (Gorsuch, 2003). As there are p variables and m factors, there will be pm loadings. The set of factor loadings can be expressed as a $p \times m$ matrix $\mathbf{A}_m$ where the element found in row i, column j (denoted by a_{ij}) represents the correlation of observed variable i and factor j—that is, the loading of variable i on factor j. A simple linear transformation of an orthonormal set of eigenvectors (the so-called principal axes of a symmetric matrix) produces the matrix of factor loadings

$$\mathbf{A} = \mathbf{V}\mathbf{D}^{1/2}, \tag{2.2}$$

where $\mathbf{D} = \text{diag}\left(d_1, d_2, \ldots, d_p\right)$ is a $p \times p$ diagonal matrix with the ordered set of eigenvalues on the main diagonal, and $\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_p]$ is a $p \times p$ augmented matrix composed of an ordered set of corresponding eigenvectors (Rummel, 1970). Note that Equation 2.2 produces the complete $p \times p$ loading matrix **A**. Since there are typically $m < p$ factors retained, the truncated $p \times m$ loading matrix $\mathbf{A}_m$ can be obtained by simply omitting the last $p - m$ columns of **A**. Because of the conditions imposed during eigenanalysis, all factor loadings derived in this manner will be in the closed interval [-1, 1] (Harman, 1976).

All of the principal-factor solutions are based on a linear model. In fact, the principal-factor solutions fall under the larger domain of the generalized linear model, the same family as linear regression (Gorsuch, 2003; Mulaik, 2010). One of the key assumptions in factor analysis is that each observed variable can be expressed as a linear function of the factors (Gorsuch, 1983). Using the factor loadings as weights, the linear model predicts (or reproduces) each of the *p* observed variables from the *m* factors:

$$\hat{z}_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m \quad (i = 1, 2, \ldots, p) \tag{2.3}$$

or, in matrix notation,

$$\hat{\mathbf{Z}} = \mathbf{X}\mathbf{A}'_m, \tag{2.4}$$

where $\hat{\mathbf{Z}}$ is the $n \times p$ reproduced raw-data matrix. Recall that the loadings represent correlations between factors and variables, and the factors are uncorrelated. Given this, and because the variables and factors are assumed to be standardized, the loadings are equivalent to standardized regression weights (Mulaik, 2010). In this way, the model can be viewed as a special case of linear regression (Thurstone, 1947). The special circumstance apparent in this method is that the predictors (i.e., the factors) are not known a priori.

The previously discussed formula for the complete loading matrix **A** (Equation 2.2) is a consequence of the fact that the observed correlation matrix is symmetric. All symmetric matrices

can be diagonalized, and a convenient pair of matrices to utilize for the diagonalization of **R** is the aforementioned **D** and **V** (Fraleigh & Beauregard, 1995). The use of **V** is justified since a set of eigenvectors from a symmetric matrix is linearly independent, which guarantees that **V** is nonsingular. Hence, **R** can be expressed in the canonical form $\mathbf{V}^{-1}\mathbf{R}\mathbf{V} = \mathbf{D}$. Note that the set of eigenvectors from a symmetric matrix can always be selected in such a manner that **V** is an orthogonal matrix (or, more correctly, any set of eigenvectors extracted from a symmetric matrix can be orthonormalized). Therefore, there exists a matrix **V** such that $\mathbf{V}' = \mathbf{V}^{-1}$, which then allows

$$\begin{aligned} \mathbf{R} &= \mathbf{V}\mathbf{D}\mathbf{V}' \\ &= \left(\mathbf{V}\mathbf{D}^{1/2}\right)\left(\mathbf{D}^{1/2}\mathbf{V}'\right) \\ &= \left(\mathbf{V}\mathbf{D}^{1/2}\right)\left(\mathbf{V}\mathbf{D}^{1/2}\right)' \\ &= \mathbf{A}\mathbf{A}' \end{aligned} \tag{2.5}$$

(Harman, 1976). As demonstrated in Equation 2.5, using the full $p \times p$ loading matrix **A** will reproduce **R** perfectly since none of the factors have been omitted. Using the $p \times m$ truncated loading matrix $\mathbf{A}_m$, however, will result in a reproduced covariance matrix $\hat{\mathbf{S}} = \mathbf{A}_m\mathbf{A}_m'$, where $\hat{\mathbf{S}}$ approximates but does not equal **R**. The reproduced covariance matrix $\hat{\mathbf{S}}$ is an attempt to reconstruct the observed correlation matrix **R** using the derived solution in $\mathbf{A}_m$ (Harris, 1975). The disparity between the observed and the reproduced matrices is represented by the residual covariance matrix

$$\mathbf{S}_{res} = \mathbf{R} - \hat{\mathbf{S}} \tag{2.6}$$

(Tabachnick & Fidell, 2001). It has become the convention to refer to these two matrices as the reproduced correlation matrix and the residual correlation matrix, respectively, although they are in fact covariance matrices (Thurstone, 1947).

Communality is a fundamental concept of the principal-factor methods. Communality (denoted by h_i^2) is defined as the proportion of variance in a given observed variable that is linearly related to the set of *m* factors (Pohlmann, 2004), with each of the *p* variables having its own

communality value. Suppose that variable *i* is correlated with a set of standardized orthogonal predictor variables—namely, the set of *m* extracted factors. Since all the factors are mutually orthogonal, the amount of variance in z_i that can be accounted for by the set of factors (the communality for variable *i*) can be expressed as

$$h_i^2 = \sum_{j=1}^{m} a_{ij}^2 \tag{2.7}$$

(Thurstone, 1947). Equivalently, the set of communalities are the elements on the main diagonal of the reproduced covariance matrix $\hat{\mathbf{S}}$. Note that communality is in fact a special instance of the squared multiple correlation (SMC) coefficient from linear regression, where z_i is the criterion variable and $x_1, x_2, \ldots, x_m$ are the predictors (Harman, 1976).

An Ontological Distinction: Factors versus Components

The principal-factor solutions are subdivided into two general forms: common-factor analysis[5] (CFA) and principal-components analysis (PCA). Ultimately, the source of the differences between these two models rests with their respective ontological assumptions regarding the directionality of influence. The common-factor model assumes that a factor is representative of a real yet unobservable construct or trait that affects the observed variables. In contrast, components are defined as nothing more than linear combinations of the observed variables, and the meaning of the concepts represented (if any) are emergent from the variables (Rummel, 1970). Of course, these opposing ontological characteristics of CFA and PCA unavoidably lead to various operational differences. A brief summary of the most important distinguishing features is presented in Table 2.1.

Table 2.1

Fundamental Differences between CFA and PCA

Feature	CFA	PCA
Ontological orientation	Reflective	Formative
Error term	Uniqueness	None
Values on main diagonal	Communalities	Unities
Factor/component scores	Estimated	Calculated

Common-factor analysis. CFA is an exercise in conceptual parsimony as its primary aim is to detect and delineate latent constructs (Everitt, 2006). The fundamental assumption is that the observed variables are reflective—that is, the observed variables are being influenced by an underlying set of latent variables (Kline, 2005). As its name suggests, CFA targets a particular type of latent trait known as a common factor, which is defined as a latent variable that has an effect on more than one observed variable (Gorsuch, 1983). Because common factors will simultaneously affect multiple variables, they are assumed to be the underlying causes of the correlations among the observed variables (Costello & Osborne, 2005; Harman, 1976). For this reason, CFA operates exclusively on the variance that can be linked to the common factors (i.e., communality).

In the common-factor model, the residual portion of the variance of an observed variable is attributed to a unique factor (Gorsuch, 1983). A unique factor (y_i) is exclusive to a particular observed variable. As there are *p* variables, there will also be *p* unique factors. The unique factors are assumed to be standardized and mutually orthogonal to one another as well as to the set of common factors (Gorsuch, 2003). With the use of unique factor loadings (u_i), the unique factors are incorporated into the linear model to give the common-factor model,

$$z_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m + u_i y_i, \tag{2.8}$$

or, in matrix notation,

$$\mathbf{Z} = \mathbf{X}\mathbf{A}'_m + \mathbf{Y}\mathbf{U} \tag{2.9}$$

where $\mathbf{Y}$ is a $n \times p$ matrix of unique factors and $\mathbf{U} = \text{diag}(u_1, u_2, \ldots, u_p)$. The portion of an observed variable's variance that is attributed to the unique factor is known as uniqueness (u_i^2). Equivalently, uniqueness is the proportion of variance in a variable that cannot be linked to the common factors (much like the residual error in linear regression). Note that $h_i^2 + u_i^2 = 1$ since each observed variable has unit variance (Harman, 1976). The linear model from Equation 2.9 also gives rise to another important equation,

$$\mathbf{R} = \mathbf{A}_m \mathbf{A}_m' + \mathbf{U}^2 \tag{2.10}$$

(proof omitted). The relationship expressed in Equation 2.10 is better known as the Fundamental Theorem of Factor Analysis (Rummel, 1970).

As CFA is a search for common factors, a basic principle is that the uniqueness associated with each observed variable should be removed from the system before extracting the common factors, thereby assuring that only the communalities of the variables will remain in the system (Mulaik, 2010). This is achieved by estimating the communalities and placing those values in their respective position on the main diagonal of the correlation matrix. What results from this operation is the reduced correlation matrix, denoted by $\mathbf{R}^*$, where the *i*th diagonal element is h_i^2, and all off-diagonal elements remain unchanged. CFA is conducted by extracting factors from the reduced correlation matrix $\mathbf{R}^*$ rather than the observed correlation matrix $\mathbf{R}$. This is the key procedural difference between CFA and PCA (Gorsuch, 1983).

A phenomenon known as factor-score indeterminacy is known to be associated with CFA (Stevens, 2002). A consequence of dealing with latent traits is that they are, by definition, unobservable; therefore, common factors can only be approximated. The term *approximated* is used not only to reflect that factors are unobservable, but also to indicate that the common-factor model is characterized by a mathematical indeterminacy. For each of the *m* common factors, there are an infinite number of alternative factors that will leave the derived values in the factor-loading matrix $\mathbf{A}_m$ unchanged. Such a scenario results from the inclusion of the *p* unique factors, which

presents an underdetermined system with $m + p$ unknowns, but only p equations. A consistent system of equations with more unknowns than equations will necessarily have an infinite number of solutions (Fraleigh & Beauregard, 1995). The potential pitfall is that all of these alternative factors may not be highly correlated among themselves, which would introduce ambiguity regarding the meaning (construct validity) of the proposed factor (Mulaik, 2010). However, such concerns are embedded within the context of a single analysis; it does not consider the potential replicability of factors across samples.

Principal-components analysis. In contrast to CFA, PCA is an exercise in empirical parsimony (Everitt, 2006). The primary function of PCA is to transform a set of observed variables into a new, often smaller set of orthogonal variables—called components—which will retain most (if not all) of the total original variance (Suhr, 2005). A component is defined simply as a linear combination of the observed variables under the condition that the amount of variance captured by each mutually orthogonal component is maximized (Tabachnick & Fidell, 2001). In this fashion, the components emerge from the observed variables; in other words, the observed variables are formative (Kline, 2005). Consequently, the principal-components model makes no assumption about the existence of any latent structure (Costello & Osborne, 2005).

Where CFA operates only on the common variance that is shared among the observed variables, PCA operates on the total variance (Widaman, 1993). Provided that each of the observed variables has unit variance, components are extracted from the correlation matrix **R** with unities on the main diagonal since the objective is to create a set of components that retains as much of the variance of the observed variables as possible (Suhr, 2005). Parenthetically, a correlation matrix with ones on the diagonal is also known as an unreduced correlation matrix (Guttman, 1954). This stands in contrast to CFA which removes the residual variance by placing estimated communalities on the main diagonal.

Although trivial components (those that capture little of the original variance) are often discarded, PCA does not include an explicit error term in the model (Pohlmann, 2004). In doing so, PCA bypasses the potential problem of factor-score indeterminacy (Rummel, 1970). Given that there are no unique factors, there are p equations with $m \leq p$ unknowns. An overdetermined ($m < p$) or just-determined ($m = p$) consistent system of equations will always have a unique solution (Fraleigh & Beauregard, 1995), and therefore no indeterminacy.

As opposed to the explanatory and conceptual roles of CFA, PCA is better suited to serve in a taxonomic capacity. It provides a convenient method to categorize variables and summarize observed data, but it is inadequate as an explanatory tool (Floyd & Widaman, 1995; Hakstain & Muller, 1973). PCA can be used in conjunction with other analytical methods, such as multiple-regression analysis (MRA) and multivariate analysis of variance (MANOVA), as a preliminary step to treat or summarize data under certain conditions (Harris, 1975). For example, PCA has often been used to deal with the problem of multicollinearity in MRA. If two or more of the predictor variables are highly correlated, the entire set of predictors can be transformed to a mutually orthogonal set of new predictors that preserves the total variance. PCA can also be used to condense an impractically large and unwieldy set of observed variables to a smaller, more manageable set, with only a minimal loss of variance (Anderson, 2003). A drawback is that the components are not necessarily substantively meaningful (Suhr, 2005).

PCA as an approximation of CFA. Although contrary to the assumptions of both models, it is not uncommon for PCA to be used as a surrogate for CFA (Mulaik, 2010). This practice began in the early 1960s when computers first became accessible to the general academic community. These machines possessed very limited processing resources, so in most cases computing a full common-factor solution was simply not feasible since the initial estimation of the communalities was prohibitively time consuming (Gorsuch, 2003). However, PCA has no such initial

computations; it could therefore serve as a compromise that would roughly approximate CFA, but without the burdensome initial communality estimates (Costello & Osborne, 2005).

Despite the ubiquity of this practice, the soundness of using PCA as an approximation for CFA has often been questioned. From a theoretical perspective, the component model should not be confused with the common-factor model (Humphreys & Ilgen, 1969). Although computationally similar, these two methods were developed for clearly different purposes. Further, CFA is more realistic because the model incorporates error and is, therefore, better suited for situations where unique variance is clearly present (i.e., moderate to low communalities; Fabrigar et al., 1999). Accordingly, the common-factor model should be applied in the social sciences when exploring latent structure because of the large amounts of error variance inherent in such research. In addition, CFA attempts to minimize that error when making estimations (Widaman, 1993). CFA also has the advantage in replicability and generalization (Floyd & Widaman, 1995). Where the results of CFA are generalizable to other sets of variables, the generalization of a PCA solution outside those particular variables of a given study is dubious (Mulaik, 1990; Widaman, 1993). The readily available power and efficiency of modern computers has served as a catalyst to further the debate. Since computing resources are no longer a limitation, there is no reason to rely on an expedient proxy rather than applying the most appropriate model (McArdle, 1990).

Although PCA has been touted as an accurate approximation of CFA, there are cases where these two methods will generate markedly divergent solutions. In general, PCA has a tendency to produce spuriously large estimates of the loading coefficients due to capitalization on chance, especially with smaller sample sizes (Comrey, 1978). Conversely, CFA has been shown to be unbiased in the estimation of loadings (Gorsuch, 1990). By extension, PCA will also tend to produce inflated communality estimates as well (Costello & Osborne, 2005). This becomes particularly evident when the number of observed variables is small to moderate. A Monte Carlo study by Snook and Gorsuch (1989) showed that the solutions for PCA and CFA only begin to converge when

there are more than 40 variables; otherwise, PCA will produce inflated estimates of the loadings. Later studies (e.g., Widaman, 1993) suggest that the variable-to-factor ratio also has an impact on estimate bias. Finally, it has been shown that when relatively low communalities are present—particularly with item analysis—CFA is much more accurate (Floyd & Widaman, 1995). However, the differences between CFA and PCA vanish as the communalities approach 1.0 (Gorsuch, 1983).

An Epistemological Distinction: Exploratory versus Confirmatory Factor Analysis

The epistemological orientation of a particular factor-analytic study can be either exploratory or confirmatory; however, this mode of classification does not exist as a pure dichotomy, but rather as a continuum (Jöreskog, 2007; Stevens, 2002). Just as the name implies, exploratory factor analysis is an inductive approach that requires little or no substantive theoretical background (Gorsuch, 1983). It is essentially a class of heuristic methods that are used for theory building, uncovering latent structure, and classification. In contrast, confirmatory factor analysis is a deductive, algorithmic approach (Suhr, 2006). These are theory-testing methods that are designed to assess the fit of a pre-existing theoretical framework to empirical data (Kline, 2005). This distinction is presented here only for completeness as the remainder of this discussion deals exclusively with exploratory factor analysis.

General Procedures in Common-Factor Analysis

The present study focuses on the common-factor model; specifically, principal-axes factor analysis—the prototypical form of CFA—is the method of interest. The following is a brief outline of the basic procedures that are necessary for conducting a typical exploratory CFA.

Preliminary Conditions and Assumptions

Before beginning an exploratory CFA, a number of preliminary conditions and assumptions should be met. First, it is assumed that the proper epistemological orientation is exploratory, and the proper ontological orientation is reflective. Although not prerequisites for CFA, it will be assumed for the purposes of the following discussion that an appropriate set of variables has been

selected, and there are no missing data. Although not a prerequisite, it is customary for all of the observed variables to be standardized (Harman, 1976). A crucial yet often overlooked assumption is that the Pearson product-moment correlation coefficient is an appropriate measure of association for the given data (Rummel, 1970). This implies two things: Each observed variable has a true quantitative, numerical metric (i.e., interval or ratio scales), and the relationships among the p observed variables are all assumed to be linear. Further, it is also assumed that the variables are linearly related to the factors (Thurstone, 1947).

The issue of sample size in factor analysis remains a particularly contentious topic. Nevertheless, CFA is a large-sample technique, so it is obviously preferable to have more subjects (Costello & Osborne, 2005). An often-cited guideline is that the minimum ratio of the number of subjects to the number of observed variables be no less than five (Floyd & Widaman, 1995; Gorsuch, 1983). So, if it is known that p variables will be used in the analysis, then at least $5p$ subjects are needed. But more recent research has indicated that absolute sample size is, in fact, more important (Stevens, 2002). While some authors recommend as few as 100 subjects, others suggest that 300 is a more sensible minimum sample size (e.g., Tabachnick & Fidell, 2001).

Multivariate normality of the set of observed variables is not a de jure assumption in CFA (Gorsuch, 1983; Rummel, 1970). However, there are some inferential tests for determining the number of factors to extract that do require the variables to be multinormal. Although no formal distributional assumptions are present, outliers are to be avoided (Tabachnick & Fidell, 2001). The Pearson product-moment correlation coefficient is known to be acutely sensitive to extreme values.

Factorability of the Correlation Matrix

Given that the various preliminary conditions are tenable, a $p \times p$ zero-order correlation matrix $\mathbf{R}$ is computed from the sample data. Since the procedure is exploratory, a natural first step is to assess the factorability of $\mathbf{R}$. That is, some evidence of structure among the observed variables is needed to justify proceeding with the analysis. Such a step is warranted as various studies have

shown that interpretable yet baseless results have been produced from completely random data (Gorsuch, 1983). A cursory inspection of the zero-order correlation coefficients among the p variables provides a useful (if informal) indication of the factorability of **R**. CFA may not be an appropriate approach if none of the correlations have an absolute value greater than .30 (Tabachnick & Fidell, 2001).

Bartlett's (1950) test of sphericity is an inferential statistic used to assess the factorability of **R**. This statistic tests the null hypothesis that the population correlation matrix is equal to an identity matrix. However, this test is known to be extremely sensitive and should be used with only relatively small samples (Tabachnick & Fidell, 2001). For this reason, Dziuban and Shirkey (1974) propose using this statistic as a conservative indicator of factorability. While a significant test is still dubious, a correlation matrix with a non-significant test should certainly not be subjected to factor analysis.

Kaiser (1970, 1981) viewed factorability as a psychometric issue, opting to use the term *sampling adequacy* to reflect the importance of sampling the right set of variables in order to detect any meaningful underlying structure. He based his work in this area on that of Guttman, who had shown that the closer the off-diagonal elements of an anti-image correlation matrix were to zero, the stronger the evidence that the data have a common-factor structure (Cerny & Kaiser, 1977; Kaiser, 1970). The anti-image of a variable is the residual portion of that variable remaining after removing the variance that can be associated with the other variables in the set (Gorsuch, 1983). The correlation between the anti-images of two observed variables (from a set of p) can conveniently be defined as the opposite value of the corresponding $(p-2)$th-order partial correlation (Rummel, 1970). Consequently, if there is in fact at least one common factor underlying a set of observed variables, then the anti-image correlations will be relatively small in absolute magnitude as compared to the zero-order correlations. Using this property, Kaiser sought to create

a non-inferential, psychometrically based indicator of factorability. What resulted from this line of thought is now better known as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy.[6]

The KMO procedure itself is straightforward, albeit computationally intensive. The inverse of the observed correlation matrix ($\mathbf{R}^{-1}$) must be computed, which, in turn, is used to produce the anti-image correlation matrix

$$\mathbf{Q} = \left[(\text{diag}\ \mathbf{R}^{-1})^{-1/2}\right]\mathbf{R}^{-1}\left[(\text{diag}\ \mathbf{R}^{-1})^{-1/2}\right] \tag{2.11}$$

where diag $\mathbf{R}^{-1}$ represents the diagonal matrix that is constructed by setting all off-diagonal elements of the inverse correlation matrix to zero. The KMO test value is computed by

$$K = \frac{\text{trace}(\mathbf{R}^2) - p}{\text{trace}(\mathbf{R}^2) + \text{trace}(\mathbf{Q}^2) - 2p}, \tag{2.12}$$

or equivalently,

$$K = \frac{\sum_i \sum_{i'} r_{ii'}^2}{\sum_i \sum_{i'} r_{ii'}^2 + \sum_i \sum_{i'} q_{ii'}^2} \quad (i \neq i') \tag{2.13}$$

where $q_{ii'}$ is the element from row i, column i' of $\mathbf{Q}$ (Kaiser & Rice, 1974). The purpose of putting these quantities in this particular formula was to ensure that the KMO would be restricted to the closed interval [0, 1].

Once the KMO value is calculated, it must then be compared to some criterion. Some sources (e.g., Mulaik, 2010; Tabachnick & Fidell, 2001) suggest that a computed KMO value greater than or equal to .60 is indicative of factorability—that is, if $K \geq .60$, then there is sufficient evidence that there exists at least one common factor underlying the observed variables. However, .60 is an arbitrary value which has been given without explicit justification. In fact, Kaiser originally recommended that the baseline criterion for factorability should be .50 because a KMO of that value would imply $\sum_i \sum_{i'} r_{ii'}^2 = \sum_i \sum_{i'} q_{ii'}^2$ (Cerny & Kaiser, 1970; Kaiser, 1981; Kaiser & Rice, 1974). Since an anti-image correlation can be thought of as a correlation between two observed variables after stripping away any linear influence from the common factors, then the equality of the mean of the squared observed correlations and the mean of the squared anti-image correlations implies that

there is no common-factor influence to be removed. Ergo, there would be no common factors relevant to the given set of observed variables. Furthermore, the use of .50 also has an empirical foundation. Shirkey and Dziuban (1976) explored the sampling characteristics of the KMO statistic by generating samples from populations where each correlation matrix was equal to an identity matrix. These populations were multinormal, and the sampling conditions were systematically varied with respect to sample size and number of observed variables. Across all conditions in this simulation, the minimum value was .470, and the maximum was .529, with the KMO becoming increasingly stable as the sample size increased. The mean value of the KMO across all conditions was .499.

Number of Factors to Extract

Perhaps the most crucial step in CFA concerns the proper number of factors to extract for a given dataset (Coovert & McNelis, 1988; Hayton, Allen, & Scarpello, 2004). As CFA is essentially an exercise in parsimony, the number of factors to extract is ideally fewer that the number of observed variables (Harris, 1975). Yet this parsimony must be tempered with plausibility. Although a solution with only a few factors is sought, there must still be enough factors present in the model to reasonably account for the observed correlations among the variables (Fabrigar et al., 1999). Hence, overextraction and underextraction can both have adverse effects on a solution (Fava & Velicer, 1992, 1996). However, it has been shown in numerous Monte Carlo studies that extracting too few factors is by far the more harmful scenario because it has the potential to distort those factors that have been retained (Gorsuch, 2003).

There are many methods for determining the most appropriate number of factors to extract for a given observed dataset, ranging from the statistical to the purely visual (the most popular approaches are discussed in greater detail in the following section). Regardless, the true criteria for the proper number of factors are ultimately replicability (consistent results with other samples of

subjects from the same population) and invariance (consistent results with other samples of variables from the same domain; Crocker & Algina, 1986).

Communality Estimation

Before computing the factor solution, an initial set of estimated communalities must be determined (Gorsuch, 1983). There are several methods available for estimating communality. A simple early method used the greatest absolute correlation associated with each observed variable (Thurstone, 1947). However, the greatest correlation method is not recommended for studies with small numbers of variables (Harman, 1976).

Another communality-estimation technique uses the squared multiple correlations (SMC) for each observed variable with the other $p - 1$ variables acting as predictors. This method is easy to implement, and it typically performs quite well (Comrey, 1978; Gorsuch, 2003). Mulaik (2010) argued that the SMC is the best systematic estimate for communality by showing that in the population the SMC for a variable (when predicted by the $p - 1$ other variables) is the greatest (or strongest) lower bound for the communality of that variable. So, the SMC is a conservative estimate of communality.

Iteration by refactoring is an ostensibly objective method for determining the initial communalities that appears to provide the most accurate solution overall (Russell, 2002; Widaman, 1993). Although computationally intensive, this estimation method is easily implemented with the use of modern statistical software such as SPSS.[7] This process begins with a set of estimated communalities (usually SMCs) placed on the main diagonal of **R**. The specified number of factors (m, as determined in the previous step) is extracted, and a new set of communality estimates is computed from that solution. These new communalities are then entered into their respective positions on the main diagonal of **R**, and the process is repeated until the communalities converge to a set of stable values. That is, refactoring stops when the maximum change in the communalities from one iteration to the next is less than some convergence criterion value (SPSS uses .001 as the

default value). An ancillary criterion may be set to limit the number of iterations allowed (SPSS uses 50 as the default value). Either of the aforementioned estimation methods (or even unities) can be used to provide a set of baseline estimates, but this ultimately makes little difference since the estimates typically converge on the same values regardless of the baseline estimates used (Floyd & Widaman, 1995). However, starting with SMCs almost always provides the fastest convergence (Harris, 1975; Rummel, 1970).

Initial Solution

Once the initial communality estimates have been calculated, the values are placed in the appropriate positions on the main diagonal of **R**—thus creating the reduced correlation matrix **R***—and an initial solution of *m* mutually orthogonal factors is extracted. Eigenanalysis is performed on the reduced correlation matrix which ensures that the solution will produce the maximal sum of squared loadings for each orthogonal factor. Note that for the initial solution (i.e., before rotating the factors) the sum of squared loadings associated with a factor is equal to its eigenvalue (Mulaik, 2010).

Final communalities. Assessment of the communalities at this stage (i.e., before rotation) is acceptable since the communalities are invariant with respect to rotation of the factor solution (Tabachnick & Fidell, 2001). Similar to the initial communality estimates, the final communalities are also SMCs. In this case, however, a communality coefficient represents the proportion of variance in an observed variable linked to the set of factors (Harman, 1976). Accordingly, the final communalities indicate how well the variables are defined by the factor solution (Rummel, 1970). Numerous sources (e.g., Fabrigar et al., 1999; Gorsuch, 1983; Stevens, 2002) classify communality estimates of .7 and above as high, while values below .4 are considered low.

Low communalities have a few potential causes. Low reliability for an observed variable can stifle communality since the sample reliability is the upper bound of communality (Mulaik, 2010). It is also possible that the variables with small communalities are outlier variables—that is,

they are unrelated to the other variables and the domain of theoretical interest. It is for this reason that the investigator should always assess the face validity of the variables to be used in a factor-analytic study. Finally, low communality estimates may simply be the result of extracting too few factors (Fabrigar et al., 1999).

Residual covariance matrix. The residual matrix for a given solution is calculated as the difference between the observed correlation and the reproduced covariance matrices (see Equation 2.5). Recall that the off-diagonal elements of both the reproduced and residual matrices are in fact covariances, not correlations. Ideally, the residual covariances will be small, which would indicate that the proposed factor loadings can closely reproduce the observed correlation matrix, and no additional factors will be needed (Thurstone, 1947). However, if there are numerous moderate residuals (.05 to .10) or even a small number of large residuals (greater than .10), extraction of at least one additional factor may be necessary (Tabachnick & Fidell, 2001). One of the primary goals of factor analysis is to find the smallest number of factors that contains the greatest amount of information. Clearly, the adequacy of the fit of a solution is directly related to the number of factors extracted. More factors will naturally improve the fit between the observed and reproduced correlation matrices, but at the cost of parsimony.

Factor Rotation and Interpretation

The ultimate objective of CFA is to reveal and understand the meaning of the extracted factors, which is accomplished through a systematic examination of the factor loadings (Thurstone, 1947). Unless the solution consists of only a single factor, the initial solution itself is rarely interpreted. The loadings are reallocated through a linear transformation so that certain properties of the solution are retained (namely, the communalities and reproduced correlations) while improving the overall substantive quality and interpretability (Harman, 1976). That is, the factors are rotated[8] to a new frame of reference in order to make the solution more meaningful with regard to the observed variables. Although the final communalities will remain unchanged, the rotated

solution will virtually always change the sum of squared loadings for each factor. Consequently, the loss of a maximized solution is the price of improved interpretability (Harris, 1975).

There are two basic classes of rotational solutions: orthogonal and oblique. Orthogonal solutions assume that the factors are uncorrelated, where oblique (or unrestricted) solutions allow for possible inter-factor correlations (Gorsuch, 2003). Note that the latter does not require factors to be correlated, it only allows for the possibility of this condition. Rotation is achieved by applying a linear transformation to the initial solution. The derivation of the rotational transformation is beyond the scope of this study and is therefore omitted (see Gorsuch, 1983, for an excellent discussion on this topic).

Where orthogonal solutions have only one loading matrix, two types of loading matrices are involved with oblique solutions. Recall that factor analysis is a special form of multiple regression where the factors serve as the predictor variables. When the predictors are uncorrelated, then each standardized regression coefficient is equivalent to the corresponding zero-order correlation between the predictor and the criterion variable; conversely, when predictors are correlated, this equivalence no longer holds. For this reason, oblique solutions transforms the initial loading matrix ($\mathbf{A}_m$) into two types of special loading matrices—namely, a pattern matrix (**B**) and a structure matrix (**C**). The $p \times m$ pattern matrix contains the regression weights for reproducing the observed variables using the new factors as predictors. The structure matrix (also $p \times m$) contains the zero-order correlations between the observed variables and the new factors. Oblique solutions also produce an $m \times m$ symmetric matrix of inter-factor correlations (**F**). Note that with an orthogonal solution, $\mathbf{B} = \mathbf{C}$ and $\mathbf{F} = \mathbf{I}$. Equation 2.10 can be modified to incorporate the pattern matrix **B** and the inter-factor correlation matrix **F**:

$$\mathbf{R} = \mathbf{BFB}' + \mathbf{U}^2. \tag{2.14}$$

Because completely independent phenomena are a rarity in the social and behavioral sciences, oblique solutions are a particularly important feature of the factor-analytic methods.

Arbitrarily imposing an orthogonal solution when an oblique is more appropriate would distort information and give misleading results (Costello & Osborne, 2005; Fabrigar et al., 1999). It is therefore advisable to rotate factors to a preliminary unrestricted solution; if the underlying factors are indeed orthogonal, it would be revealed in the results (Russell, 2002).

Interpretation of the factors naturally requires some familiarity with the domain of interest (Stevens, 2002). This knowledge, coupled with an inspection of which observed variables have large (or salient) loadings on a factor—and, just as importantly, which variables do not—provides insight into the nature of the factor. Tabachnick and Fidell (2001) aptly described the interpretation and naming of factors as "a process that involves art as well as science" (p. 625).

Determining the Appropriate Number of Factors to Extract

Horn and Engstrom (1979) recounted a personal communication with Henry Kaiser: He once quipped that finding "a solution to the number-of-factors problem in factor analysis is easy, that he used to make one up every morning before breakfast" (p. 283). The challenging part, of course, is discovering *the* solution. Indeed, there has been a host of procedures invented for the purpose of divining the proper number of factors to extract from a correlation matrix. Yet, an inspection of the literature reveals that only a small handful of those methods are used with any regularity (Mumford, Ferron, Hines, Hogarty, & Kromrey, 2003). Some of the better known number-of-factor tests are the proportion of variance extracted, Bartlett's chi-square test, the Kaiser-Guttman criterion, Cattell's scree test, Horn's parallel analysis, and Velicer's minimum average partial correlation test. The procedures, rationales, advantages, and disadvantages for these methods will be discussed in detail in the following section. Because the specific dimensionality test used by an investigator should be theoretically compatible with the general analytical approach (i.e., CFA or PCA; Hakstain & Muller, 1973; Humphreys & Ilgen, 1969; Velicer, Eaton, & Fava, 2000), a particularly important theme intertwined in the discussion is the model

with which each method is theoretically aligned—specifically, the common-factor model or the component model.

Proportion of Variance Extracted

Although it is known to be one of the original methods for determining dimensionality, the exact origins of the proportion-of-variance criterion are unclear. However, an early reference to this approach can be found in Thurstone (1947). Regardless, the proportion-of-variance criterion is certainly one of the oldest approaches to the question of dimensionality that has any contemporary relevance (Gorsuch, 1983).

Procedure. This method is based on the component model, so the procedure begins by extracting the ordered set of all p eigenvalues from the unreduced correlation matrix **R** (Velicer et al., 2000). Recall that an eigenvalue (d_j) is equal to the sum of the squared loadings for a given component from the initial (unrotated) maximized solution. It therefore represents the amount of variance in the set of all p variables that can be linked to the given component. Since the components are extracted from the correlation matrix with unities on the diagonal, total variance refers to the sum of the variances of all p observed variables. Assuming that the variables are all standardized, the total variance will then equal p, the number of variables. So, the proportion of variance extracted for a solution consisting of the first t components is then

$$P_t = \frac{1}{p}\sum_{j=1}^{t} d_j \tag{2.15}$$

(Mulaik, 2010). This quantity is calculated for each possible number of components to retain—that is, there will be a P_t value for $t = 1, 2, \ldots, p$. After these cumulative percentages have been computed—with each percentage value representing the proportion of total variance accounted for by all the components extracted up to that point—the number of components to extract is indicated at the point where a predetermined percentage has been achieved (Hakstain & Muller, 1973). The particular cutoff value is fairly arbitrary, but many sources suggest using at least 75% as the

criterion (e.g., Gorsuch, 1983; Harman, 1976). So, for example, if $P_{t-1} < .75$ and $P_t \geq .75$, then $m = t$.

Rationale. The logic behind the proportion-of-variance criterion is straightforward as it operates solely on the idea of practical significance (Harman, 1976). The basic premise is that only those components that account for a substantial proportion of the observed variance will be of any substantive or practical value; the remaining components are trivial. The challenge is, of course, deciding what amount of variance constitutes a substantial proportion.

Advantages and disadvantages. This method is simple and intuitive, and it represents a clear and practical outcome. Still, the arbitrary nature of the criterion is a troublesome limitation: What may be considered an inconsequential amount of variance in one study may not be so trivial in another (Coovert & McMelis, 1988). The subjectivity in determining a cutoff value is made apparent by the wide range of recommended values found in the literature (Velicer et al., 2000). Since this criterion directly indicates the amount of variance accounted for in the set of observed variables, it is tempting for an investigator to retain a greater number of factors, but these additional factors may be meaningless.

Bartlett's Chi-Square Test

Another classic dimensionality test is the inferential statistic introduced by Bartlett (1950, 1951). This technique is based on the principal-components model (Gorsuch, 1973; Linn, 1968) and should be used only as a proxy indicator if applied to CFA (Harman, 1976). This test is logically analogous to another well-known inferential number-of-factors test by Lawley, which was designed for use with maximum-likelihood factor analysis (Bartlett, 1950; Horn & Engstrom, 1979). Only Bartlett's test will be presented here to avoid redundancy.

Procedure. Bartlett's test is conducted as a sequence of up to $p - 1$ null-hypothesis statistical tests. As this is an inferential statistic, the parametric assumption of multivariate normality must first be tenable (Hong et al., 2006). If normality can be assumed, then the p

eigenvalues are extracted from the unreduced correlation matrix. The order of extraction should be maintained as they will be extracted in order of decreasing magnitude ($d_1 \geq d_2 \geq \cdots \geq d_p$). Now let *t* be the number of potential components to retain at any given step ($t = 0, 1, \ldots, p-2$). Each of the null hypotheses tests for the equality of the last $p - t$ population eigenvalues (δ_j). These null hypotheses are of the form

$$H_0^{(t)}: \delta_{t+1} = \delta_{t+2} = \cdots = \delta_p \tag{2.16}$$

and are tested in sequential order beginning with $t = 0$. The proper number of components to retain is indicated at the first failure to reject a null hypothesis (Gorsuch, 1973). So, if there is a failure to reject the *t*th null hypothesis, then $m = t$. For each value of *t*, the statistic

$$B_t = -\left[n - 1 - \tfrac{1}{6}(2p + 5) - \tfrac{2}{3}t\right] \log W_t \tag{2.17}$$

is computed, where

$$W_t = \frac{(d_{t+1})(d_{t+2}) \ldots (d_p)}{\left(\frac{d_{t+1} + d_{t+2} + \cdots + d_p}{p - t}\right)^{p-t}} \tag{2.18}$$

(Bartlett, 1950, 1951). This statistic is approximately distributed as a chi-square random variable with $(p - t - 1)(p - t + 2)/2$ degrees of freedom. Bartlett used the usual nominal significance level of .05 for evaluating the statistic. Since PCA is a large-sample technique, Horn and Engstrom (1979) proposed the use of a conservative adjustment to the nominal significance level (.001, for example) in order to correct for the increased power associated with larger samples.

Rationale. Bartlett's test is based upon the premise that after extracting the appropriate number of components, the remaining eigenvalues are due to random error. Random error is expected to be fairly homogeneous, so each random-error component should account for the same amount of variance in the set of observed variables (Horn & Engstrom, 1979). Failure to reject the *t*th null hypothesis implies that when *t* components have been extracted, each of the remaining $p - t$ components account for roughly the same amount of variance in the set of observed variables, and this homogeneity is interpreted as an indication of random error. Interestingly, Mulaik (2010)

noted that although it can be directly applied only to PCA, the basic rationale for Bartlett's test is more consistent with the common-factor model.

The derivation of the formula given by Equations 2.17 and 2.18 is omitted here because it is quite lengthy and provides little additional insight to the rationale behind this technique. It is worth noting that Bartlett's number-of-components test is very closely related to Bartlett's test of sphericity (a measure of factorability; Gorsuch, 1983). In fact, the test performed at $t = 0$ is precisely this factorability test (Bartlett, 1950). This provides a different way to view the rationale for Bartlett's number-of-components test: The tth test in the sequence can be thought of as a test of factorability given that the first t components have been removed from the system.

Advantages and disadvantages. It is ironic that Bartlett's method requires a large sample in order to ensure that the test statistic is asymptotically distributed as a chi-square random variable because such a large sample very often makes the test much too sensitive to be of any practical use (Hong et al., 2006). With larger samples, smaller differences among the roots are deemed statistically significant; yet, it is often the case that they are not substantively significant (Gorsuch, 1973). Bartlett (1950) himself admitted that statistical significance does not necessarily imply substantive significance, as the test may lead to the extraction of meaningless or transient components. Hence, overextraction is an imminent concern with the Bartlett test. When used with PCA as a proxy for CFA, Bartlett's test should be considered only as an upper bound for the number of common factors (Gorsuch, 1973, 1983).

Zwick and Velicer (1982) offered a tentative justification for the use of the Bartlett test: While an increase of sample size does lead to greater power, it also leads to greater accuracy (which should lead to smaller differences among the trivial roots). Hence, these two currents may offset one another. This position was subsequently abandoned (Zwick & Velicer, 1986), although Zwick and Velicer noted that the Bartlett chi square may remain useful in situations where minor

components are not so readily dismissed, such as the preliminary phases of exploratory research in a new domain.

Kaiser- Guttman Criterion

Based on the work of Guttman (1954), Kaiser introduced a simple and immensely popular number-of-factors rule in 1960. A subsequent publication in 1961 fully elaborated upon the use of a well-known theorem by Guttman in the creation of this rule. Due to the simplicity of its application—as well as the intuitive appeal of one of its theoretical rationales—the Kaiser-Guttman criterion is one of the most commonly used number-of-factors rules (Rummel, 1970; Stevens, 2002). This method has a number of aliases, such as the eigenvalue-greater-than-one rule, the Kaiser-One (or K1) rule, and Guttman's weakest lower bound. It is the default number-of-factors rule in many software packages, including SPSS (SPSS, Inc., 2009).

Procedure. Application of the Kaiser-Guttman criterion is quite simple. The ordered set of all p eigenvalues is extracted from the unreduced correlation matrix **R**. The lower bound for the number of factors to retain (not necessarily *the* number of factors) is indicated by the number of eigenvalues that are greater than unity (Gorsuch, 1983).

Since this procedure uses the eigenvalues from an unreduced correlation matrix, it would appear that it was intended for PCA. However, as will be shown, the justification for this method was originally based on the common-factor model. Nevertheless, many authors (e.g., Everitt, 2006; Mulaik, 2010; Russell, 2002; Velicer et al., 2000; Yeomans & Golder, 1982) consider the Kaiser-Guttman rule to be more theoretically congruent with the component model and therefore appropriate for use with PCA regardless of the original rationale.

Rationale. Guttman's (1954) focus was on the relationship between the various lower bounds for the number of common factors and the type of communality estimation used rather than seeking a rule for determining the number of factors per se. However, Kaiser (1960) believed that Guttman's lower bounds held promise as indicators of dimensionality. The third (or strongest)

lower bound for the number of factors is the number of positive eigenvalues from the correlation matrix with SMCs on the main diagonal. Despite its potential, Kaiser (1960) found that the strongest lower bound consistently extracted far too many factors (almost always more than $p/2$ factors). This motivated the use of the first (or weakest) lower bound (i.e., the number of eigenvalues of **R** greater than unity), which was found to provide much more reasonable results.

Despite the uncertain theoretical orientation of the Kaiser-Guttman rule, the original rationale for this procedure is indeed focused solely on the common-factor model (Hakstain & Muller, 1973). It makes use of Thurstone's (1947) assertion that in the *population*, the proper number of common factors is equal to the rank of the reduced correlation matrix, defined as $\mathbf{R}^* = \mathbf{R} - \mathbf{I} + \mathbf{H}^2$, where **R** is the correlation matrix, **I** is an identity matrix, and $\mathbf{H}^2$ is the diagonal matrix of the true (yet unknown) communalities. Thurstone postulated that in the population the reduced correlation matrix should be Gramian: $\mathbf{R}^* = \mathbf{A}_m\mathbf{A}_m'$. That is, **R*** is the product of the loading matrix and its transpose. Recall that the order of $\mathbf{A}_m$ is $p \times m$. Assuming $m < p$, then m is the rank of $\mathbf{A}_m$. Because the ranks of a matrix and its corresponding Gramian product are equal (Fraleigh & Beauregard, 1995), m is also the rank of **R***.

In his classic work, Guttman (1954) introduced a particularly useful lemma given here without proof: Let **G** be a Gramian matrix, and let $\mathbf{T}_1$ be a diagonal matrix such that $\mathbf{G}_1 = \mathbf{G} - \mathbf{T}_1$ is also Gramian. Now let $\mathbf{T}_2$ be a diagonal matrix such that every diagonal element of $\mathbf{T}_2$ is greater than the corresponding diagonal element of $\mathbf{T}_1$, and let $\mathbf{G}_2 = \mathbf{G} - \mathbf{T}_2$. If the scalar g_1 represents the rank of $\mathbf{G}_1$, and g_2 represents the non-negative index of $\mathbf{G}_2$ (i.e., the number of non-negative eigenvalues of $\mathbf{G}_2$), then $g_1 \geq g_2$.

Using Thurstone's (1947) conjecture along with this lemma, Guttman gave the following proof for the weakest lower bound for the rank of **R*** in the population: The population correlation matrix **R** is a Gramian matrix. Let $\mathbf{U}^2 = \mathbf{I} - \mathbf{H}^2$ be a diagonal matrix of unique variances such that the reduced correlation matrix $\mathbf{R}^* = \mathbf{R} - \mathbf{U}^2$ is Gramian and of minimum rank, and let $\mathbf{R}_0 = \mathbf{R} - \mathbf{I}$

(i.e., a correlation matrix with zeros on the main diagonal). Now let the scalar m be the rank of $\mathbf{R}^*$ and the scalar m_0 be the positive index for $\mathbf{R}_0$ (i.e., the number of positive eigenvalues for $\mathbf{R}_0$). Assuming all the communalities are greater than zero, then $u_i^2 < 1$ for all i. By the previously given lemma, m is then greater than or equal to the non-negative index of $\mathbf{R}_0$. But the positive index of a matrix will always be less than or equal to the non-negative index, so $m \geq m_0$. Now let $\mathbf{V}$ be an orthogonal eigenvector matrix and $\mathbf{D}$ be the diagonal matrix of eigenvalues for $\mathbf{R}$. Since

$$\begin{aligned} \mathbf{V}'\mathbf{R}_0\mathbf{V} &= \mathbf{V}'(\mathbf{R} - \mathbf{I})\mathbf{V} \\ &= \mathbf{V}'\mathbf{R}\mathbf{V} - \mathbf{V}'\mathbf{V} \\ &= \mathbf{D} - \mathbf{I} \end{aligned} \tag{2.19}$$

(see Equation 2.6) and $\mathbf{D} - \mathbf{I}$ is diagonal, $\mathbf{V}$ and $\mathbf{D} - \mathbf{I}$ are by definition the eigenvector and eigenvalue matrices (respectively) for $\mathbf{R}_0$. This further implies that m_0 (the positive index for $\mathbf{R}_0$) also represents the number of eigenvalues for $\mathbf{R}$ that are greater than unity. Hence, the number of eigenvalues for the population correlation matrix $\mathbf{R}$ that are greater than unity is a lower bound for the rank of the reduced correlation matrix $\mathbf{R}^*$, and thus a lower bound for the number of common factors by Thurstone's (1947) model.

Kaiser (1970) later offered a somewhat different justification for the eigenvalue-greater-than-one rule that is more intuitively appealing than the original rationale (Velicer et al., 2000; Yeomans & Golder, 1982). Given a set of observed variables, the expected value of the sum of squared loadings associated with some random factor is equal to 1.0. Recall that an eigenvalue represents the total amount of variance in the observed variables that is linearly related to a factor. Those factors that account for more variance than this baseline value are more likely to be meaningful and should therefore be retained. In other words, if a factor accounts for more variance than that contained in a single variable (which has unit variance), then it is by definition a common factor (Gorsuch, 2003). Interestingly, Mulaik (2010) speculated that the justification for this rule

was ultimately Kaiser's own personal experiences using it with PCA rather than a mathematical or psychometric principle.

There was actually a third justification proposed for the Kaiser-Guttman rule. It was thought that any principal component with a corresponding eigenvalue greater than one would necessarily have positive Kuder-Richardson (internal consistency) reliability (Kaiser, 1960). However, this was later shown to be incorrect by Cliff (1988).

Advantages and disadvantages. As confirmed by numerous Monte Carlo studies (e.g., Linn, 1968; Velicer et al., 2000; Zwick & Velicer, 1982, 1986), the Kaiser-Guttman method is considered to be one of the poorest number-of-factor tests (Costello & Osborne, 2005). Overextraction is a particularly troublesome drawback that plagues both CFA and PCA when using the Kaiser-Guttman criterion (Comrey, 1978; Fabrigar et al., 1999; Zwick & Velicer, 1982, 1986), especially with larger sets of observed variables (Russell, 2002). The Kaiser-Guttman criterion has also been criticized for being rather arbitrary and inflexible (O'Connor, 2000; Turner, 1998). Consider, for example, two sequential factors with eigenvalues of 1.03 and .97. There is practically no meaningful difference in the amounts of variance accounted for, yet one factor will be retained while the other will not. Such an interpretation implies the realistically untenable assumption that the standard errors of each sample eigenvalue is virtually zero (Lambert, Wildt, & Durand, 1990).

Nevertheless, there are rare endorsements of the Kaiser-Guttman rule, although they are usually contingent upon a rather restrictive set of preconditions. For example, Hakstian, Rogers and Cattell (1982) claimed that the Kaiser-Guttman criterion performs better in scenarios where the sample is relatively large ($n > 250$) and the mean communality is .60 or greater. Stevens (2002) claimed that the test performs better when the number of variables is less than 30 and each of the communalities is greater than .70.

Cattell's Scree Test

One of the more whimsically named techniques, the scree test (Cattell, 1966), was devised as a purely visual procedure. The scree test differs from other procedures in that it is a type of heuristic (rather than algorithmic) approach that provides a method for researchers to visually differentiate the factors that are ostensibly worthy of retention from those that are trivial. It has become one of the more widely used methods for determining the number of factors (Mumford et al., 2003). Although historically the Kaiser-Guttman criterion has been more popular, the savvy researcher will often prefer the scree test because it is more adaptable with respect to the idiosyncratic nature of sample data (Zoski & Jurs, 1996). This procedure was originally designed for the common-factor model (Cattell, 1966), yet it is still applicable in PCA (Velicer et al., 2000). For obvious reasons, the scree test is also referred to as the visual scree test and the subjective scree test in textbooks and research literature.

Procedures. The scree test begins with the ordered set of all p eigenvalues that were extracted from the unreduced correlation matrix **R**. These values are then plotted on a two-dimensional Cartesian graph with the ordinal positions on the horizontal axis and the numerical values of the eigenvalues on the vertical axis (Figure 2.1). The ordinal position refers to the order of extraction, so the first eigenvalue will be the largest, the second will be the next largest, and so on. This will produce a graph of a sequence of points that appears to be monotonically decreasing and converging on zero. With most real datasets the graph will begin with each successive eigenvalue showing a marked decrease in magnitude. This trend will eventually give way to a sequence of eigenvalues in a relatively straight line with a slight downward slope (Cattell & Vogelmann, 1977). Incidentally, the distinctive appearance of this graph resembles the profile of a mountain jutting upward out of a field of scree (the accumulation of rubble and rock debris at the base of a mountain), which is the basis of the metaphorical name. The basic premise is that the eigenvalues of the trivial factors conform to the scree line, while those of the non-trivial factors will

show a pronounced separation from such a uniform trend. The critical step then is to identify the break (or "elbow") of the graph—that is, the point at which the graph separates into a region of sharp decline and a region of a gradually decreasing linear trend (Hong et al., 2006). Cattell (1966) described the scree line as beginning "at a definite point, but smoothly, as does a tangent leaving a circle" (p. 268). Rather than looking for the most prominent break in the graph explicitly, a strategy more apropos to Cattell's original idea is to look for the longest, most stable line of eigenvalues with a slight angle to the horizontal axis (Hakstian et al., 1982). Retaining those factors associated with the points strictly to the left and above the scree line generally provided the best solution (Stellefson & Hanik, 2008; Zoski & Jurs, 1996). Using the graph presented in Figure 2.1 as an example, the most stable scree line would be that running from the fourth position to the last, which suggests that three factors should be extracted.

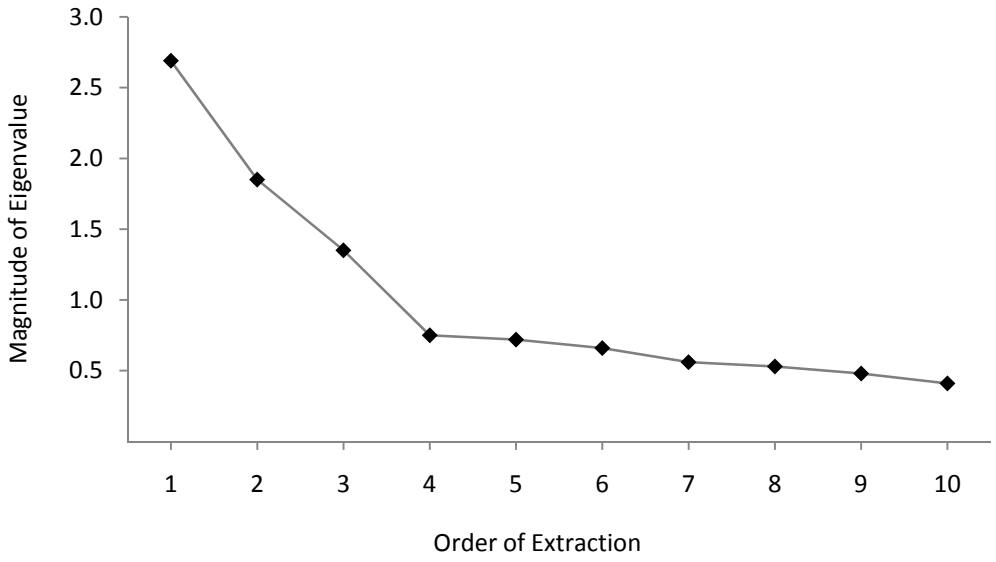

Figure 2.1. Example of a basic scree plot.

Although it is ideal to have only a single scree line, there are occasions when the data will produce multiple breaks in the graph—that is, two or more scree lines in a tiered progression. The

standard rule is to use the numerically lowest number of factors suggested by the various scree lines (Cattell, 1966). In the example shown in Figure 2.2, there are two apparent breaks; the first scree line indicates four factors, while the second points to thirteen. The given rule suggests a four-factor solution. However, this is essentially an atheoretical rule, so it should not be blindly applied without reference to the domain of interest. Pre-existing theory or prior research should be consulted as a guide when possible (Nasser et al., 2002). When there are different numbers of factors suggested by a scree test, they are typically so disparate that such a priori knowledge (if not common sense) should indicate which values are reasonable.

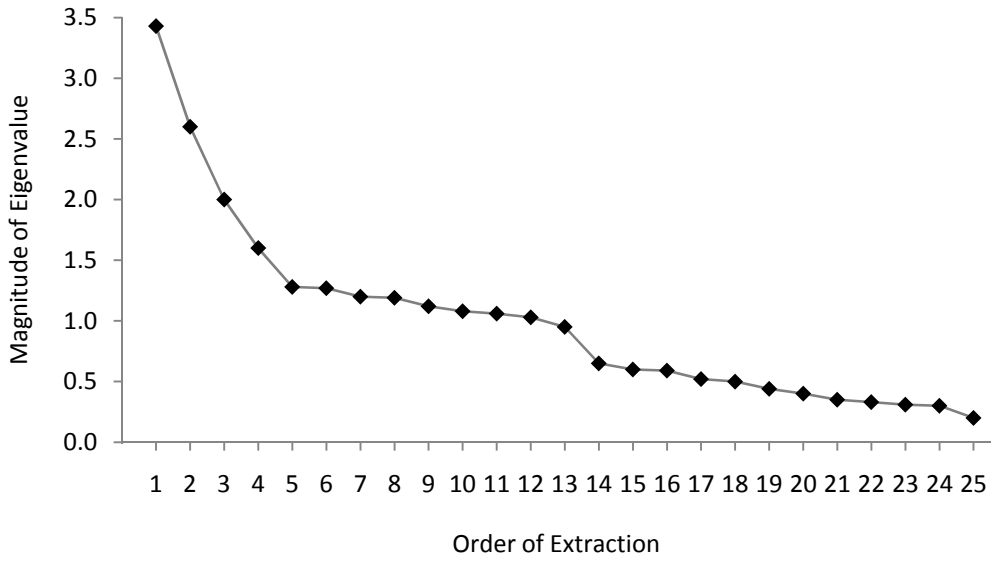

Figure 2.2. Example of a scree plot with two break points.

Given that human judgment plays a key role in this process, consistency in the application of the scree test relies on some level of familiarity and proficiency. Acknowledging the need for training, Cattell (1966) remarked that "even the test as simple as this requires the acquisition of *some* art in administering it" (p. 256). In order to provide a basic level of training, Cattell and

Vogelmann (1977) have offered 15 examples with answers, from which Zoski and Jurs (1993) compiled into a set of basic guidelines for the proper use of the scree test.

Objective versions of the scree test. Because of the inherently subjective nature of the scree procedure, a number of objective (if mechanical) versions of this test have been proposed (Mumford et al., 2003). Such a procedure obviates the training which is required for proper use of the subjective scree, and it can also be implemented using a computer algorithm—a very useful characteristic when performing a Monte Carlo study.

Nelson and Gorsuch (1981; as cited in Gorsuch, 1983) developed an objective, regression-based procedure known as the Cattell-Nelson-Gorsuch (CNG) scree test. In essence, this is an attempt to quantitatively define the elbow from the visual scree test. This method is performed by regressing eigenvalues on the ordinal positions from the scree plot, three adjacent points at a time. The process begins by subtracting the slope computed from the eigenvalues at ordinal positions 1, 2, and 3 from the slope computed from the eigenvalues at ordinal positions 4, 5, and 6. This is repeated for points 2, 3, 4 and 5, 6, 7, and so on. In general, this can be described as a sequence of $p - 5$ values, each determined by the difference between the slope from points $t - 2$, $t - 1$, and t and the slope from points $t + 1$, $t + 2$, and $t + 3$, where $t = 3, \dots, p - 3$. The number of factors is indicated at the point t that has produced the greatest difference in slopes.

However, the CNG has been criticized for two notable shortcomings. First, it uses only six eigenvalues at each step of the procedure, so not all available information is being used (Nasser et al., 2002). Second, since the smallest number of factors the CNG can detect is three, this procedure is not appropriate if a one- or two-factor solution is plausible for a given dataset (Zoski & Jurs, 1993).

Rationale. Cattell's scree test is unique among the number-of-factors tests presented here in that it was developed inductively (Cattell & Vogelmann, 1977). Rather than beginning with an abstract mathematical theorem or theoretical model, Cattell (1966) devised the scree test after

noticing a particular pattern emerge when applying factor analysis to real datasets. This pattern, of course, was that the number of factors which was deemed to be correct or the most appropriate almost always matched the break in the graph of eigenvalues. Further, it appeared this phenomenon was not specific to any particular field of study, as it was seen to be consistent in both the social and natural sciences. Subsequent studies using plasmodes gave additional support to Cattell's conjecture.

As the scree test was developed empirically and inductively, it was not explicitly based upon a pre-existing theoretical system. Nevertheless, inductive reasoning is a theory-building process, so a theoretical framework was formulated as an explanation for the scree phenomenon. Such a theory is not necessary for the practical application of the procedure, however. Cattell (1958, 1966) and Cattell and Vogelmann (1977) develop the theoretical framework by expanding upon the basic postulates of the common-factor model. The theory holds that in a sample there are three major types of common factors in play: systematic factors, incidental factors, and common-error factors. Systematic factors are substantive, scientifically meaningful dimensions that have an influence on some subset of the observed variables. These reflect the constructs that are the focus of a given study. The incidental factors are those influences that arise from instrumental and methodological errors or otherwise transient causes. These are indeed systematic yet have no real scientific meaning within the context of a study as they are sample specific and will not replicate. The concept of a common-error factor is seemingly contradictory to the axioms of the common-factor model. While the common-factor model does specify that errors are uncorrelated in the population, sampling and measurement errors will lead to some chance correlations among the error terms in the sample (what Thurstone, 1947, called *experimental dependence*). That is, given that any two error terms are independent in the population, the corresponding sample correlation will vary within an interval centered at zero. Thus, the two error terms will have some shared variance in the sample, which will in turn give the appearance of a common-error factor.

A fundamental tenet in the theory holds that there are far more factors than there are observed variables for any given study (Cattell, 1966). Given the existence of these numerous (possibly infinite) latent sources of influence, there is no such thing as the "true" number of factors underlying a set of observed variables in this framework. So, a more practical objective would be to determine which of the factors are the most meaningful and important within the context of the study. This theory operates on the presumption that the systematic factors will account for a markedly greater proportion of the variance in the observed variables, and this will in turn be recognizable as an emergent trend in the eigenvalue plot—that is, a break from the scree line. Those factors that are deemed trivial, on the other hand, should correspond to the common-error factors. Since the common-error factors are the product of the random error introduced by sampling and measurement, they will be fairly uniform in the amount of variance they account for in the observed variables, thereby producing a relatively steady line of eigenvalues in the graph. The slight downward slope of the scree line is then an artifact of the variance-maximizing extraction procedure (namely, eigenanalysis).

This theory also accounts for the possibility of spurious results. Naturally, there is the possibility of common-error factors with eigenvalues on the same order of magnitude as the systematic factors. In addition, it cannot be assumed that incidental factors will always behave the same as the trivial terms. The scree test will, by design, always place factors that have the appearance of substantive factors (by virtue of inordinately large eigenvalues) in the non-trivial set (Cattell, 1966). Conversely, systematic factors can show characteristics of random error, which will lead to their being misclassified as trivial. It is the nature of exploratory research to encounter unexpected common (systematic) factors. Often, the observed variables in a study are selected purposely to ensure those substantive factors that are anticipated will account for relatively greater amounts of variance; however, the unexpected substantive factors will have eigenvalues with

noticeably lesser magnitudes (Horn & Engstrom, 1979). The latter type of factor can fall close enough to the scree line to be misidentified.

Cattell and Vogelmann (1977) admit that the theoretical cause of multiple scree lines is still rather unclear. However, the working theory proposes that each distinct scree line is a homogeneous cluster of trivial factors. For example, the first scree line could contain the smaller common-error factors, while the subsequent scree may hold only incidental factors.

Cattell's (1966) justifications for the use of the unreduced correlation matrix (despite this test being designed for use with the common-factor model) are of a practical nature. First, it is simply a matter of convenience to use unities rather than communalities in the preliminary steps of a factor analysis. More importantly, Cattell felt that the additional random error left in the system by virtue of unities in the main diagonal introduce a stabilizing and homogenizing effect that would help delineate the scree line. However, Tucker, Koopman, and Linn (1969) reported that the scree test works just as well and indicates the same number of factors when using communalities in the main diagonal.

Advantages and disadvantages. One of the most often-cited limitations of the scree test is its inherent subjectivity (Kaiser, 1970; Mumford et al., 2003). Turner (1998) described the interpretation of a scree plot as "more an art than a science" (p. 542). With such visual techniques there are naturally concerns that researchers will "[see] what they want to see in the data" (Zoski & Jurs, 1993, p. 5). The issue of subjectivity becomes truly problematic when faced with a dataset that produces multiple breaks or a graph with such a gradual change in slope that it lacks any discernible break (Stellefson & Hanik, 2008; Velicer et al., 2000; Zwick & Velicer, 1982). Further, since the investigator is a key element of the procedure, a certain amount of familiarity and training are required in order for the scree test to be an effective and reliable tool (Zoski & Jurs, 1996). All of these concerns with subjectivity inevitably lead to issues with the inter-rater reliability of the test (Zwick & Velicer, 1986). As a practical matter, there is an inherent limitation with this method

as it is a heuristic algorithm, and therefore cannot be implemented with computer software (Gorsuch, 1983).

Nevertheless, the scree test should not be immediately discounted simply because of its subjective nature; in fact, this subjectivity can also be counted as an asset in certain situations. For example, the scree test has been described as a graphical analog of the Bartlett's chi-square test since both seek to identify and eliminate a set of factors that account for equal amounts of variance (Lambert et al., 1990). Due to the flexibility afforded by its visual approach, the scree test can often eliminate the excessive factors that would be retained by Bartlett's test and the Kaiser-Guttman criterion as those factors will often lack visual prominence (Velicer et al., 2000).

Parallel Analysis

Conceived by Horn (1965) as a sample-based correction for the Kaiser-Guttman criterion, parallel analysis is one of the more promising number-of-factors methods (Gorsuch, 2003). Paradoxically, it is also one of the most underutilized (Hayton et al., 2004). This is in part due to the fact that this method is not a native procedure in any of the major statistical software packages (namely, SPSS and SAS), which makes its implementation potentially more difficult. Fortunately, O'Connor (2000) has published parallel-analysis command programs written for various software packages, including SPSS.

Procedure. Much like the Kaiser-Guttman criterion and the scree test, basic parallel analysis begins by extracting the ordered sequence of p eigenvalues from the unreduced correlation matrix (Horn, 1965). Again, suppose that the observed correlation matrix has been derived from observed data for n subjects on p variables. Now let there be p sets of n observations randomly drawn from a normal distribution—that is, an $n \times p$ random raw-data matrix is generated. The idea is to generate data that are like (or parallel to) the observed data in every possible respect except that they lack an underlying factorial structure. A correlation matrix is then computed from this randomly generated raw-data matrix, followed by eigenanalysis. This process of generating

random parallel matrices and extracting eigenvalues is replicated an arbitrary number of times, and the mean of the eigenvalues at each ordinal position is calculated. The factors corresponding to observed eigenvalues that are greater than their respective mean random eigenvalues are retained.

A number of modifications on the basic procedure have been proposed.[9] Although Horn (1965) developed parallel analysis for use with the common-factor model, Humphreys and Ilgen (1969) suggested using communality estimates such as SMCs rather than unities in the main diagonal of the correlation matrices. Arguing that the means of the random eigenvalues are too liberal to be used as cutoff values, Longman, Cota, Holden, and Fekken (1989) and Glorfeld (1995) proposed using the 95th percentile of the random eigenvalues as the selection criterion. Buja and Eyuboglu (1992) suggested using column-wise permutations of the raw data matrix in order to match potential distributional peculiarities (e.g., non-normal skewness and kurtosis) rather than generating random data from an unrealistically pristine normal distribution. However, it would seem that parallel analysis is quite robust to distributional inconsistencies between the real and random data. In fact, there is no appreciable loss in accuracy when using random parallel datasets generated from a simple uniform distribution (Glorfeld, 1995).

There is no set rule for the number of random parallel matrices to generate. When using the mean random eigenvalue as the criterion, some sources (e.g., Gorsuch, 2003; Hayton et al., 2004) recommend a fairly large number of replications (50 or more) should be used. Yet Velicer et al. (2000) claimed that very few replications (even just one) will suffice without any appreciable loss in accuracy. Of course, if the 95th percentile were to be used as the criterion, thousands of replications are required in order to generate an empirical distribution. Given the efficiency of modern computers, a few thousand replications is not an unreasonable amount (Glorfeld, 1995; O'Connor, 2000).

Rationale. The reasoning behind the Kaiser-Guttman criterion is theoretically sound in the population, but in finite samples the combination of sampling error and capitalization of chance will

consistently inflate eigenvalues (Gorsuch, 2003; Zwick & Velicer, 1986). Recognizing this limitation, Horn (1965) developed parallel analysis as a sample-based procedure to apply the rationale of Kaiser-Guttman test. In particular, parallel analysis attempts to model the bias that is absent in Kaiser and Guttman's method. Since determining the distributions of the eigenvalues analytically is not feasible, a Monte Carlo simulation is a logical alternative.

Horn (1965) presented the following scenario: Consider a Cartesian-coordinate graph similar to that constructed for the scree test (i.e., ordinal number on the abscissa, magnitude of eigenvalue on the ordinate) with the four curves depicted in Figure 2.3. Curve P_{T} represents the eigenvalues of the target population that arise from a set of p observed variables. Curve S_{T} shows the eigenvalues for a given empirical sample of n subjects from the target population. Note that S_{T} roughly mimics the path of the curve from the target population, yet it is consistently higher on the graph with respect to the vertical axis. This difference (e_j) between the two curves represents the inflation introduced by sampling error and capitalization on chance. Now consider P_{N}, an eigenvalue curve that comes from a population with no underlying factorial structure, and all p observed variables are mutually independent (i.e., a null population). Such a curve would be a perfectly straight, horizontal line that intersects the vertical axis at 1.0. Coincidentally, this line also represents the cutoff value for the Kaiser-Guttman rule. Let there also be a curve S_{N} for the mean eigenvalues from a set of random samples of n subjects from this null population. Since sampling bias and capitalization on chance will be present here as well, the random curve will be somewhat higher than the null curve with respect to the vertical axis. For the sake of notational simplicity and correspondence with the graph, the four curve designators will also be used to represent functions that map ordinal positions to eigenvalue magnitudes—for example, $S_{\mathrm{T}}(j)$ represents the eigenvalue from the target sample at position j.

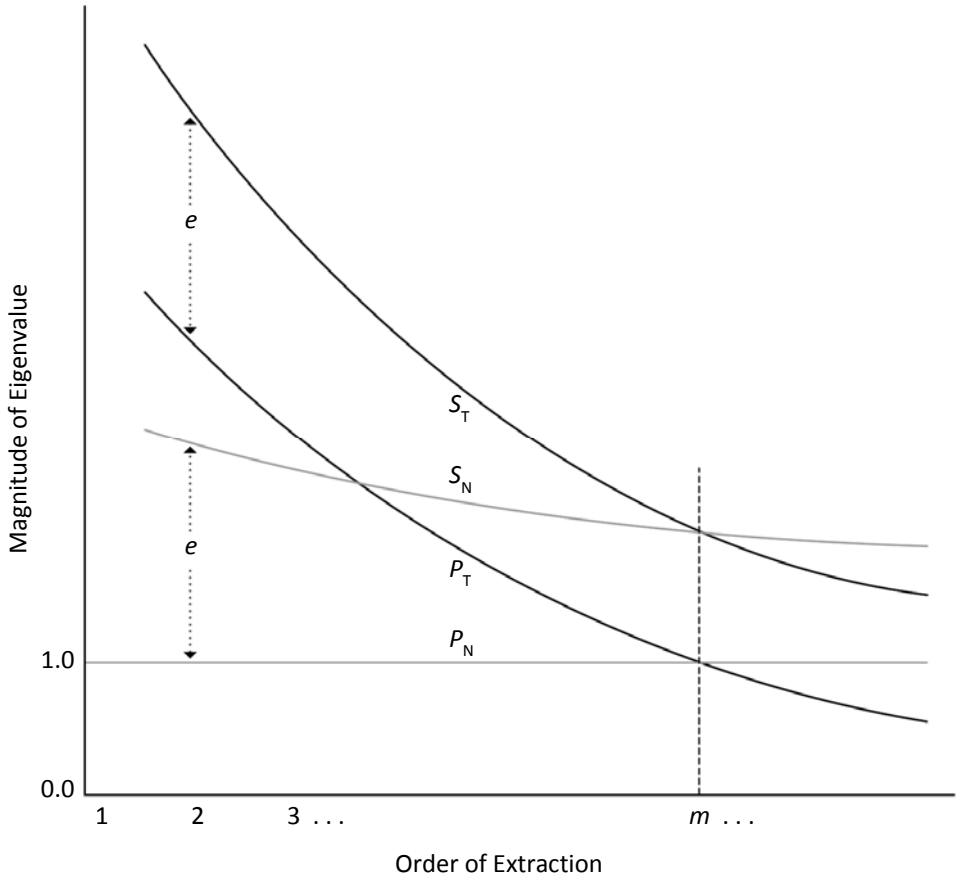

Figure 2.3. Population and sample eigenvalue curves.

The first premise maintains that the Kaiser-Guttman criterion is valid when applied to the population, and therefore the proper lower bound for the number of factors to retain is indicated by the greatest integer that is less than or equal to the point on the horizontal axis where curve P_{T} intersects curve P_{N}—that is, where $P_{\mathrm{T}}(j) = P_{\mathrm{N}}(j)$. The second premise is that the upward biases added to both the target and null samples at any given point are equal—that is, $S_{\mathrm{T}}(j) - P_{\mathrm{T}}(j) = S_{\mathrm{N}}(j) - P_{\mathrm{N}}(j)$. If these assumptions are tenable, then $S_{\mathrm{T}}(j) = S_{\mathrm{N}}(j)$ implies $P_{\mathrm{T}}(j) = P_{\mathrm{N}}(j)$. So, the intersection of the empirical sample curve and the randomly generated curve indicates the intersection of the target and null population curves, which in turn indicates the lower bound for the number of factors (Horn, 1965).

Each of the modified versions of parallel analysis has its own unique justification. The variant that uses communality estimates on the main diagonal of the correlation matrices begins with a different rationalization for parallel analysis. Although it lacks the elegance of Horn's original rationale, this simple and intuitive alternative line of reasoning holds that substantive factors should be those with eigenvalues greater than the eigenvalues generated by random parallel datasets (Humphreys & Ilgen, 1969; Humphreys & Montanelli, 1975). Note the correspondence between this rationale and the alternative rationale offered by Kaiser (1970) regarding the eigenvalue-greater-than-one rule. Because Horn's original method uses only unities in the main diagonal of the observed and randomly generated correlation matrices, Humphreys and Ilgen (1969) concluded that parallel analysis is applicable only to PCA. If parallel analysis is to be applied to the common-factor model, then reduced correlation matrices should be used.

Use of the 95th percentile of the randomly generated eigenvalues as a criterion is a modification of parallel analysis that rests upon both a theoretical and a practical justification. Though parallel analysis is not a true inferential test, Harshman and Reddon (1983; as cited in Glorfeld, 1995) recognized that using the mean of the randomly generated eigenvalues is akin to setting the nominal significance level to $\alpha = .50$. Acknowledging this flaw, Longman et al. (1989) argued for the use of a more appropriate cutoff such as the 95th percentile. Aside from this theoretical argument, the use of a slightly more conservative criterion is also a way to offset the tendency of parallel analysis to overextract (Cota, Longman, Holden, Fekken, & Xinaris, 1993; Glorfeld, 1995).

Advantages and disadvantages. Although underutilized by researchers (Hayton et al., 2004), parallel analysis is one of the more accurate tests of dimensionality as demonstrated in numerous simulation studies (e.g., Mumford et al., 2003; Velicer et al., 2000; Zwick & Velicer, 1986). This accuracy transcends model type, as it has been shown to be accurate with both CFA (Humphreys & Montanelli, 1975) and PCA (Zwick & Velicer, 1986). Furthermore, parallel analysis

has been found to be quite robust to distributional differences between the real data and their parallel counterparts (Glorfeld, 1995). Consequently, it is one of the very few tests generally recommended in the literature (O'Connor, 2000).

Despite these accolades, parallel analysis is not without shortcomings. Although it is undoubtedly an upgrade to the Kaiser-Guttman procedure, parallel analysis inherited some of the criticisms of its forerunner, such as the ambiguity regarding which analytical model is more appropriate (Velicer et al., 2000) and the arbitrary nature of the retention criterion (Fabrigar et al., 1999). In developing parallel analysis, Horn (1965) relied on the basic premises of Guttman (1954), which were based on the common-factor model. However, some authors (e.g., Humphreys & Ilgen, 1969; Humphreys & Montanelli, 1975) contend that the basic form of parallel analysis is appropriate only for PCA since it employs an unreduced correlation matrix. Parallel analysis also has a particularly conspicuous shortcoming when using the 95th percentile as the criterion. Because the total variance in any system is fixed, a large common factor that loads on many variables will deflate subsequent eigenvalues. In turn, this creates the potential for underextraction (Turner, 1998), which numerous Monte Carlo studies have confirmed (Mumford et al., 2003). Conversely, there is a tendency to retain one or more trivial factors when the mean of the randomly generated eigenvalues is used as the cutoff (Zwick & Velicer, 1986).

Until recently, parallel analysis had been criticized for being computationally prohibitive. O'Connor (2000) demonstrated that this procedure (including many of the modifications) is now easily executed with relatively small command programs in various software packages (including SPSS). Nevertheless, parallel analysis is still considered to be more difficult to administer than the classic tests (Hayton et al., 2004). This viewpoint is partially due to the fact that parallel analysis is not yet an integrated "point-and-click" procedure in popular statistical software.

Minimum Average Partial Test

Velicer (1976) proposed the minimum average partial (MAP) test, a novel but still theoretically germane approach to determining the number of components to retain. Although it is based on the component model (and, in fact, can be directly applied only to PCA), the rationale was unquestionably influenced by the theoretical underpinnings of the common-factor model (Velicer, 1976; Zwick & Velicer, 1982). Conceptually, the MAP test is rather simple; computationally, however, it is fairly convoluted and laborious. This computational barrier—paired with the fact that the MAP is not an integrated part of most statistical software packages—has led to this procedure being largely ignored in the past (O'Connor, 2000). But like parallel analysis, the MAP test has recently experienced a renaissance due to the power of modern computers (Stellefson & Hanik, 2008). Along with the aforementioned parallel analysis programs, O'Connor (2000) has also provided MAP test command programs written for SPSS.

Procedure. This method is an iterative process involving p ordered steps numbered from 0 to $p - 1$ (one step for each possible number of components to retain). Let t represent the potential number of components being considered at any given iteration (t = 0, 1,..., $p - 1$). At the tth step the linear influence of the first t components is removed from the observed variables in order to produce a set of partial correlations; specifically, the matrix of tth-order partial correlations $\mathbf{R}_{(t)}$ is computed. Each of the off-diagonal elements of $\mathbf{R}_{(t)}$ is a partial correlation coefficient for a pair observed variables (z_i and $z_{i'}$, say) while controlling for the subset of the first t components $(x_1, x_2, \dots, x_t)$, denoted $r_{ii'\cdot 12\dots t}$ $(i, i' = 1, 2, \dots, p;\ i \neq i')$. The tth-order partial-correlation matrix is given by

$$\mathbf{R}_{(t)} = \left(\mathbf{S}_{(t)}^{-1/2}\right)(\mathbf{R} - \mathbf{A}_t\mathbf{A}_t')\left(\mathbf{S}_{(t)}^{-1/2}\right), \tag{2.20}$$

where

$$\mathbf{S}_{(t)} = \text{diag}(\mathbf{R} - \mathbf{A}_t\mathbf{A}_t') \tag{2.21}$$

(Velicer, 1976). Each step makes use of a $p \times t$ unrotated loading matrix from a t-component solution ($\mathbf{A}_t$). As this is for PCA only, the loading matrices are derived from the unreduced correlation matrix **R**. Conceptually, this can be thought of as computing a solution with t components at each step—that is, extracting one new component per iteration. In practice, however, only the full solution with p components need ever be calculated. At each iteration $\mathbf{A}_t$ can be determined simply by using only the first t columns of the complete $p \times p$ loading matrix **A**:

$$\mathbf{A}_t = \mathbf{A}\begin{bmatrix}\mathbf{I}_t \\ \mathbf{O}_t\end{bmatrix}, \tag{2.22}$$

where the $p \times t$ augmented matrix is composed of the $t \times t$ identity matrix ($\mathbf{I}_t$) and the $(p - t) \times t$ null matrix ($\mathbf{O}_t$). Notice that Equation 2.20 uses the residual covariance matrix $\mathbf{R} - \mathbf{A}_t\mathbf{A}_t'$ (which is also a matrix of partial variances and covariances), and that matrix is pre- and post-multiplied by a diagonal matrix of (partial) standard deviations. This is simply the formula for a matrix of correlation coefficients, only in this case it uses variables that have been residualized on a set of t components (Anderson, 2003).

Once the partial-correlation matrix $\mathbf{R}_{(t)}$ has been obtained, the mean of the squared partial correlations for a solution with t components is then computed:

$$V_t = \frac{1}{p(p-1)}\sum_{i}\sum_{i'} r^2_{ii'\cdot 12\ldots t} \qquad (i \neq i') \tag{2.23}$$

(Velicer, 1976), or, equivalently in matrix-based form,

$$V_t = \frac{1}{p(p-1)}\left[\text{trace}\left(\mathbf{R}^2_{(t)}\right) - p\right] \tag{2.24}$$

(Hong et al., 2006). This process is repeated for 0 to $p - 1$ possible components. The number of components to retain is indicated by the value of t that corresponds to the minimum value of V_t (Velicer et al., 2000).

Note that at the first iteration of this process (i.e., at $t = 0$), the quantity is simply the mean of the squared zero-order correlations. This initial step essentially serves the same purpose as a factorability test (e.g., KMO). At this point, no components have been extracted, and if $V_1 > V_0$, no

components should be extracted at all. Such a result would indicate that the given observed correlation matrix is not suitable for further analysis (O'Connor, 2000).

Rationale. The initial terms in the sequence of V_t values will be decreasing from one step to the next as additional components are partialled out of the set of observed variables. The corresponding components represent common variance, so the removal of their influence will cause the correlations among the (residualized) observed variables to decrease (Velicer et al., 2000). This downward trend will continue until there is a component that is highly correlated with only one of the observed variable but virtually uncorrelated with the remaining variables in the set. Such a component is analogous to a unique factor (Velicer, 1976). Once this "unique component" is partialled out of the observed variables, the test value will increase due to the suppressor effect. Since the initial segment of the sequence of V_t values is monotonically decreasing, the minimum value will be the one immediately before the increase. Hence, the minimum value of the average squared partial correlations indicates the final component that represents common variance (O'Connor, 2000; Zwick & Velicer, 1982). Incidentally, the partial correlation matrix $\mathbf{R}_{(t)}$ (which is also the residual correlation matrix for a solution with t components) most closely approximates an identity matrix at the point where V_t is at a minimum (Hong et al., 2006; Stellefson & Hanik, 2008). Note that the MAP test is somewhat similar to Bartlett's test of sphericity (an indicator of factorability) in that they both indicate when a matrix of correlations resembles an identity matrix.

This phenomenon can also be explained in terms of variances and covariances (Gorsuch, 2003). Consider the basic formula for the correlation between any two variables,

$$r_{ii'} = \frac{s_{ii'}}{s_i s_{i'}}, \tag{2.25}$$

where s_i and $s_{i'}$ are the standard deviations of the two variables, and $s_{ii'}$ is the covariance. When a third variable (in this case, a component) that is highly correlated with only one of the variables is partialled out, the variance of that variable will be decreased while the variance of the other

variable and the covariance will remain relatively unchanged. If s_i (for example) decreases, and $s_{i'}$ and $s_{ii'}$ do not change, then $r_{ii'}$ will increase.

Advantages and disadvantages. The MAP test has been shown to be one of the more accurate dimensionality tests currently available (Stellefson & Hanik, 2008), which is why it has come to be one of only two tests that are widely recommended for general use (O'Connor, 2000). Even though it is relatively accurate, the MAP test is known to have a marked propensity to underextract (Mumford et al., 2003; Velicer et al., 2000; Zwick & Velicer, 1986). This is particularly troublesome since underextraction is generally regarded as a more serious error than overextraction (Fava & Velicer, 1996). One possible reason for this tendency is that components with only a few salient loadings can cause the sequence of MAP test values to increase prematurely (Gorsuch, 1983).

The MAP test also has the benefit of an absolute and unambiguous decision criterion (Stellefson & Hanik, 2008). Although the minimum value is unequivocal, there are nevertheless instances when the minimum and one of the adjacent test values in the sequence are very close in magnitude. In such a scenario the decision as to the number of components to extract is being predicated upon only a very slight difference (Gorsuch, 2003).

A very important yet often overlooked strength of the MAP test is that it has one of the clearest and most consistent theoretical orientation with respect to the concept of latent traits (Velicer, 1976; Zwick & Velicer, 1982). In fact, Velicer described the MAP test as having a rationale that "reflects the original rationale used by Spearman" (Velicer et al., 2000, p. 55). Ironically, the method that is the most theoretically compatible with the common-factor model cannot actually be directly applied to CFA (Velicer et al., 2000). Specifically, when the MAP test is used with a reduced correlation matrix, the test values do not show an increase until well past the $t = p/2$ mark, if at all (Piccone, 2009). Although it cannot be applied directly to CFA, the MAP test (in conjunction with a

preliminary PCA) has been used as an initial step in CFA as a proxy indicator for the number of factors to extract (Zwick & Velicer, 1986).

Empirical Comparisons of the Number-of-Factors Methods

Although the general strengths and weaknesses of each individual method have been presented in the preceding discussion, this reveals little about their performance characteristics relative to one another. There are only a few comprehensive Monte Carlo studies that compare the most prominent number-of-factor techniques in a head-to-head format. Unfortunately, none of these studies investigate all six of the methods presented here. Nevertheless, Bartlett's chi square (BCS), Cattell's scree test (CST), the Kaiser-Guttman criterion (KG), parallel analysis (PA), and the minimum average partial test (MAP) are each covered by at least one of the major published studies. The proportion of variance extracted was the only method excluded from these studies.

The results of these studies were typically reported in terms of accuracy and precision. In this context, accuracy generally refers to the proportion of generated samples wherein a given dimensionality test recovered the correct number of factors built into the population. Some studies use the amount of discrepancy between the mean number of factors estimated by a given method and the actual number of factors is reported as an indicator of accuracy. Precision refers to the amount of variance observed in the number of factors estimated by a given method. Accuracy and precision are essentially generalizations of validity and reliability, respectively.

Zwick and Velicer (1986)

A classic factor-analytic Monte Carlo study focusing on the problem of dimensionality is that of Zwick and Velicer (1986). This study investigated the performances of BCS, CST, KG, MAP, and PA by manipulating the number of observed variables (36 and 72), the true number of factors (3, 6, and 9), sample size (2p and 5p), and the absolute magnitude (or saturation) of the loadings (.5 and .8). BCS was tested at a typical nominal significance level (.05) and also at two more conservative levels (.001 and .0005). Only one form of PA was tested using the mean eigenvalues from 50

random parallel datasets as the criteria. Since only 10 random samples were drawn from each manufactured population, CST also could be tested. The scree plot of each sample was judged by two trained raters.

KG was observed to have a strong tendency to overextract; in fact, it never underextracted in any of the samples. Furthermore, the amount of upward bias increased as the number of observed variables increased. BCS did not perform much better as it was shown to be the least precise of all the methods in the study. It also had problems with accuracy, showing a tendency to overextract with more observed variables, greater levels of saturation, and larger sample sizes. Further, the use of conservative significance levels with larger sample sizes was shown to be an ineffective strategy for BCS. CST showed better accuracy and precision than either KG or BCS, yet it was correct in only 57% of the samples. It tended to be more accurate and precise with greater levels of saturation, but the number of variables had no influence. While it was observed to overextract, the most critical issue with the CST was the modest levels of inter-rater reliability. The MAP test was consistently more accurate and precise than KG, BCS, and CST, and its performance appeared to be unaffected by sample size, but there was, however, some influence by the degree of saturation. When the population saturation level was low, the MAP test would be correct or in error by only ±1 for 78% of the samples; with a high saturation, this same result occurred in 100% of the samples. When the MAP test is incorrect, it tends to underextract (90% of its errors were underestimates). The most accurate and precise method in this simulation was PA. Accuracy and precision of PA both improved with increases in sample size, saturation, and the variable-to-factor ratio. With low levels of saturation, PA was found to be either correct or off by only ±1 for 97% of the cases. This occurred for 100% of the cases when there were high levels of saturation. However, PA was shown to have an upward bias as 65% of its errors were overextractions.

Based upon these results, Zwick and Velicer (1986) recommend MAP and PA (using the mean as a cutoff) as the preferred methods for general use. While CST should not be trusted as a

stand-alone method, it can be used as an auxiliary indicator in conjunction with MAP or PA. Finally, KG and BCS should be avoided altogether.

Velicer, Eaton, and Fava (2000)

Although not a broad as the Zwick and Velicer (1986) study, an in-depth comparative review and Monte Carlo by Velicer et al. (2000) studied KG, MAP, and PA. Not only were these methods compared on the basis of their observed performances, but also with respect to their rationales. Two forms of PA were tested, both using the mean eigenvalues generated from random parallel datasets as the criteria. However, the first version used a small number of parallel datasets (5 replications; PA-5), while the second version used moderate number (100 replications; PA-100). The sample size (75, 150, and 300), number of variables (24, 48, and 72), true number of factors (3, 6, 9, 12, and 18), and saturation (.4, .6, and .8) were the parameters for this simulation.

KG was found to be the least accurate method. Of all the methods being studied, KG produced the lowest proportion of correct estimates, and it overextracted by the greatest margin. It was also observed to be the least precise method with a standard deviation more than two times greater than the next largest standard deviation. The MAP technique performed well, as did both of the PA techniques. The PA tests were both quite accurate and precise, with PA-100 performing only marginally better than PA-5. Overall, the PA methods were only slightly better than the MAP test. However, the rationale of the MAP test was judged to be superior to that of PA or KG. Based upon this study, the authors strongly recommend both MAP and PA (in its various forms), while the use of KG is strongly discouraged.

Mumford, Ferron, Hines, Hogarty, & Kromrey (2003)

Mumford et al. (2003) performed a Monte Carlo study that is particularly noteworthy because, unlike the previous studies which used PCA as the method of extraction, this study focused exclusively on the common-factor model. Moreover, its innovative use of the omega-squared statistic (a measure of effect size used in ANOVA) provides a useful new analytical perspective. This

study examines KG, MAP, and two forms of PA by manipulating the true number of factors (3, 5, and 7), number of variables ($3m$, $5m$, and $10m$), sample size ($3p$, $5p$, $10p$, $20p$, and $40p$), magnitude of communalities (a random mix of either .6, .7, .8 or .2, .3, .4), and the level of inter-factor correlation (0, .3, and .5). The first version of PA is the basic form, using the set of mean eigenvalues as the criterion (PA-M), while the second form used a more conservative cutoff value by adding .1 to each of the mean eigenvalues (PA-C). Note that the PA-C is somewhat similar to the form of PA that uses the 95th percentile as the cutoff; however, the PA-C uses an arbitrary value to set the conservative upper limit. Mumford et al. (2003) offered no rationale behind the use of this criterion in place of the 95th percentile. Both PA-M and PA-C used 1000 parallel samples in this study.

The results regarding the mean number of estimated factors are quite similar to the previous studies; however, the more revealing results are those based upon the analysis of the omega-squared values. This type of analysis is possible since this study bears an unmistakable resemblance to a five-way factorial ANOVA. The population conditions (actual number of factors, sample size, number of observed variables, communality, and factor inter-correlation) were treated as independent variables, and the number of estimated factors served as the dependent variable. For each number-of-factors method, an omega-squared statistic ($\widehat{\omega}^2$) was calculated for each possible interaction and main effect. For example, an omega squared was calculated for the MAP test with respect to the main effect of the true number of factors with three levels. As for the meaning to be gained from such an analysis, recall that the omega-squared statistic can be interpreted as the proportion of variance in the dependent variable that is related to the different levels of the independent variable (Keppel & Wickens, 2004). Ideally, differences in the true number of factors would have the greatest influence, while the effects of the other population conditions (and their interactions) would be negligible.

For the omega-squared statistics associated with the actual-number-of-factors main effect, PA-C had the greatest value by a wide margin ($\widehat{\omega}^2 = .456$). The next greatest omega-squared value

was that of PA-M ($\hat{\omega}^2 = .305$), followed by KG ($\hat{\omega}^2 = .271$). Surprisingly, the MAP test had the poorest performance ($\hat{\omega}^2 = .258$) of all the methods. As previously noted, the actual-number-of-factors main effect would ideally have the most influence. Although this was true for PA-M and PA-C, the greatest omega-squared values for KG and MAP were both associated with the number of observed variables ($\hat{\omega}^2 = .380$ and $\hat{\omega}^2 = .361$, respectively). The second largest values for the PA-M and PA-C techniques were also from the number-of-observed-variables main effect. The main effect of the level of communality was shown to have a moderate influence on PA-M ($\hat{\omega}^2 = .128$) and MAP ($\hat{\omega}^2 = .162$), but it had only a small impact on PA-C ($\hat{\omega}^2 = .045$) and KG ($\hat{\omega}^2 = .056$). The omega-squared values for the remaining main effects and interactions were negligible for all of the techniques.

Piccone (2009)

Most recently, Piccone (2009) conducted one of the most extensive factor-analytic Monte Carlo simulations to date in terms of the variety of parameters and number of conditions used. This study manipulated the number of factors in the population (1 to 5, 8, and 10), the number of indicator variables per factor (3, 5, 8, 10, and 15), the magnitude of the factor loadings (.3 to .5, .6 to .8, and .3 to .8), sample size (250, 500, and 1000), and the magnitude of inter-factor correlations (0, .2, and .4). This simulation also explored the effects of the presence of unique indicators (observed variables that do not load onto any factor in the population and therefore introduce random error into the system). This study examined MAP, KG, and two versions of PA. One version of PA used unreduced correlation matrices (PA-U), while the other version used reduced correlation matrices (PA-R). Both forms of PA used the 95th percentile as the cutoff. This study also examined two forms of the standard-error scree (SES), which is another type of objective scree test.

In this simulation, KG performed poorly with less than 25% overall accuracy and a tendency to overextract by a margin of nearly six factors. The performance of KG declined as saturation levels decreased and with the introduction of unique indicator variables. KG also displayed an

unusual characteristic as its performance deteriorated as the variable-to-factor ratio increased. Although better than KG, neither form of SES was very accurate overall (less than 50%), and both forms overextracted by at least four factors. SES was found to be fairly resilient to larger inter-factor correlations, but it was sensitive to an increase in the number of factors and to the presence of unique indicators. Overall, MAP performed much better than KG or SES as it was observed to be reasonably accurate (66%); however, it was shown to have a tendency to underextract by more than one factor. Unlike the other dimensionality procedures in the study, MAP was observed to be completely unaffected by the presence of unique indicators. MAP was found to be quite sensitive to the number of salient loadings per factors. Specifically, the accuracy of MAP was less than 10% under the condition of three variables per factor. With an overall accuracy greater than 90% and a very slight tendency to underextract, PA-U was clearly the top-performing procedure in the study. PA-R was not far behind with 83% accuracy and a slight tendency to overextract. PA-R was remarkably stable across the levels of inter-factor correlation, but PA-U was not quite as robust. In particular, PA-U had trouble with higher inter-factor correlations when eight or more factors were present. Both forms of PA were shown to be fairly resilient to the presence of unique indicators.

Summary

The initial sections of this chapter reviewed the developmental history, key theoretical concepts, and procedural aspects of factor analysis. This facilitated an in-depth discussion of six of the most prominent dimensionality tests and a summary of the three most comprehensive number-of-factor Monte Carlo studies.

This review and discussion of the literature has revealed rational arguments for researchers to choose the most appropriate factor-analytic approach with respect to the purposes of their particular studies, especially since computationally expedient shortcuts and approximations are no longer relevant given the speed and efficiency of modern computers. Specifically, researchers (social and behavioral scientists in particular) should use CFA rather than a PCA proxy when

exploring latent constructs. Moreover, many authors warn against eclecticism when selecting a method for determining the proper number-of-factors to retain. Unfortunately, the two most promising number-of-factors methods are either theoretically ambiguous (parallel analysis) or simply inapplicable to CFA (MAP test). The latter is particularly frustrating given that the MAP test is based on a rationale that is considered to be the most theoretically consistent with the common-factor model.

Conclusions from the Literature

Clearly, there has been a long and interesting history of efforts to resolve the question regarding the proper number of common factors to retain. Unfortunately, some of the most promising guidelines for determining dimensionality are either theoretically incongruous (or at least ambiguous) with respect to the common-factor model or pragmatically incompatible with the procedures of CFA. Yet, there appeared to be an opportunity to further extend the line of thought regarding the application of Velicer's (1976) rationale to the common-factor model. Gorsuch (1990) and Piccone (2009) extended the basic rationale of Velicer by applying the MAP test to the reduced (rather than the unreduced) correlation matrix. However, Piccone demonstrated that this common-factor version of the MAP test is quite inaccurate as an indicator of the number of factors, noting that it "flatly blew up" (p. 86). When applied to a reduced correlation matrix, the incremental removal of common factors from the system (rather than components) typically results in a gradual monotonic decline of the MAP-test values with no increase (and no minimum value) until well past the $p/2$ mark (if at all).

Nevertheless, there may yet be a way to apply Velicer's rationale to the common-factor model by incorporating a seemingly unrelated procedure—specifically, the Kaiser-Meyer-Olkin (KMO) indicator of factorability. Known as the sequential KMO (SKMO) procedure, this proposed approach applies the MAP test to a reduced correlation matrix, and the KMO test is then applied to the residual correlation matrix at each sequential stage of the MAP test. The following is a more

detailed elaboration of the procedural steps and the theoretical rationale of the SKMO indicator of dimensionality.

Procedure

Much like the MAP test, this proposed method consists of a sequence of p iterations. Let t represent the number of factors being considered at each step ($t = 0, 1, \ldots, p-1$). At the beginning of each iteration, the set of communalities are estimated in order to produce the reduced correlation matrix $\mathbf{R}^*$, which is subsequently subjected to eigenanalysis. This will yield a solution with t common factors with a $p \times t$ loading matrix $\mathbf{A}_t$. This in turn is used to produce the tth-order residual correlation matrix

$$\mathbf{R}_{(t)} = \left(\mathbf{S}_{(t)}^{-1/2}\right)(\mathbf{R} - \mathbf{A}_t\mathbf{A}_t')\left(\mathbf{S}_{(t)}^{-1/2}\right), \tag{2.26}$$

where

$$\mathbf{S}_{(t)} = \text{diag}(\mathbf{R} - \mathbf{A}_t\mathbf{A}_t'). \tag{2.27}$$

Note that $\mathbf{R} - \mathbf{A}_t\mathbf{A}_t'$ is a partial covariance matrix, and $\mathbf{S}_{(t)}$ is a diagonal matrix of partial variances; this formula is of the same general form as that of a zero-order correlation matrix (Anderson, 2003). In this case, however, the elements of the matrix $\mathbf{R}_{(t)}$ represent the correlations of the observed variables which have all been residualized on the set of the first t factors. It is at this point in each iteration that the KMO test is applied to the residual correlation matrix $\mathbf{R}_{(t)}$. This begins by computing the anti-image correlation matrix for the residualized variables:

$$\mathbf{Q}_{(t)} = \left[\left(\text{diag}\,\mathbf{R}_{(t)}^{-1}\right)^{-1/2}\right]\left(\mathbf{R}_{(t)}^{-1}\right)\left[\left(\text{diag}\,\mathbf{R}_{(t)}^{-1}\right)^{-1/2}\right]. \tag{2.28}$$

The KMO test value for the tth-order residual correlation matrix $\mathbf{R}_{(t)}$ is given by

$$K_t = \frac{\text{trace}\left(\mathbf{R}_{(t)}^2\right) - p}{\text{trace}\left(\mathbf{R}_{(t)}^2\right) + \text{trace}\left(\mathbf{Q}_{(t)}^2\right) - 2p}, \tag{2.29}$$

or equivalently,

$$K_t = \frac{\sum_i \sum_{i'} r_{ii'(t)}^2}{\sum_i \sum_{i'} r_{ii'(t)}^2 + \sum_i \sum_{i'} q_{ii'(t)}^2} \quad (i \neq i'), \tag{2.30}$$

where $r_{ii'(t)}$ is the partial correlation coefficient of variables *i* and *i*′ after being residualized on the first *t* factors, and $q_{ii'(t)}$ is the anti-image correlation for variables *i* and *i*′ which have also been residualized on the set of the first *t* factors. Other than being applied to the residual correlation matrix $\mathbf{R}_{(t)}$ rather than **R**, the use of the KMO procedure here is no different than the usual application as described in a previous section (compare Equations 2.28, 2.29, and 2.30 to Equations 2.11, 2.12, and 2.13). This process is repeated for each of the possible number of factors to retain, producing a sequence of KMO values.

Beginning with the value from the first iteration where there are no factors extracted (i.e., K_0), the KMO test values are evaluated in sequential order with respect to some criterion value (e.g., .60). If the KMO test value at step *t* indicates that the residualized correlation matrix $\mathbf{R}_{(t)}$ is factorable—that is, there exists at least one factor underlying the residualized correlation structure—then there must be at least $t + 1$ factors, and the test value from the next iteration is inspected. If K_t indicates otherwise, then there are *t* factors, and the procedure is concluded. As noted in a previous section of this chapter, the current standard for the conventional use of the KMO is that a value of .60 or greater indicates a factorable correlation matrix. Although the exact cutoff criterion has yet to be determined for the SKMO, preliminary trials with a number of well-known plasmodes indicate that .50 may be a more reasonable choice.

To better explain the basic SKMO procedure, an annotated example is provided. This demonstration will use the example output found in Table 2.2. These results are based on a dataset composed of 24 observed variables and 301 subjects, and .60 will be the specified criterion. Beginning with the initial step at $t = 0$, the test value of $K_0 = .90$ is compared to the cutoff value of .60. Since the value passes, there is at least one factor to be retained. The value at $t = 1$ is then inspected. The test value $K_1 = .74$ is also greater than .60, so there are at least two factors in the solution. The next test value, $K_2 = .64$, also passes, revealing that there are at least three factors in

the solution. Moving on to $t = 3$, the test value $K_3 = .59$ does not pass, and the process stops. The SKMO procedure indicates that three factors should be retained.

Table 2.2

Abbreviated Example of SKMO Output

t	K_t
0	.90
1	.73
2	.64
3	.59
4	.50
5	.48
6	.47
⋮	⋮
22	.49
23	.49

Rationale

This proposed approach uses the rationale of the KMO procedure (Kaiser, 1970, 1981) to augment the ideas from Piccone's (2009) application of the MAP test (Velicer, 1976) to the common-factor model. As previously stated, the basic idea behind this procedure is to apply the MAP test to a reduced correlation matrix, and the KMO test is then applied to the residual correlation matrix at each sequential stage of the MAP test. Now consider iteration *t* of the aforementioned MAP procedure on a reduced correlation matrix. A solution of *t* factors has been extracted from $\mathbf{R}^*$, and these factors are then partialled out of the set of *p* observed variables. The residual correlation matrix $\mathbf{R}_{(t)}$ reflects the correlations among the variables which have now been residualized on the set of *t* factors. Now suppose that this residual correlation matrix was simply any other another correlation matrix, only this correlation matrix comes from a domain barren of those latent traits represented by the set of *t* factors that has been previously partialled out. As it is a correlation matrix, it can therefore be assessed for factorability. Recall that assessing the

factorability of a correlation matrix is tantamount to evaluating evidence for the existence of at least one factor underlying the observed correlation structure. So, if the *t*th-order residual correlation matrix $\mathbf{R}_{(t)}$ passes the test of factorability (in this case, the KMO), then there is empirical evidence to support the supposition that $\mathbf{R}_{(t)}$ has at least one factor, and by extension, $\mathbf{R}^{*}$ has at least $t + 1$ factors. Conversely, if the factorability test indicates that $\mathbf{R}_{(t)}$ is not a viable candidate for factoring, then there are no further factors to extract from $\mathbf{R}^{*}$ beyond the set of t factors.

Notice that once the residual system has been depleted of common factors, the sequence of K_t values should theoretically approach .50, which would imply $\sum_i \sum_{i'} r^2_{ii'(t)} = \sum_i \sum_{i'} q^2_{ii'(t)}$. Since an anti-image correlation for two variables that have been residualized on a set of *t* factors can be thought of as a correlation between the same two residualized variables after further stripping away the linear influence from any remaining the common factors, then the equality of the sum of the squared residual correlations and the sum of the squared anti-image correlations of residualized variables implies that there is no remaining common-factor influence to be removed from the system.

Footnotes

[1] To be historically accurate, this model was originally known as the two-factor model because the variance of an observed variable was attributed to two independent sources—namely, a common factor (in this case, *g*) and a unique factor (Rummel, 1970). It has since become standard practice to distinguish these models in terms of the relative number of common factors. For example, Thurstone (1947) refers to Spearman's approach as the single-common-factor model.

[2] While Thurstone (1947) acknowledged the work of Truman Kelley and of Harold Hotelling (the originators of principal-components analysis) for their innovative computational algorithms used in factoring a correlation matrix, he gave credit to Walter Bartky (a colleague in the Department of Astronomy at the University of Chicago) for the original suggestion to perform eigenanalysis on a correlation matrix.

[3] Indeed, a misunderstanding of the difference in the respective purposes of these two methods led to animosity between Thurstone and Hotelling for a time. It would seem, however, that this subsided by the time Thurstone joined the faculty at the University of North Carolina where Hotelling was already a professor (Jones, 2007).

[4] The other major division within factor analysis has no commonly used name. They are the collection of vestigial methods that served as a computationally expedient approximation for a principal-factor solution (Mulaik, 2010). Examples include the diagonal and centroid methods. These approaches are mentioned here for completeness and will not be discussed further.

[5] There are numerous variants of CFA that have additional model assumptions, such as maximum-likelihood, alpha, and image factor analyses. The prototypical form of CFA is known as

principal-axes factor analysis (PAFA); indeed, many authors use the term PAFA synonymously with CFA (Gorsuch, 1983). This will also be the convention adopted here as this study focuses on only the basic CFA model.

[6] Although they were not listed as co-authors of either paper, Kaiser credited both Meyer and Olkin as co-originators of this method.

[7] Iteration by refactoring with SMCs is the default estimation method when performing CFA with SPSS. The program will automatically use the greatest absolute correlations if the SMCs cannot be computed, but this is not a method that can be requested explicitly.

[8] The term *rotation* is a reference to the linear transformations of the vectors representing the factors with respect to the variable subspace. In a Cartesian framework, these vectors appear to rotate about the origin.

[9] It should be acknowledged that until only recently the literature on parallel analysis has been filled with two other alternative approaches to the Monte Carlo method: regression equations to predict the probable mean of random eigenvalues, and eigenvalue tables for interpolation (Cota et al., 1993). These were invented as computationally less intensive shortcut methods (Velicer, Eaton, & Fava, 2000). These methods have been made obsolete by modern computing power, so they will not be presented here.

CHAPTER III

METHODS

Based on the issues discussed in Chapter II, this study was designed to examine the performance of a novel criterion for the number of factors to be retained in common-factor analysis (CFA). The method under investigation is the one described at the conclusion of the previous chapter—namely, the sequential Kaiser-Meyer-Olkin (SKMO) procedure. A Monte Carlo simulation was designed to determine the accuracy, precision, and bias of the SKMO method in reproducing known factor structures under various conditions.

Manipulation of the Simulation Parameters

The specific levels of each parameter used in this simulation were selected in accordance with relevant current standards, scenarios that typically arise in applied research, and previous factor-analytic Monte Carlo studies. In particular, the design of the present study draws heavily from the simulation conducted by Piccone (2009) since it is one of the most recent and most extensive Monte Carlo studies dealing with the number-of-factors issue. Thus, five general elements of the factor problem were manipulated in the present study—namely, the actual number of factors, variable-to-factor ratio, sample size, magnitude of the loadings, and magnitude of the inter-factor correlations. These parameters and their selected levels are described in the following section.

Actual Number of Common Factors

As this study deals with the methods for determining dimensionality, the true number of factors present in the population is obviously a key ingredient in this simulation. Thus, this

simulation accommodated a wide range for the number of common factors underlying the observed variables. Specifically, there were 1, 2, 3, 4, 5, 8, and 10 factors. The most basic factor model is unidimensional, thus the smallest number of factors in this study was one (although it is theoretically possible for there to be no factors). A maximum value of 10 was used as theoretical frameworks in the social sciences (and their corresponding measurement instruments) will typically not have more than 10 distinct dimensions (Piccone, 2009). Further, many of the previous Monte Carlo studies employ no more than 10 factors (e.g., Fava & Velicer, 1992, 1996; Humphreys & Montanelli, 1975; Mumford et al., 2003; Piccone, 2009; Yeomans & Golder, 1982; Zwick & Velicer, 1986).

Variable-to-Factor Ratio

Since it has been found to influence the accuracy and stability of factor solutions, the number of salient variables per factors is an important consideration in factor analysis (Velicer et al., 2000; Yeomans & Golder, 1982; Widaman, 1993). The ratios of 3:1, 5:1, 8:1, and 10:1 were the specific set of values used in this study. The ratio of three variables per factor is considered to be the minimum for stable factors (Gorsuch, 2003; Russell, 2002; see Kline, 2005, for an in-depth discussion on this topic). Velicer et al. identified eight salient loadings per factor as a fairly large ratio for the purposes of producing stable factors. Ten variables per factor served as the upper limit in this simulation since only a few previous studies explore ratios greater than that value.

Since the true number of factors has been fixed, setting the variable-to-factor ratio also determined the number of observed variables (see Table 3.1). When crossing the number of factors with the variables-to-factor ratio, some of the resulting numbers of variables are too small to be useful in factor analysis. For example, only six variables are required when the number of factors is set at two and the number of variables per factor is three. For this reason, only those combinations using eight or more variables were considered in this study.

Table 3.1

Number of Variables as a Function of the True Number of Factors and the Variable-to-Factor Ratio

Number of Variables per Factor	Number of Factors						
	1	2	3	4	5	8	10
3	-	-	9	12	15	24	30
5	-	10	15	20	25	40	50
8	8	16	24	32	40	64	80
10	10	20	30	40	50	80	100

Note. The levels of the true number of factors and the levels of variable-to-factor ratio were not completely crossed. A dash (-) indicates if a particular parameter combination was not used.

Pattern-Magnitude Interval

The absolute magnitude of the factor loadings has been identified as one of the most crucial aspects of a factor solution (Hakstain, Rogers, & Cattell, 1982; Velicer et al., 2000; Yeomans & Golder, 1982; Zwick & Velicer, 1982, 1986). Note that manipulation of loading magnitudes is tantamount to manipulating the levels of communality. In this study, ranges of values were used in order to avoid unrealistically homogeneous sets of factor loadings. So, three conditions were created by defining minimum and maximum pattern coefficients (denoted by $\beta_{\min}$ and $\beta_{\max}$, respectively). Note that this must be constructed specifically in terms of the pattern coefficients since the factors are not necessarily orthogonal. This simulation used the following saturation intervals: low (.3 to .5), high (.6 to .8), and wide (.3 to .8).

An absolute magnitude of .3 is a commonly cited lower bound for salience (Comrey, 1978; Gorsuch, 1983; Zwick & Velicer, 1982). Tabachnick and Fidell (2001) specifically recommend the use of .32 as the lower limit for salience since that will account for 10% of the variance in an observed variable. A loading of .8 is considered to be a large value (Fava & Velicer, 1992, 1996; Linn, 1968; Velicer et al., 2000); encountering a loading greater than .8 is quite rare (Costello & Osborne, 2005; Zwick & Velicer, 1982). Thus, the low-salience and high-salience intervals were set

at .3 to .5 and .6 to .8, respectively. Realistically, however, most applications of factor analysis to real data will find a mixture of high and low loadings (and, subsequently, high and low communalities). Thus, a third condition was constructed to reflect such a wide range of values by using the lower limit of the low range (.3) and the upper limit of the high range (.8).

In order to produce a (relatively) heterogeneous set of loadings for any of the three intervals, the population pattern matrices (denoted by **B** [beta]) used in this simulation were constructed in an incremental fashion according to the number of variables per factor. In this study, p/m is always a positive integer, so each factor has exactly p/m non-zero pattern coefficients ranging in value from $\beta_{\min}$ to $\beta_{\max}$ by uniform increments. The increment (denoted by Δ_β) between any two adjacent pattern coefficients on the same factor is given by

$$\Delta_\beta = \frac{\beta_{\max} - \beta_{\min}}{{}^{p}/_{m} - 1}. \tag{3.1}$$

Other simulation studies (e.g., Mumford et al., 2003) have used randomly selected loadings from similarly defined intervals. However, the incremental approach was selected for this study as it ensures heterogeneity where random selection does not.

Sample Size

The sample sizes of the generated datasets will be manipulated by using 250 (small), 500 (medium), and 750 (large) observations. A sample size of 250 was set as the minimum for this simulation since that number is marginally less than the minimums recommended by the most recent guidelines (Gorsuch, 2003; Tabachnick & Fidell, 2001). Subsequent values are at increments of 250 in order to provide a moderate level of resolution.

Inter-Factor Correlation

Factor solutions are not necessarily orthogonal; in fact, it is more reasonable to assume that a solution is oblique (Costello & Osborne, 2005). Furthermore, it has been shown that the amount of factor correlation can distort the estimated number of factors to retain for some dimensionality

tests (Piccone, 2009). Therefore, this simulation varied the magnitudes of the factor correlations by using the values of 0, .2, and .4. A zero correlation was used in this simulation since it is the minimum possible value, and it represents the basic orthogonal factor model. Factors tend to become indistinguishable when they are too highly correlated; therefore, factor correlations are typically restricted to absolute values of .4 or less (Gorsuch, 1983). This value was used as the upper limit. The value of .2 was selected as a value representative of a moderate level of factor correlation. This was implemented by setting each of the off-diagonal elements of the factor-correlation matrix (denoted by $\mathbf{\Phi}$) to the same value.

Description of the Monte Carlo Design

This simulation manipulated the five parameters according to the preceding profiles. The assigned sets of values for the parameters are summarized in Table 3.2. A model was constructed for each unique combination of parameter values, which was in turn used to generate 100 random datasets each. Using 100 replications per condition appears to be a consistent benchmark, although there exist previous Monte Carlo studies that used as few as four (Yeomans & Golder, 1982) and as many as 10,000 (Mumford et al., 2003) replications. The basic output quantities for this study are summarized in Table 3.3.

Table 3.2

Profile of Monte Carlo Conditions

Parameter	Levels
Actual number of factors[a]	1, 2, 3, 4, 5, 8, 10
Variable-to-factor ratio[a]	3:1, 5:1, 8:1, 10:1
Pattern-magnitude interval	[.3, .5], [.6, .8], [.3, .8]
Sample size	250, 500, 750
Inter-factor correlation	0, .2, .4

[a]All levels of the actual number of factors and the variable-to-factor ratio are not completely crossed. There are 25 combinations of these two parameters that will be used in this study.

Table 3.3

Basic Output Quantities for the Monte Carlo Simulation

Quantity	Value
Number of conditions	(25)(3)(3)(3) = 675
Number of replications per condition	100
Number of datasets generated	67,500

Data-Generation Procedure

In order to implement the generation of the random datasets, an original suite of command macros was written for the IBM SPSS statistical software package (version 19) on the Microsoft Windows 7 operating system (64-bit). These procedures for generating the data were based upon previous simulation models by Cattell and Vogelmann (1977), Piccone (2009), Snook and Gorsuch (1989), and Velicer et al. (2000).

The data-generation process begins by establishing a set of matrices at the population level. A brief description of each of these matrices is given in Table 3.4. The values of the elements and the orders of these matrices are determined by a specific combination of the five conditional elements discussed in the previous section of this chapter.

Table 3.4

Population Matrices Used for Data Generation

Matrix	Order	Description
$\mathbf{A}$	$p \times m$	Unrotated (maximized) loading matrix (alpha)
$\mathbf{B}$	$p \times m$	Rotated pattern matrix (beta)
$\mathbf{P}$	$p \times p$	Unreduced correlation matrix (rho)
$\mathbf{P}^*$	$p \times p$	Reduced correlation matrix (rho)
$\mathbf{\Phi}$	$m \times m$	Factor correlation matrix
$\mathbf{\Psi}$	$p \times p$	Unique loading matrix

The generated data was from a population that conforms to the common-factor model. Accordingly, the population correlation matrix is given by $\mathbf{P} = \mathbf{B\Phi B}' + \mathbf{\Psi}^2$. For each unique

combination of conditions as defined by the five parameters, there was a set of input quantities—namely, sample size, actual number of factors, variable-to-factor ratio, pattern-magnitude range ($\beta_{\min}$ and $\beta_{\max}$), inter-factor correlation, and the number of replications. Once these values were set, the following steps were used to produce a set of raw-data matrices to which the number-of-factor procedures discussed in Chapter II were applied:

1. Setting values and dimensions for the pattern (**B**) and inter-factor correlation (**Φ**) matrices will give the population reduced correlation matrix $\mathbf{P}^* = \mathbf{B\Phi B}'$.
2. Perform eigenanalysis on **P*** to produce **A**.
3. Calculate the population unique loading matrix $\mathbf{\Psi} = (\mathbf{I} - \text{diag}\ \mathbf{AA}')^{1/2}$.
4. Generate the factor-score matrices $\widetilde{\mathbf{X}} = [\tilde{x}_{kj}]$ and $\widetilde{\mathbf{Y}} = [\tilde{y}_{ki}]$ where $\tilde{x}_{kj}$ and $\tilde{y}_{ki}$ are randomly selected from a standard normal distribution (zero mean, unit variance).
5. Compute the raw-data matrix $\mathbf{Z} = \widetilde{\mathbf{X}}\mathbf{A}' + \widetilde{\mathbf{Y}}\mathbf{\Psi}$.
6. Compute the sample correlation matrix $\mathbf{R} = \mathbf{Z}'\mathbf{Z} - \frac{1}{n}\mathbf{Z}'\mathbf{JZ}$ where **J** is an $n \times n$ matrix of unities.
7. Submit the set of generated sample correlation matrices to the analysis macros.

Data Analysis

An original set of command syntax programs were developed in SPSS (version 19) for the proportion of variance extracted, Bartlett's chi square, the Cattell-Nelson-Gorsuch (CNG) scree test, the Kaiser-Guttman criterion, and the SKMO. The parallel analysis and MAP programs were adapted from those written by O'Connor (2000).

Specifications for the Number-of-Factors Procedures

All of the number-of-factor methods discussed in Chapter II (including the SKMO) were applied to each of the 100 datasets generated for each possible combination of conditions per the aforementioned process. The Kaiser-Guttman rule, the CNG scree test, and the MAP test each have only one version to be evaluated. The proportion of variance extracted was applied using 75% as the cutoff criterion, and Bartlett's chi-square test was performed with a nominal significance level

of .01. Four different versions of parallel analysis were tested. The first variant used the unreduced correlation matrix **R** in conjunction with the mean-eigenvalue criterion. The second also used **R**, but with the 95th percentile as the cutoff value. The third and fourth variants both used the reduced correlation matrix **R*** with the mean-eigenvalue and 95th-percentile criteria, respectively. Each of these variants of parallel analysis used 50 random uniform datasets to calculate the criterion value for each analysis. Note that the parallel datasets were generated using uniform rather than normal distributions because these two types of distributions produce virtually identical results (Glorfeld, 1995). This equivalence was also confirmed in a pilot test. Four different versions of the SKMO were explored. These variants arise from two different communality estimation methods (SMCs with and without iteration) and two different criteria values (.50 and .60). For the two versions of the SKMO that use iteration by refactoring, iteration ceased when the convergence criterion of .001 had been met or 25 iterations had been processed (whichever occurs first); these values were selected since they are the default settings in SPSS (SPSS, 2010). The command syntax programs for these tests were written in accordance with the procedures outlined in Chapter II.

Performance Indicators

All performance indicators used in this study are based upon the discrepancy between the predicted number of factors to retain and the true number of factors present in the population. The following indicators were adapted from those used by Zwick and Velicer (1986) and Piccone (2009).

Accuracy. The proportion of trials that result in a dimensionality test predicting the correct number of factors to retain (i.e., the hit rate) served as the primary measure of accuracy. The hit rate is equivalent to the proportion of the samples with discrepancies equal to zero.

Bias. Closely related to accuracy is the concept of bias. The bias of a particular number-of-factors method refers to the direction and magnitude in which it tends to err when predicting the

number of factors to retain—that is, the tendency of a given technique to underextract or overextract and by what margin. The mean of the discrepancy between the predicted and true number of factors was used as the indicator for bias.

Precision. The precision (or consistency) of a dimensionality test was measured by the standard deviation of the discrepancies between the predicted and true number of factors. This provides an indication of the relative amount of variability associated with each of the dimensionality procedures. A smaller standard deviation indicates greater precision.

CHAPTER IV

RESULTS

As outlined in Chapter III, there were 675 population correlation matrices representing all the combinations of conditions set for the Monte Carlo simulation. One-hundred sample correlation matrices were generated from each set of population conditions in order to evaluate 13 different number-of-factors procedures.

This chapter provides a comprehensive report of the results of the Monte Carlo simulation, and it is organized in such a manner as to facilitate consideration of the research questions presented in Chapter I. The initial section provides the overall results for each of the dimensionality tests across all conditions. The report is then organized into sections dealing with each of the five conditional elements that were manipulated for this simulation—namely, the actual number of factors, variable-to-factor ratio, pattern-magnitude interval, sample size, and inter-factor correlation. Each of these sections is further organized into subsections representing the number-of-factor techniques being evaluated. Specifically, the 13 different tests have been divided into four distinct families of methods: sequential Kaiser-Meyer-Olkin (SKMO) tests, classical methods, parallel analysis, and minimum average partial test. Within each section, the first subsection summarizes the performances of the four variants of the SKMO test, which allows for the consideration of the first research question. The remaining three subsections briefly summarize the other methods and compare the SKMO to those methods, which allows for consideration of the second research question. For convenience, Table 4.1 provides an index for all the number-of-factors tests with their abbreviations, families, and other specifications.

Table 4.1

Details of the Number-of-Factors Procedures under Investigation

Abbreviation	Family	Procedure Details	Criterion
SKMO-N50	Sequential Kaiser-Meyer-Olkin Tests	Noniterated communalities	.50 cutoff
SKMO-I50	Sequential Kaiser-Meyer-Olkin Tests	Iterated communalities	.50 cutoff
SKMO-N60	Sequential Kaiser-Meyer-Olkin Tests	Noniterated communalities	.60 cutoff
SKMO-I60	Sequential Kaiser-Meyer-Olkin Tests	Iterated communalities	.60 cutoff
PVE	Classical Methods	Proportion of variance extracted	75% cutoff
BCS	Classical Methods	Bartlett chi-square test	$\alpha = .01$
CNG	Classical Methods	Cattell-Nelson-Gorsuch scree test	
KG	Classical Methods	Kaiser-Guttman test	
PA-RM	Parallel Analysis	Reduced correlation matrices	Mean eigenvalue
PA-RP	Parallel Analysis	Reduced correlation matrices	95th percentile eigenvalue
PA-UM	Parallel Analysis	Unreduced correlation matrices	Mean eigenvalue
PA-UP	Parallel Analysis	Unreduced correlation matrices	95th percentile eigenvalue
MAP	Minimum Average Partial Test		

The accuracy, bias, and precision of each procedure was assessed using three indicators. Each of these performance indicators operates on the discrepancy between the number of factors that is predicted by a procedure for a given sample and the actual number of factors present in the population. First, the hit rate (accuracy) is the proportion of the samples for which a procedure correctly identified the true number of factors (i.e., the proportion wherein the discrepancy is zero). The discrepancy mean (bias) and discrepancy standard deviation (precision) are simply the mean and standard deviation of the discrepancy values, respectively.

Overall Performance

These results represent the performance of each dimensionality test aggregated across all possible conditions that were imposed in the Monte Carlo simulation. The overall results are presented in Table 4.2. Except for the figures given for the CNG, each value in Table 4.2 is based upon 67,500 observed samples. The CNG is based on 54,000 samples because it was excluded from the conditions with one or two true factors.

Table 4.2

Overall Performance Indicators

Procedure	Hit Rate	Discrepancy	
		M	*SD*
SKMO-N50	78%	-0.58	1.75
SKMO-I50	71%	-0.47	1.99
SKMO-N60	59%	-1.85	2.94
SKMO-I60	57%	-1.96	3.00
PVE	0%	11.76	8.94
BCS	22%	5.57	8.47
CNG	10%	-1.42	3.05
KG	43%	3.55	5.60
PA-RM	73%	0.47	1.15
PA-RP	86%	0.04	0.83
PA-UM	90%	-0.15	0.86
PA-UP	89%	-0.30	1.17
MAP	63%	-1.79	2.87

Sequential Kaiser-Meyer-Olkin Tests

The two SKMO variants using the .50 criterion consistently produced much more accurate and precise results than those using the .60 criterion. While the SKMO-N50 and SKMO-I50 had overall hit rates above 70% and discrepancy standard deviations below 2.00, the SKMO-N60 and SKMO-I60 displayed hit rates lower than 60% and standard deviations at or near 3.00. Further, the .50-criterion methods tended to underextract by well under one factor, where the .60-criterion methods tended to underextract by nearly two factors.

Surprisingly, the versions utilizing the noniterated communality estimation method (SKMO-N50 and SKMO-N60) performed better than their iterative counterparts (SKMO-I50 and SKMO-I60). While the difference in performance between SKMO-N60 and SKMO-I60 was slight, the SKMO-N50 performed noticeably more accurate than the I50.

Of the four SKMO variants, the SKMO-N50 had the best overall performance as it produced a very good hit rate (78%) and discrepancy standard deviation (1.75). The SKMO-N50 tended to underextract; however, this was only by a very small margin (-0.58).

Classical Methods

In short, PVE, BCS, CNG, and KG all performed poorly. Among these four methods, KG showed the best hit rate (43%), but it tended to overextract by more than three factors. KG also showed relatively poor precision with a discrepancy standard deviation well over 5.00. CNG had the best discrepancy mean (-1.42) and standard deviation (3.05) of the four classical tests, yet it had a very low hit rate (10%). BCS performed poorly with a hit rate of only 22% and a tendency to overextract by more than five factors. BCS also showed very poor precision with a standard deviation over 8.00. PVE was the worst of the four classical methods with a hit rate of 0% and a tendency to overextract by more than 11 factors. Precision for PVE was poor as well.

Overall, every version of the SKMO had a better hit rate than any of the classical methods by a wide margin. SKMO-I60, the poorest performing SKMO variant, had a hit rate of 57%, whereas

the best hit rate among the classical methods was only 43%. Both the SKMO-N50 and SKMO-I50 showed better precision and margins of bias than any of the classical methods.

Parallel Analysis

The various forms of parallel analysis performed extremely well. PA-RM, the poorest performer among the four variants of parallel analysis, still produced good results with an overall hit rate of 73%. PA-RP, PA-UM, and PA-UP all performed particularly well. PA-UM had the best overall hit rate (90%), yet PA-RP showed a slightly better discrepancy mean (0.04) and standard deviation (0.83). With a hit rate of 89%, PA-UP was only slightly less accurate than PA-UM. Although PA-RP had the best mean and standard deviation, both PA-UM and PA-UP had better hit rates. In general, the variants that used unreduced correlation matrices (PA-UM and PA-UP) were more accurate than their counterparts that used reduced correlation matrices (PA-RM and PA-RP). Further, the unreduced methods tend to underextract by a small margin, where the reduced methods have a slight tendency to overextract.

In general, the various forms of parallel analysis outperformed the SKMO methods. Nevertheless, SKMO-N50 had a better hit rate (78%) than PA-RM (73%), marking the only time in this study that a method outperformed any of the parallel-analysis variants. PA-RM did have a slightly better discrepancy mean and standard deviation than SKMO-N50. SKMO-I50, however, did have exactly the same margin of bias (in absolute magnitude) as PA-RM.

Minimum Average Partial Test

The MAP test had a respectable performance in terms of accuracy (*HR* = 63%) and precision (*SD* = 2.87). However, it tended to underextract by nearly two factors. The MAP test did outperform the two versions of SKMO using the .60 criterion by a slim margin. However, the two versions which used the .50 criterion were clearly superior to the MAP test in all aspects of overall performance.

Actual Number of Factors

This Monte Carlo was designed to have seven different possible conditions for the true number of factors—specifically, 1, 2, 3, 4, 5, 8, and 10 factors were used. With the exception of the one- and two-factor conditions, the performance indicators for each condition are based upon 10,800 samples. Because of the exclusion of the conditions with fewer than eight observed variables, the one-factor condition has 5,400 samples, and the two-factor condition has 8,100 samples. The performance indicators for each dimensionality test at each condition are presented in Table 4.3. The general trend for most of the number-of-factors tests in this study was a negative relationship between the number of actual factors and the performance criteria. That is, most tests performed better when there were fewer factors present, and the level of performance tended to diminish as the number of factors increased. This general pattern can be seen in Figures 4.1 and 4.2.

Table 4.3

Performance Indicators by the Actual Number of Factors

Procedure	Actual Number of Factors											
	1			2			3			4		
	HR	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*
SKMO-N50	77%	0.28	0.56	92%	0.09	0.35	94%	-0.04	0.31	89%	-0.16	0.53
SKMO-I50	60%	0.67	0.93	74%	0.45	0.94	86%	0.13	0.70	80%	0.01	0.76
SKMO-N60	100%	0.00	0.00	92%	-0.08	0.29	68%	-0.68	1.08	60%	-1.10	1.51
SKMO-I60	100%	0.00	0.00	92%	-0.09	0.30	67%	-0.70	1.08	59%	-1.14	1.53
PVE	0%	4.45	0.98	0%	6.55	2.41	0%	7.65	4.19	0%	9.64	5.49
BCS	37%	2.70	2.52	34%	4.21	4.42	28%	4.52	5.77	24%	5.36	7.19
CNG[a]							49%	0.58	0.62	1%	0.26	1.08
KG	71%	0.35	0.60	51%	1.02	1.33	51%	1.51	2.09	46%	2.34	3.02
PA-RM	93%	0.08	0.31	88%	0.15	0.48	82%	0.22	0.59	77%	0.32	0.77
PA-RP	99%	0.01	0.07	98%	0.02	0.16	94%	0.01	0.30	91%	0.02	0.43
PA-UM	100%	0.00	0.04	99%	0.01	0.10	96%	-0.01	0.23	94%	-0.04	0.36
PA-UP	100%	0.00	0.00	100%	0.00	0.06	95%	-0.07	0.33	92%	-0.13	0.52
MAP	98%	-0.02	0.14	82%	-0.29	0.65	62%	-0.95	1.28	60%	-1.31	1.70

Procedure	Actual Number of Factors								
	5			8			10		
	HR	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*
SKMO-N50	82%	-0.32	0.81	63%	-1.22	2.12	51%	-2.09	3.09
SKMO-I50	78%	-0.21	0.95	63%	-1.25	2.28	51%	-2.28	3.27
SKMO-N60	54%	-1.59	1.99	38%	-3.35	3.38	27%	-4.80	4.27
SKMO-I60	53%	-1.66	1.99	36%	-3.56	3.37	25%	-5.11	4.23
PVE	0%	11.66	6.64	0%	17.10	9.79	0%	20.28	11.73
BCS	20%	6.03	8.37	12%	7.08	10.95	9%	7.33	11.99
CNG	1%	-0.34	1.59	1%	-3.05	2.87	1%	-4.54	3.73
KG	40%	3.21	3.95	31%	6.09	6.81	26%	8.12	8.71
PA-RM	71%	0.44	0.94	60%	0.78	1.50	55%	1.00	1.84
PA-RP	86%	0.04	0.58	74%	0.06	1.12	67%	0.08	1.56
PA-UM	91%	-0.08	0.51	82%	-0.28	1.13	76%	-0.50	1.65
PA-UP	89%	-0.22	0.75	80%	-0.58	1.55	75%	-0.89	2.15
MAP	58%	-1.71	2.12	54%	-3.06	3.50	52%	-3.93	4.41

[a]The minimum value for the CNG procedure is three, so it was excluded from the one- and two-factor conditions in this simulation.

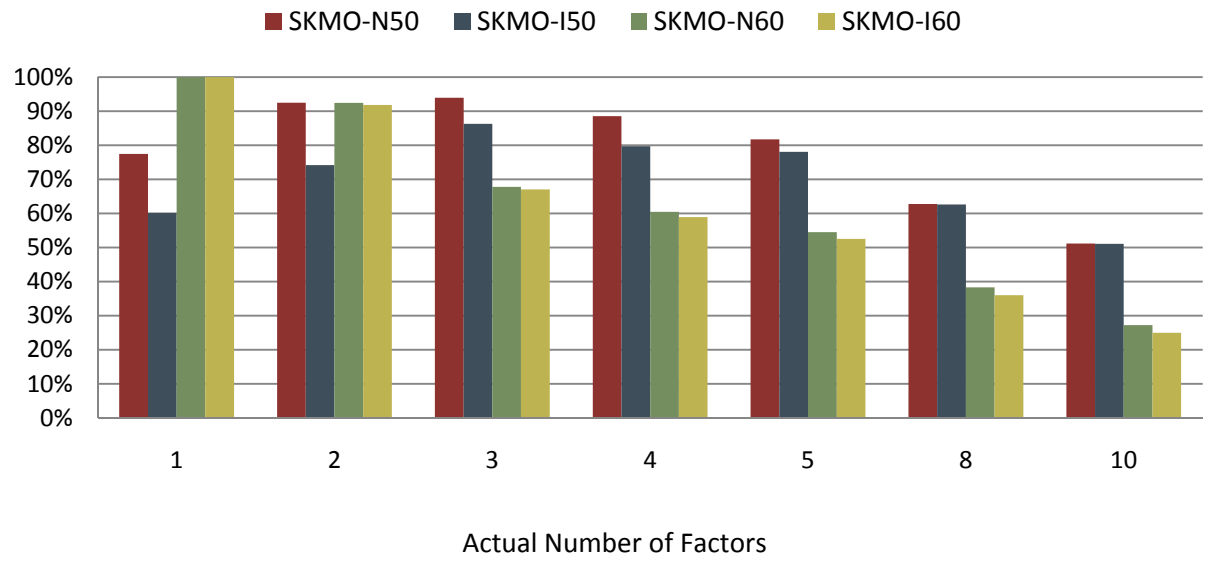

Figure 4.1. Hit rates for the four SKMO variants by the actual number of factors.

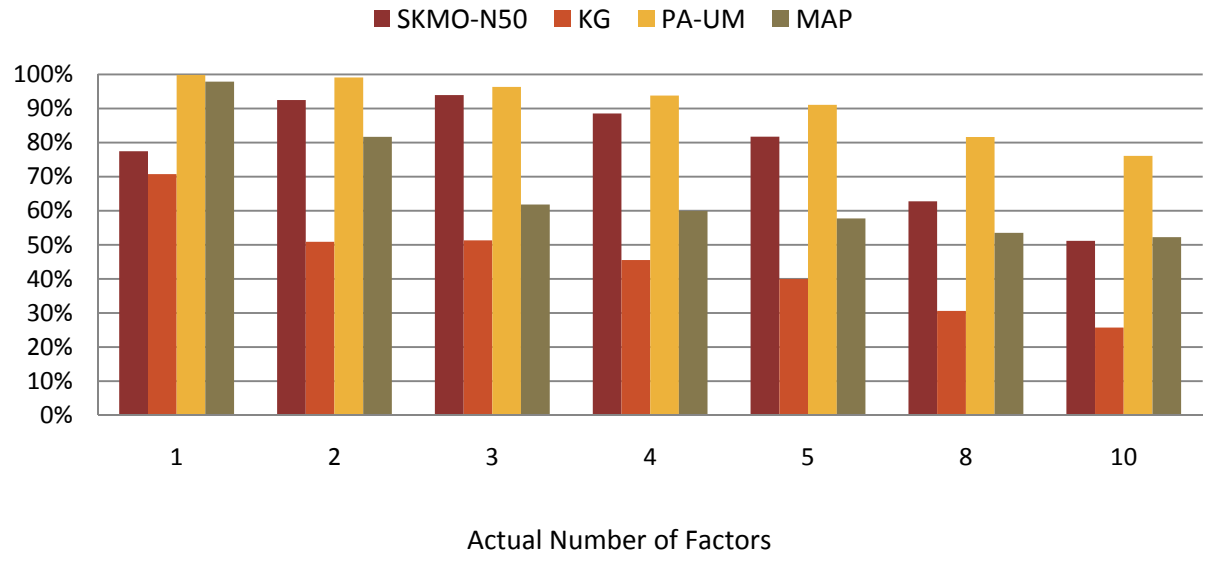

Figure 4.2. Hit rates for the top-performing procedures by the actual number of factors.

Sequential Kaiser-Meyer-Olkin Tests

The two SKMO variants using the .50 criterion deviated slightly from the general performance pattern. As Figure 4.1 shows, the SKMO-N50 had a relatively poorer hit rate at the

one-factor condition (77%), but it improved and stabilized across the two- to five-factor conditions (92%, 94%, 89%, and 82%, respectively). Also, the SKMO-N50 had a slight tendency to overextract at the one- and two-factor conditions. This bias became negative at three factors and increased in magnitude with each subsequent condition. Although SKMO-I50 did not perform nearly as well as SKMO-N50, the former followed the same basic trend as the latter. Like all other methods in this study, the accuracy and precision of the SKMO-N50 and SKMO-I50 were dramatically decreased in the eight- and ten-factor conditions.

The two versions of SKMO that employed the .60 criterion followed the general monotonic trend, performing best at one factor and worst at ten factors. Other than at the one-factor condition, the SKMO-N50 was observed to be clearly superior to both of the .60-criterion versions. Although SKMO-N60 and SKMO-I60 both had flawless performances at the one-factor level, their performances quickly diminished as the true number of factors increased.

Classical Methods

KG consistently had a better hit rate than PVE, BCS, and CNG across all levels of the actual number of factors. Although KG began with a good hit rate with a single factor (71%), its performance quickly diminished with each additional factor. KG also maintained a tendency to overextract at each level. CNG maintained decent discrepancy means and standard deviations across these conditions, but it had terrible hit rates (most at 1%). The lone exception was the 49% hit rate at the three-factor mark.

The SKMO-N50 and SKMO-I50 were clearly more accurate and precise than any of the classical tests across these conditions. The only exception was that KG outperformed SKMO-I50 when there was a single factor present.

Parallel Analysis

All parallel-analysis variants had excellent accuracy and precision when there were few factors—in fact, PA-RP, PA-UM, and PA-UP were virtually perfect when dealing with one and two

factors. All versions followed the general monotonic trend of decreased performance with increased number of factors. However, the hit rates for PA-UM and PA-UP only dropped to 76% and 75%, respectively, at the 10-factor condition.

Most versions of parallel analysis performed better than any of the SKMO methods. However, the SKMO-N50 performed nearly as well as PA-RP, PA-UM, and PA-UP under the two- to five-factor conditions. Except at the extreme conditions (one and ten factors), SKMO-N50 was more accurate and precise than PA-RM.

Minimum Average Partial Test

The MAP test was nearly perfect at the single-factor condition and very good at the two-factor condition. However, the performance of the MAP declined steadily with additional factors, dropping to a 60% hit rate at only four factors.

The SKMO-N50 surpassed the MAP under almost all levels of the true number of factors. One exception was the better accuracy and precision of the MAP when there was only a single factor. MAP and SKMO-N50 had virtually the same hit rate at 10 factors, but the margin of bias for the MAP nearly doubled that of the SKMO-N50 under this condition.

Variable-to-Factor Ratio

This simulation allowed for four different ratios of the number of observed variables to the number of actual factors: 3:1, 5:1, 8:1, and 10:1. The conditions defined by the larger factor-to-variable ratios (8:1 and 10:1) both had 18,900 samples. Because of the exclusion of the condition combinations that result in fewer than eight observed variables, the 3:1 condition has 13,500 samples, and the 5:1 condition has 16,200 samples. The basic results for each level of variable-to-factor ratio are presented in Table 4.4 and Figures 4.3 and 4.4. As these results show, accuracy and precision of a dimensionality test typically become better as the variable-to-factor ratio increases. The most dramatic improvement in performance generally occurs between the 3:1 and 5:1 levels.

Table 4.4

Performance Indicators by Variable-to-Factor Ratio

Procedure	Variable-to-Factor Ratio											
	3:1			5:1			8:1			10:1		
	HR	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*
SKMO-N50	60%	-1.26	2.23	81%	-0.50	1.57	83%	-0.35	1.52	82%	-0.40	1.60
SKMO-I50	63%	-0.97	2.44	70%	-0.33	1.90	74%	-0.30	1.79	75%	-0.39	1.85
SKMO-N60	27%	-3.84	3.31	58%	-1.73	2.86	70%	-1.20	2.50	71%	-1.19	2.49
SKMO-I60	28%	-3.90	3.34	57%	-1.86	2.94	68%	-1.31	2.57	68%	-1.29	2.56
PVE	0%	4.37	2.09	0%	8.78	4.59	0%	13.84	8.36	0%	17.50	10.51
BCS	14%	1.61	3.79	19%	4.94	6.65	25%	6.81	9.25	27%	7.71	10.25
CNG	12%	-1.54	3.09	10%	-1.43	3.05	9%	-1.36	3.04	9%	-1.34	3.03
KG	64%	0.52	0.83	43%	2.08	2.69	38%	4.32	5.58	32%	6.22	7.61
PA-RM	60%	0.56	1.35	68%	0.61	1.31	79%	0.41	1.06	82%	0.33	0.89
PA-RP	72%	-0.20	1.44	84%	0.09	0.77	90%	0.10	0.51	92%	0.09	0.42
PA-UM	72%	-0.63	1.61	90%	-0.10	0.71	96%	0.00	0.33	97%	0.02	0.25
PA-UP	66%	-1.11	2.11	89%	-0.25	0.99	97%	-0.05	0.43	98%	-0.02	0.29
MAP	6%	-5.31	2.79	61%	-1.93	2.94	80%	-0.69	1.74	90%	-0.27	1.00

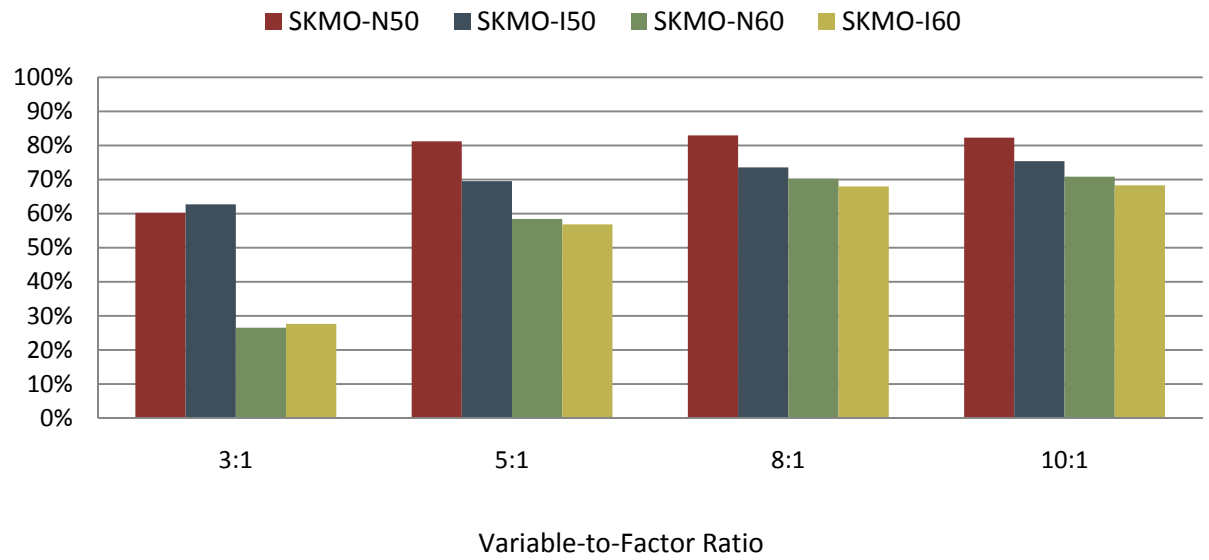

Figure 4.3. Hit rates for the four SKMO variants by variable-to-factor ratio.

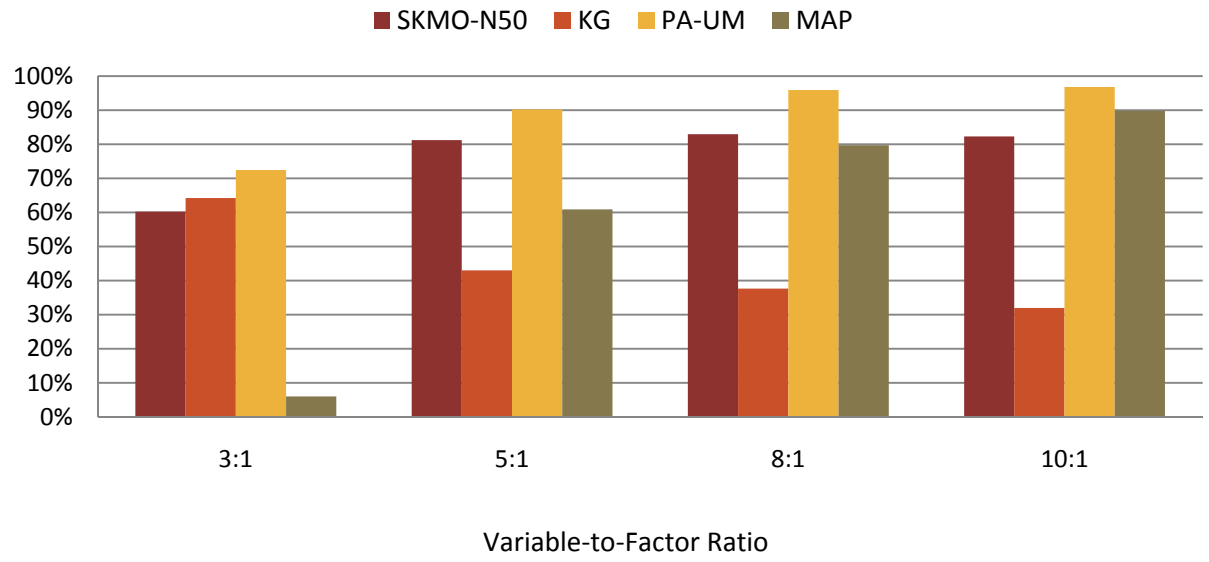

Figure 4.4. Hit rates for the top-performing procedures by variable-to-factor ratio.

Sequential Kaiser-Meyer-Olkin Tests

The .50-criterion versions were consistently better than the .60-criterion versions regarding accuracy, precision, and bias at all four levels of the factor-to-variable ratio. This was particularly noticeable at the 3:1 condition. The noniterated procedures usually performed better than the iterated versions at each level. The only exception to this was at the 3:1 level, but the differences in performance were very slight.

Although SKMO-N50 was typically the best performer among the SKMO variants, the SKMO-I50 had a small advantage in accuracy over SKMO-N50 at the 3:1 level. The SKMO-N50 also displayed a noteworthy characteristic: Like all other dimensionality tests, the SKMO-N50 began with relatively low accuracy (60%) at the 3:1 level. However, the performance of the SKMO-N50 improved greatly at the 5:1 level (81%), and this performance remained practically constant across the subsequent levels of the factor-to-variable ratio (see Figures 4.3 and 4.4).

Classical Methods

The KG method was consistently more accurate than PVE, BCS, and CNG across all four factor-to-variable conditions. KG began at the 3:1 condition with a respectable 64% hit rate, but its performance steadily deteriorated as the factor-to-variable ratio increased. This was contrary to the general performance trend of the other number-of-factors tests.

With a few exceptions, every version of SKMO outperformed every classical method under all four conditions. The most notable of these exceptions occurred at the 3:1 level where KG was slightly more accurate and much more precise than SKMO-N50 and SKMO-I50.

Parallel Analysis

All four versions of parallel analysis followed the general trend of improved performance as the factor-to-variable ratio increased. The unreduced methods generally performed better than the reduced methods at each level. PA-UM and PA-UP were particularly accurate and precise at the 8:1 and 10:1 levels, producing hit rates above 95% and standard deviations less than 0.50. PA-UM and PA-UP typically performed about the same at each level except at 3:1 where PA-UM was noticeably better.

The SKMO-N50 was just as accurate as or better than PA-RM under each condition. The remaining three versions of parallel analysis were more accurate and precise than SKMO-N50 at each level. However, this was never by any extreme margin. In fact, there were instances where the performance of the SKMO-N50 came very close (e.g., SKMO-N50 and PA-UP at 3:1).

Minimum Average Partial Test

The MAP performed terribly at the 3:1 condition, where it had only a 6% hit rate, and it tended to underextract by more than five factors. However, the MAP steadily improved as the variable-to-factor ratio increased, reaching a 90% hit rate at the 10:1 level. An inspection of Table 4.4 and Figure 4.4 shows that the MAP is strongly affected by the variable-to-factor ratio—more so than any other dimensionality test.

The SKMO-N50 had a markedly better performance than the MAP at 3:1 and 5:1, and the SKMO-N50 was slightly better than the MAP at 8:1. However, the MAP was slightly better at the 10:1 level. The most important difference between the SKMO-N50 and the MAP is the fact the variable-to-factor ratio has a strong influence on the performance of the MAP, where this is only minimal for the SKMO-N50.

Pattern-Magnitude Interval

Three different pattern-magnitude intervals were used in this study. For any given population pattern matrix, all loadings were within a predetermined interval. The low level allowed for population pattern coefficients to range from .3 to .5, the wide level from .3 to .8, and the high level from .6 to .8. These results are based on 22,500 samples (18,000 samples for the CNG) per condition. The most common trend across these three conditions was for a procedure to perform relatively poorly with low saturation, somewhat better with the wide range, and even better with the high-saturation levels. The results are presented in Table 4.5 and Figures 4.5 and 4.6.

Table 4.5

Performance Indicators by Pattern-Magnitude Interval

Procedure	Pattern-Magnitude Interval								
	Low			Wide			High		
	HR	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*
SKMO-N50	53%	-1.57	2.63	85%	-0.20	0.86	96%	0.03	0.27
SKMO-I50	49%	-1.46	2.88	78%	-0.11	1.28	86%	0.17	0.65
SKMO-N60	24%	-3.77	3.37	62%	-1.63	2.74	91%	-0.16	0.72
SKMO-I60	23%	-3.83	3.36	59%	-1.80	2.86	90%	-0.24	1.01
PVE	0%	14.95	10.53	0%	11.82	8.37	0%	8.49	6.15
BCS	55%	-1.09	1.81	0%	12.89	9.92	11%	4.92	3.90
CNG	12%	-1.60	3.07	11%	-1.39	3.04	8%	-1.26	3.04
KG	10%	6.72	7.14	27%	3.66	4.59	91%	0.28	1.15
PA-RM	47%	0.97	1.66	74%	0.42	0.86	100%	0.00	0.05
PA-RP	68%	-0.01	1.36	89%	0.13	0.46	100%	0.00	0.01
PA-UM	78%	-0.28	1.28	94%	-0.13	0.71	99%	-0.03	0.25
PA-UP	77%	-0.64	1.69	92%	-0.22	0.97	98%	-0.04	0.36
MAP	30%	-3.11	3.04	76%	-1.32	2.66	84%	-0.94	2.38

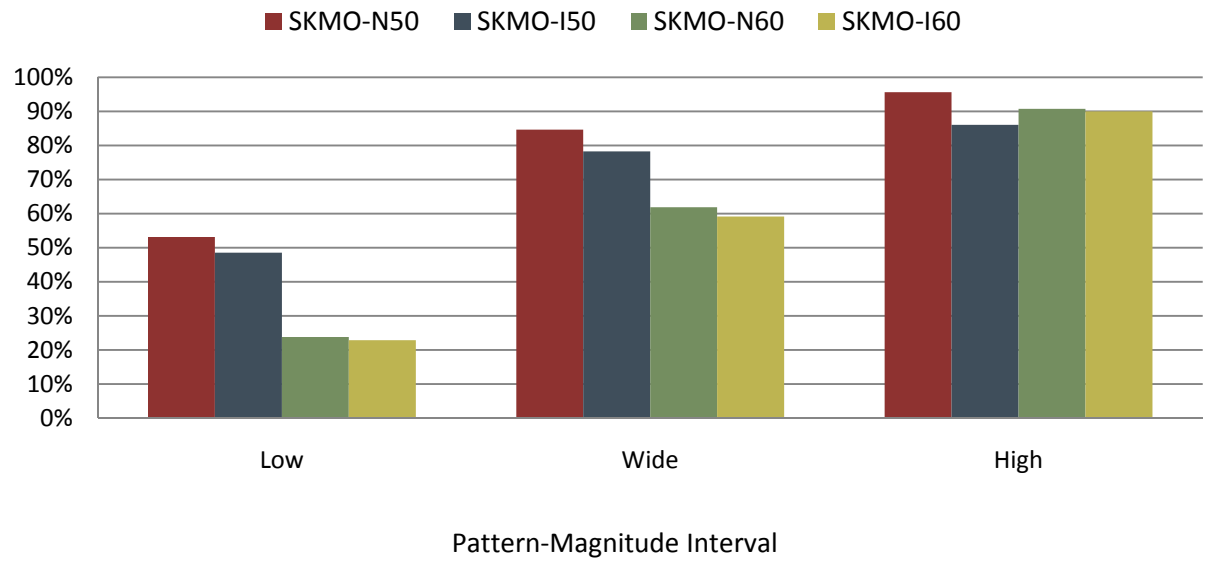

Figure 4.5. Hit rates for the four SKMO variants by pattern-magnitude interval.

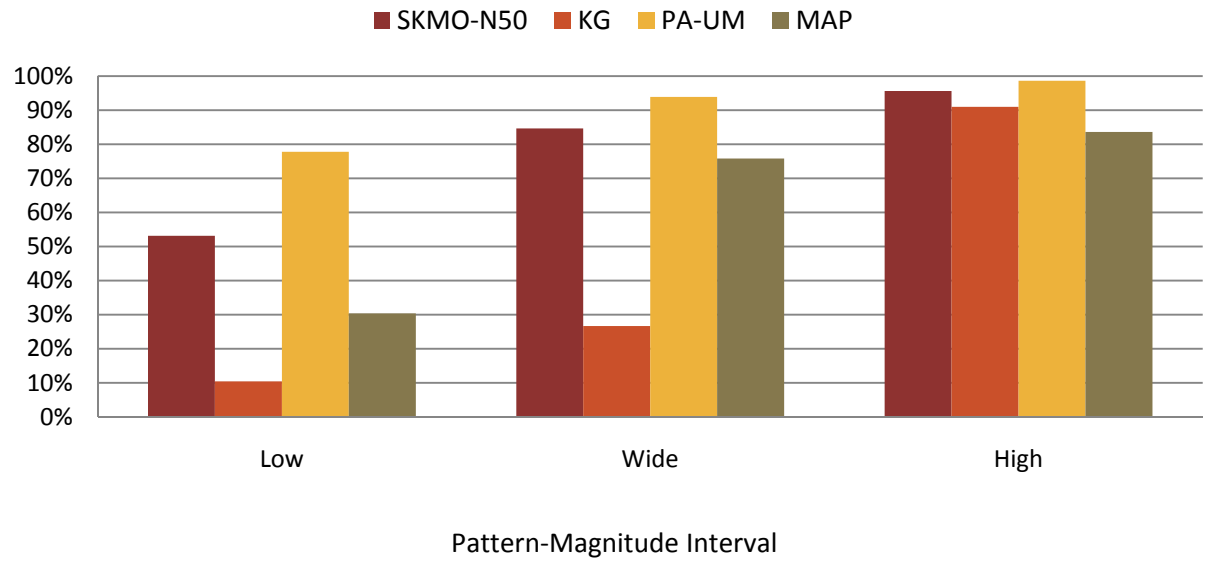

Figure 4.6. Hit rates for the top-performing procedures by pattern-magnitude interval.

Sequential Kaiser-Meyer-Olkin Tests

The .50-criterion versions of SKMO were generally more accurate and precise than their .60-criterion counterparts across all three levels of pattern magnitude; particularly noticeable was the wide gap in performance at the low level. The noniterated SKMO tests both performed slightly better than their iterated counterparts at each saturation level. The differences in performance among the four versions became less pronounced at the high-saturation condition. However, with a hit rate of 96% and standard deviation of 0.27, SKMO-N50 was clearly superior at the high level. In fact, SKMO-N50 performed the best under all three conditions. Particularly noteworthy is that the SKMO-N50 had a very good hit rate (85%) and standard deviation (0.86) at the wide-saturation level. Also, all four SKMO variants almost always had a slight tendency to underextract. The only exception is that SKMO-N50 and SKMO-I50 had a small positive bias at the high-saturation level.

Classical Methods

The BCS demonstrated a rather peculiar behavior as it showed a relatively good hit rate (55%) at the low-saturation condition but became progressively less accurate and precise at the

wide- and high-magnitude levels. Among the classical methods, KG showed the best performance under the wide- and high-saturation conditions. In fact, KG produced an impressive 91% hit rate under high saturation.

Although the BCS has a slight edge on SKMO-N50 at the low level (55% to 53%), the SKMO-N50 was far superior under the wide- and high-magnitude conditions. Under the high-saturation condition, KG performed as well or better than all SKMO variants except for SKMO-N50. The SKMO-N50 had a slightly better hit rate and a much better standard deviation than KG at the high level.

Parallel Analysis

All versions of parallel analysis followed the general trend with better performances at the high level and relatively poorer performances at the low level. In fact, all four versions were virtually perfect with high saturation. Among the parallel analysis variants, PA-UM and PA-UP were the strongest performers at the low and wide conditions, while PA-RM and PA-RP had a very slight edge at the high level.

SKMO-N50 was better than PA-RM at the low and wide levels and was very close at the high level. Although SKMO-N50 did not beat any of the other parallel analysis variants, it did come very close, particularly with wide- and high-magnitude intervals.

Minimum Average Partial Test

The MAP test showed very poor performance under the low-saturation condition with a hit rate of only 30%, but it had very good accuracy at the wide (76%) and high (84%) levels. Its precision did not improve by any great margin, however.

The SKMO-N50 was consistently better than MAP across all three levels with respect to accuracy, precision, and margin of bias. Although not as good as SKMO-N50, the SKMO-I50 also beat the MAP at each level by slight margin.

Sample Size

The Monte Carlo simulation allowed for three different sample sizes. The generated samples consisted of either 250 (small), 500 (medium), or 750 (large) observations for each observed variable. There were 22,500 samples (18,000 for the CNG) per condition. Typically, a number-of-factors test will become more accurate and precise as the sample size increases, with the greatest improvement usually occurring between the small and medium sample sizes.

Table 4.6

Performance Indicators by Sample Size

Procedure	Sample Size								
	250			500			750		
	HR	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*
SKMO-N50	67%	-1.12	2.47	81%	-0.42	1.34	86%	-0.20	0.92
SKMO-I50	60%	-1.18	2.70	75%	-0.27	1.61	78%	0.04	1.10
SKMO-N60	47%	-2.41	3.18	61%	-1.70	2.86	68%	-1.45	2.69
SKMO-I60	44%	-2.61	3.25	60%	-1.77	2.89	67%	-1.49	2.72
PVE	0%	10.66	7.59	0%	12.02	9.14	0%	12.59	9.83
BCS	21%	1.93	4.33	20%	6.10	7.84	25%	8.69	10.58
CNG	11%	-1.55	3.05	10%	-1.37	3.05	9%	-1.32	3.05
KG	32%	4.37	6.02	44%	3.42	5.55	52%	2.87	5.07
PA-RM	67%	0.59	1.45	74%	0.48	1.09	79%	0.33	0.81
PA-RP	81%	-0.07	1.13	86%	0.09	0.74	90%	0.09	0.48
PA-UM	82%	-0.24	1.15	92%	-0.12	0.77	96%	-0.08	0.56
PA-UP	81%	-0.52	1.53	91%	-0.23	1.03	95%	-0.14	0.80
MAP	61%	-1.81	2.83	64%	-1.78	2.88	65%	-1.78	2.90

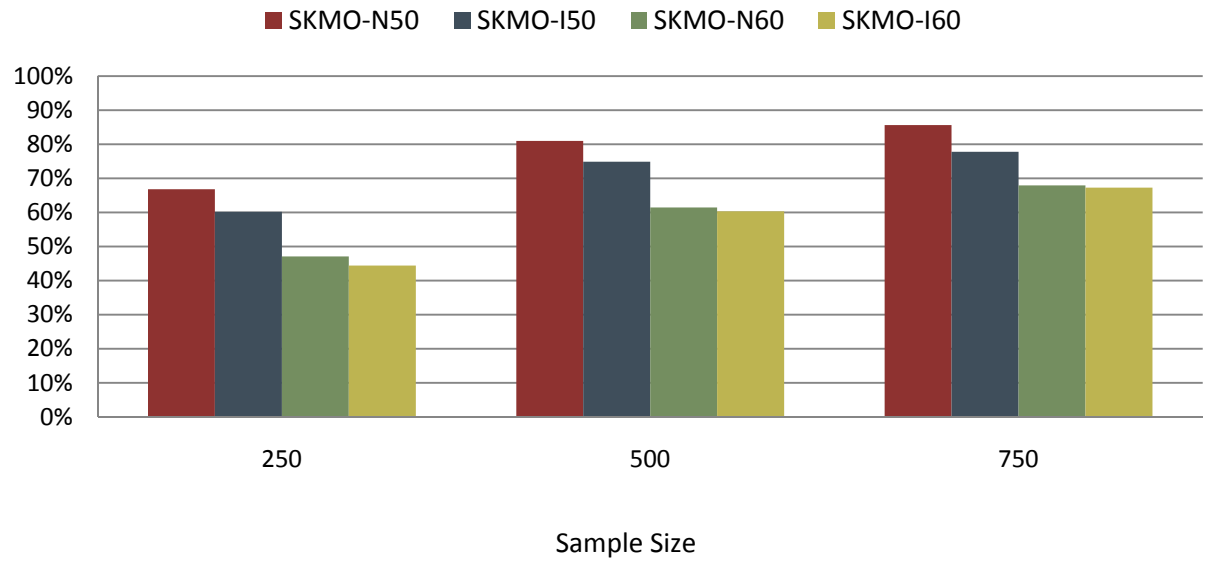

Figure 4.7. Hit rates for the four SKMO variants by sample size.

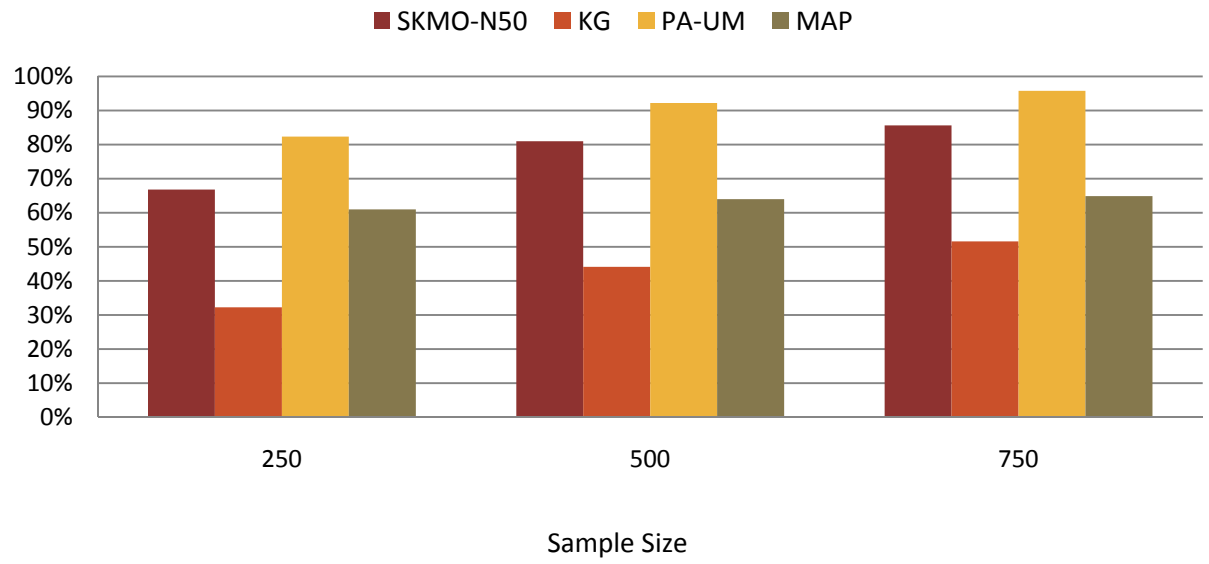

Figure 4.8. Hit rates for the top-performing procedures by sample size.

Sequential Kaiser-Meyer-Olkin Tests

All SKMO variants followed the general trend with improved performances as the sample size increased. The .50-criterion versions were consistently better than the .60-criterion versions

across all levels, and the noniterated forms always performed slightly better than the iterated at each level. All SKMO variants almost always have a negative bias at each level, with the lone exception being SKMO-I50 at the large-sample condition. Of the four versions, SKMO-N50 consistently had the best performance at each level.

Classical Methods

KG is by far the most accurate of the four classical methods at each possible sample size. However, its best performance (under the large-sample condition) was accurate only 52% of the time. KG also tended to overextract by nearly three factors at the large-sample condition. All versions of the SKMO are considerably more accurate, more precise, and have a smaller margin of bias than KG across all three sample-size conditions.

Parallel Analysis

All versions of parallel analysis follow the general trend of improved performance as the sample size increased, and all versions perform quite well at each level. The unreduced versions typically perform better than their reduced counterparts, although PA-RP performs just as well as PA-UM and PA-UP at the small-sample condition. PA-UM performs virtually the same as PA-UP at each level.

SKMO-N50 was as accurate as PA-RM at the small sample size and more accurate at the medium and large sample sizes. The remaining three versions of parallel analysis were more accurate and precise than SKMO-N50 at each level. However, the performance of SKMO-N50 was closer to that of PA-RP, PA-UM, and PA-UP than any other method.

Minimum Average Partial Test

The performance of the MAP test was fairly consistent across all levels. In particular, the hit rate was in the 60% to 65% range under all three conditions. Also, the MAP tended to underextract by nearly two factors regardless of sample size.

Only the versions of SKMO using the .50 criterion performed as well or better than the MAP test. The SKMO-N50 performed better than the MAP at all three levels. Although the SKMO-I50 was slightly less accurate than the MAP under the small-sample condition, the former performed better than the latter with larger sample sizes.

Inter-Factor Correlation

The population correlations among the factors assumed three different levels: no correlation (= .00), moderate correlation (= .20), and high correlation (= .40). All factor correlations under a given condition were set to the same value. There were 22,500 samples (18,000 for the CNG) per condition. As seen in Table 4.7 and Figures 4.9 and 4.10, the common trend in performance for these procedures was a relative decline in accuracy as the factors become more oblique. The relative precision of most procedures is only slightly reduced as the inter-factor correlations become larger.

Table 4.7

Performance Indicators by Inter-Factor Correlation

Procedure	Inter-Factor Correlation								
	.00			.20			.40		
	HR	*M*	*SD*	*HR*	*M*	*SD*	*HR*	*M*	*SD*
SKMO-N50	83%	-0.40	1.55	78%	-0.53	1.66	72%	-0.81	1.99
SKMO-I50	75%	-0.30	1.77	71%	-0.40	1.90	67%	-0.70	2.25
SKMO-N60	65%	-1.68	2.96	60%	-1.80	2.93	51%	-2.07	2.93
SKMO-I60	63%	-1.79	3.01	59%	-1.91	2.99	50%	-2.17	2.97
PVE	0%	11.78	8.95	0%	11.76	8.94	0%	11.73	8.93
BCS	26%	5.73	8.37	23%	5.62	8.43	18%	5.37	8.62
CNG	1%	0.92	1.29	9%	-2.17	3.31	20%	-3.00	2.61
KG	42%	3.60	5.62	42%	3.58	5.61	44%	3.48	5.56
PA-RM	73%	0.54	1.21	73%	0.50	1.13	73%	0.36	1.10
PA-RP	87%	0.15	0.62	86%	0.09	0.68	84%	-0.13	1.09
PA-UM	94%	0.04	0.34	93%	-0.04	0.47	83%	-0.44	1.33
PA-UP	96%	-0.05	0.51	92%	-0.16	0.76	79%	-0.70	1.74
MAP	70%	-1.62	2.93	66%	-1.61	2.72	53%	-2.14	2.91

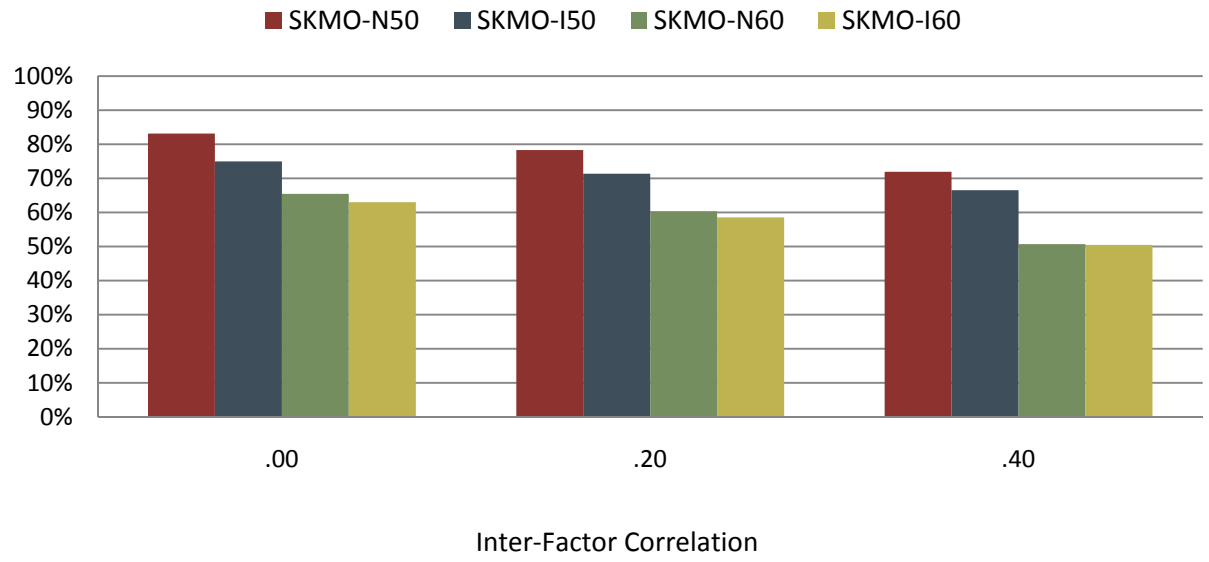

Figure 4.9. Hit rates for the four SKMO variants by inter-factor correlation.

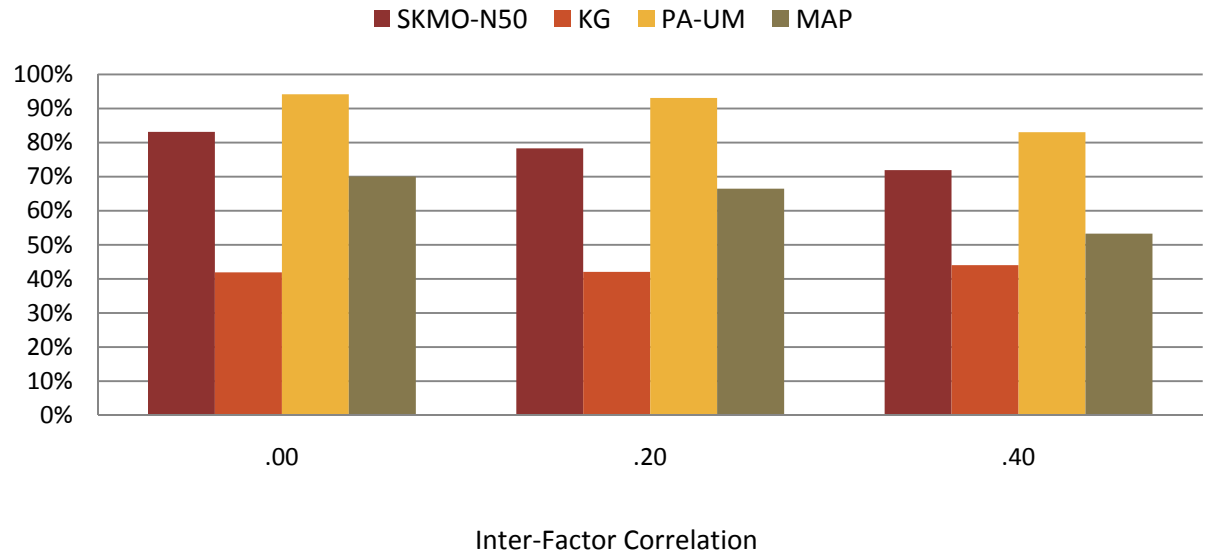

Figure 4.10. Hit rates for the top-performing procedures by inter-factor correlation.

Sequential Kaiser-Meyer-Olkin Tests

All SKMO variants followed the general trend, becoming less accurate as the inter-factor correlations increased in magnitude. All SKMO had a tendency to underextract at every level. The

SKMO-N50 was clearly the best SKMO variant at each of the three conditions: It never dropped below a 70% hit rate or exceeded a standard deviation of 2.00.

The .50-criterion versions of the SKMO performed noticeably better than the .60-criterion versions at each level of inter-factor correlation. SKMO-N50 and SKMO-I50 had consistently higher hit rates and smaller standard deviations than SKMO-N60 and SKMO-I60 under each condition. With the exception of SKMO-I50 at the high-correlation condition, the .50-criterion methods always had hit rates above 70% and standard deviations below 2.00. In contrast, the .60-criterion versions were consistently less than 70% accurate and had standard deviations at or very close to 3.00. Further, the SKMO-N50 and SKMO-I50 had considerably smaller margins of bias than SKMO-N60 and SKMO-I60 under each condition. Neither of the .50-criterion methods ever produced a mean discrepancy greater than 1.00 in absolute magnitude, where the .60-criterion methods always had a mean well above 1.00.

Although the differences were sometimes slight, the noniterated versions always had better accuracy and precision than their iterated counterparts under each condition. The accuracy and precision of the SKMO-N50 was noticeably better than the SKMO-I50 under each condition, but the SKMO-N60 and SKMO-I60 were virtually identical in terms of performance at each level.

Classical Methods

At each level of factor correlation the KG procedure was by far the best of the four classical methods. However, the KG still performed poorly under every condition as it never produced a hit rate above 50%, always had a standard deviation greater than 5.00, and tended to overextract by more than three factors. Nevertheless, KG was observed to be unaffected by the different levels of factor correlation as all aspects of its performance remained virtually constant across all three conditions. The PVE, BCS, and CNG all performed quite poorly at all three levels.

The entire SKMO family was superior to all of the classical methods. The poorest performing SKMO variant easily surpassed the best of the classical methods with respect to all aspects of performance at each level of factor correlation.

Parallel Analysis

All four members of the parallel analysis family performed quite well at each level. None of the four variants had a hit rate under 70%, a standard deviation above 2.00, or a mean above 1.00 (in absolute magnitude) under any of the three conditions. PA-UM and PA-UP showed hit rates above 90% under the zero and moderate inter-factor correlation conditions. However, the accuracy and precision of these unreduced variants decreased dramatically at the high level. Although they did not typically perform as well as their unreduced counterparts, the versions of parallel analysis using reduced correlation matrices (PA-RM and PA-RP) were observed to be quite resilient to changes in inter-factor correlations, producing remarkably consistent results across the conditions.

In general, the parallel analysis procedures were generally more accurate and precise than the SKMO-N50 at all three levels. However, there are notable exceptions. First, the SKMO-N50 was just as accurate (if not more so) than PA-RM under each condition. In addition, SKMO-N50 was more accurate than PA-RP at the zero-correlation level.

Minimum Average Partial Test

The MAP test had a decent performance at each level of inter-factor correlation. The MAP had a fairly good accuracy at the zero-correlation level (70%), but steadily declined as the inter-factor correlations increased in magnitude. The MAP tends to underextract by just over 1.5 factors at the first two levels, but this jumps to over 2 factors at the .40 level. The MAP shows little change in its precision with a standard deviation just under 3.00 in each of the conditions.

Both the SKMO-N50 and SKMO-I50 outperform the MAP test with respect to accuracy, precision, and bias at each level of inter-factor correlation. Particularly noteworthy was that under

the high inter-factor correlation condition, the SKMO-N50 had a hit rate of 72% where the MAP had only 53%. Further, the SKMO-N50 tended to underextract by less than one factor, but the MAP test underextracted by over two factors at that extreme condition.

CHAPTER V

DISCUSSION

The purpose of this study was to determine the basic performance characteristics of the sequential Kaiser-Meyer-Olkin (SKMO) procedure, a new number-of-factors test designed specifically for use with the common-factor model. This study also sought to compare the SKMO to other well-known dimensionality tests. Both goals were accomplished by using an extensive Monte Carlo simulation where five different conditional elements were manipulated—specifically, the actual number of factors, variable-to-factor ratio, sample size, pattern-magnitude interval, and inter-factor correlation (see Table 3.2 for a summary of the parameter values for each conditional element). Thirteen different dimensionality tests (including four variants of the SKMO) were examined under each possible combination of conditions (see Table 4.1 for a complete listing of the tests). A full report of the accuracy, precision, and bias regarding the estimated number of factors to retain for each dimensionality procedure under the various conditions was presented in Chapter IV.

This final chapter provides a discussion of the interpretation and implications of the results of the study. Beginning with the major findings, the results are discussed in the context of the research questions, including tentative recommendations for the use of the SKMO. The major limitations of this study are also discussed, followed by suggestions for future research.

Major Findings

The major findings from the results of this study are discussed below. The presentation of this information is organized around the two research questions. For convenience, each research question will be reviewed at the beginning of its corresponding discussion.

Research Question 1

The first research question for this study inquires about the performance characteristics of the SKMO procedure with respect to the accuracy, bias, and precision of the predicted number of factors to retain. The following discussion will investigate this question primarily from two perspectives. First, this question is discussed in terms of the criterion values used in the SKMO procedure (.50 and .60). The discussion will then focus on the communality estimation methods used in this new test (noniterated and iterated).

Criterion values: .50 versus .60. Overall, the SKMO variants that use .50 as the cutoff criterion (SKMO-N50 and SKMO-I50) were decidedly more accurate than the variants that use .60 (SKMO-N60 and SKMO-I60) when determining the correct number of factors to retain. In particular, SKMO-N50 had an overall hit rate of 78%, where SKMO-N60 was only 59%. The overall precision and margin of bias were clearly superior in the .50-criterion versions as well. Furthermore, SKMO-N50 and SKMO-I50 were almost always better than SKMO-N60 and SKMO-I60 under each of the various conditions imposed in the simulation.

A critical consideration in the comparison of the .50 and .60 cutoff values is the magnitude of bias associated with each. Underextraction represents an omission of information which can distort a factor solution, rendering it uninterpretable. The level of distortion depends on the magnitude of the misspecification; more specifically, underextracting by only one factor does not alter a solution by any great degree, but deterioration of a factor solution becomes increasingly severe when underextracting by two or more factors (Fava & Velicer, 1996). All SKMO variants have a tendency to underextract, but the margins of bias were quite small for SKMO-N50 and

SKMO-I50 (M = -0.58 and -0.47, respectively). In stark contrast, both of the .60-criterion versions tended to underextract by nearly two factors (M = -1.85 and -1.96). These results provide a particularly strong argument for the use of the .50 cutoff.

These results are consistent with the original rationale for the KMO (see Cerny & Kaiser, 1970; Kaiser, 1981; Kaiser & Rice, 1974) as well as the findings from the Monte Carlo performed by Shirkey and Dziuban (1976). However, this does stand in contrast with the contemporary use of .60 when performing a conventional KMO to determine the factorability of a correlation matrix.

The most prominent exception to this general profile of the SKMO criterion values occurred with the number of true factors at the single-factor condition, where SKMO-N60 and SKMO-I60 both performed flawlessly. In contrast, SKMO-N50 and SKMO-I50 had hit rates of only 77% and 60%, respectively, under the same condition. This was the only instance in the entire simulation where either of the .60-criterion versions was more accurate than SKMO-N50. The performance of SKMO-N50 improved at the two-factor condition (92%), which was the same as the hit rates for SKMO-N60 and SKMO-I60. Beyond the two-factor level, SKMO-N50 and SKMO-I50 were clearly superior.

Related to this exception, the SKMO-N50 and SKMO-I50 both showed an unusual trend with respect to the true number of factors in the population. Where most other procedures produced the greatest hit rate at the single-factor level with a steady, monotonic decline as the number of factors increased, the SKMO-N50 and SKMO-I50 did not reach a maximum until the three-factor level. SKMO-N50 had a very slight positive bias under one and two factors; this bias became negative at three factors and consistently grew in magnitude as the number of factors increased. SKMO-I50 behaved similarly with its bias becoming negative at four factors.

Both of these issues can be better understood with anecdotal evidence from the preliminary plasmode trials. When dealing with data from a population characterized by a single-factor structure, the first SKMO value (which is equivalent to the conventional KMO) is typically large, well

above .60, and the second SKMO value (after partialling out the first factor) is also somewhat elevated (usually in the .55 to .57 range). In this situation, use of the .50 criterion would cause the test to overextract, where the .60 cutoff would correctly indicate a single factor. Again, this is conjecture based on anecdotal evidence, and a theoretical justification for this anomaly has not yet been formulated. Capitalization on chance would be a reasonable initial hypothesis, however.

Communality estimation methods: Noniterated versus iterated. Cumulatively, the noniterated versions of SKMO consistently performed better than their iterated counterparts, though not by any great degree. The more conspicuous difference occurred between the overall hit rates of SKMO-N50 (78%) and SKMO-I50 (71%). There was little difference in the margins of bias or standards deviations between each iterated version and its noniterated counterpart. This general pattern was also found in nearly all of the conditions in the simulation. Although the performance of SKMO-N60 was always very similar to that of SKMO-I60 across the various conditions, the former consistently had a slight advantage over the latter. SKMO-N50 usually had a somewhat better showing than SKMO-I50 at each condition.

There were a few minor exceptions to the preceding characterization of the two communality estimation methods used in the SKMO. SKMO-I50 was marginally more accurate than SKMO-N50 when the variable-to-factor ratio was 3:1 (63% to 60%). The performances of SKMO-N50 and SKMO-I50 became increasingly similar as the true number of factors increased; in fact, these two tests were virtually identical at eight and ten factors. These observations suggest that SKMO-N50 may be more vulnerable to certain extreme conditions than SKMO-I50.

Clearly, the most unexpected outcome was the consistently superior performance of each of the noniterated forms over its iterated counterpart. The relevant research literature and textbooks largely promote the use of iterative communality estimation in CFA. Also known as refactoring, this approach has been reported to be the most accurate method of estimation (Floyd & Widaman, 1995; Widaman, 1993). It was therefore expected that the versions of SKMO utilizing the iterative

communality estimation approach would have an advantage over the versions that were simply using noniterated SMCs.

Although the iterative approach is generally accepted, there is not a consensus in the research community. Guttman (1956) cautioned that there had been no formal proof that refactoring would necessarily converge to the proper values (if at all). More recent literature confirms that a proof has still not been found (e.g., Mulaik, 2010). This criticism of refactoring is not without an empirical basis. Comrey (1978) claimed that iterative procedures can artificially inflate communality estimates. Moreover, Gorsuch (2003) noted that under certain conditions this process would produce unrealistic values (Heywood cases). Gorsuch (1983) also reported of instances where refactoring failed to converge on a single set of estimates. Arguing that these observations violate the mathematical definition of iteration, Gorsuch refers to the process as *pseudoiteration*.

Another possible issue involves the convergence criteria used in the refactoring process. The rule currently used by most researchers call for iteration to cease when the maximum change in the set of estimates from one round to the next is .001 or less (Widaman & Herringer, 1985). An ancillary criterion may be set to limit the number of iterations allowed (usually 50). However, a recent study suggests that the current standard criteria (.001 and 50) may not be stringent enough to consistently produce stable estimates (Ramsey & Gibson, 2006).

There is also a strong possibility that misspecification is interfering with refactoring. Communalities are inextricably linked with the number of factors. When estimating communalities by refactoring, the number of common factors must first be assumed and held constant until convergence (Floyd & Widaman, 1985; Gorsuch, 1983; Harman, 1976). The SKMO utilizes this iterative process at each step in the sequence, thus a set of communality estimates are produced for every possible number of factors (1 to p). Obviously, all but one of these solutions would be based upon an incorrect number of factors. This is potentially a problem because the refactoring process

may not necessarily produce a stable or accurate set of estimates if it is based upon the wrong number of factors (Widaman & Herringer, 1985).

Conclusions regarding the first research question. The results of this Monte Carlo simulation suggest that the SKMO-N50 is the superior version of the SKMO. Performance highlights for the SKMO-N50 include a very good overall accuracy (*HR* = 78%), a small margin of bias (*M* = -0.58), and a small discrepancy standard deviation (*SD* = 1.75). SKMO-N50 performed extremely well at the wide saturation (.3 to .8) level, which is a more realistic condition. Particularly impressive was the fact that the SKMO-N50 perform just as well with a 5:1 variable-to-factor ratio as it did with 8:1 and even 10:1. It was also observed to be fairly robust with respect to oblique solutions.

There is little doubt that .50 is the more appropriate cutoff value in general. But the relatively poor performance of SKMO-N50 at the single-factor level (coupled with the flawless performance of SKMO-N60 under the same condition) merits further investigation. Additional simulations should be conducted, and a theoretical justification for this phenomenon should be pursued.

These preliminary results indicate that the use of SMCs without iteration is the preferable method of communality estimation to incorporate in the SKMO. However, it would appear that further investigation is warranted regarding the use of iterated versus noniterated communality estimation. This is based upon four key issues. First, much of the literature suggests that iterated factor solutions should be more accurate. Second, the noniterated SKMO versions never outperformed the iterated SKMO versions by any great margin. Third, there is a conspicuous lack of references (or even a basic explanation) for the refactoring algorithms used by the software packages, so it is possible that a better algorithm exists (or the wrong one has been used). Finally, the iteration process can be unstable when the incorrect number of factors is specified.

Even though the SKMO is still in the early stages of development, preliminary recommendations for its usage are warranted. Of the four variants of the SKMO presented in this study, the SKMO-N50 is clearly the most likely candidate for general use with CFA at the present time. This noniterated version of the SKMO can be used even if the final factor solution will be iterated. Since the stability and accuracy of the iterated communality estimates are dependent upon the correct number of factors being specified, the most accurate version of the SKMO (namely, the SKMO-N50) should be used. But given its novelty, the author strongly recommends that the SKMO be used in conjunction with other dimensionality tests.

A tentative modification to the SKMO is now presented as a final thought in the discussion of the first research question. This modification is a noniterated SKMO where .60 is the criterion for the first two potential factors and .50 is the criterion for the rest. Using the .60 at the second step of the sequence is the crucial element in this modification (recall the slightly elevated SKMO value at the second step of the procedure when there is a single factor). Using .60 as the cutoff at the second step would prevent overextraction when there is only one true factor, and it would not disrupt performance when two factors are present since SKMO-N50 had exactly the same hit rate as SKMO-N60 and SKMO-I60 at the two-factor level. A trial run of this hybrid SKMO on the data generated for this study resulted in an overall accuracy of 80%.

Research Question 2

The second research question in this study explores how well the SKMO procedure performs in comparison to well-known tests of dimensionality (namely, the proportion of variance extracted, Bartlett's chi-square test, the Cattell-Nelson-Gorsuch scree test, the Kaiser-Guttman criterion, parallel analysis, and the minimum average partial test) with respect to the accuracy, precision, and bias of the predicted number of factors to retain. Since the SKMO-N50 was the best performing version of SKMO in this simulation, it will serve as the primary representative of the SKMO family for the purposes of the following discussion.

Classical methods. Overall, the SKMO procedures were clearly superior to the classical methods (PVE, BCS, CNG, and KG). The accuracy, precision, and margin of bias for each of the classical methods were nowhere near that of the SKMO-N50. In fact, SKMO-I60 (the least accurate version of SKMO) had a hit rate of 57%, where KG (the most accurate of the classical tests) had a hit rate of 43%. Furthermore, the SKMO-N50 was almost always more accurate and precise than all of the classical methods under the various conditions of the simulation.

There were two minor exceptions to the generality regarding the SKMO and the classical methods. First, the KG was slightly more accurate than SKMO-N50 with a 3:1 variable-to-factor ratio (64% to 60%). KG displayed a counterintuitive trend as it became more accurate and precise when the variable-to-factor ratio decreased, but the same general trend was also observed in the study by Piccone (2009). This is, of course, linked to the known sensitivity of KG to the number of observed variables (see Russell, 2002). The second exception was that BCS performs slightly better than SKMO-N50 at the low saturation level (55% to 53%). BCS was unexpectedly more accurate at the low-saturation level than it was at the wide or high levels. Corroborating evidence was unavailable since .50 was the smallest pattern coefficient used in the Zwick and Velicer (1986) study. Assuming this is typical behavior for the BCS, the most logical explanation is that the loss of power resulting from low saturation offset the tendency for BCS to overextract, hence improved accuracy (see Gorsuch, 1973; Hong et al., 2006). Thus, the KG and BCS performed better under these specific conditions simply because the primary sources of their weaknesses were attenuated.

The results for the classical methods from the present study are corroborated by previous Monte Carlo studies. The simulation by Zwick and Velicer (1986) showed the same general performance characteristics for the BCS, such as an increase in the margin of overextraction with an increase in the saturation, number of variables, and sample size. In the present study KG was observed to overextract, which has also been reported by numerous Monte Carlo studies (e.g., Costello & Osborne, 2005; Velicer et al., 2000; Zwick & Velicer, 1986). Most recently, a study by

Piccone (2009) reported an overall accuracy for KG just above 45%,[1] which was very similar to the accuracy of KG observed in the present study (43%). The results for the CNG matched those from Nasser et al. (2002). Specifically, if the factors are orthogonal in the population, then CNG overextracted by about one factor; however, if the factors are moderately oblique (= .4 in this case), then CNG consistently indicated three factors.

Parallel analysis. Given its reputation, it is not surprising that the various forms of parallel analysis (PA-UM, PA-UP, PA-RM, and PA-RP) all performed extremely well in this simulation. In general, SKMO-N50 was observed to perform better than PA-RM. Specifically, SKMO-N50 was observed to have a better overall accuracy than PA-RM, and the former was more accurate under most of the conditions of this simulation. However, the remaining three variants (PA-UM, PA-UP, and PA-RP) each produced an overall accuracy of 86% or greater, which was somewhat better than the accuracy of SKMO-N50 (78%). These three variants of parallel analysis also performed better at a majority of the simulation design conditions. The most noticeable differences in performance occurred with eight or more true factors in the population and with low saturation. Given that the SKMO method is still in the early stages of development, it is important to note that the differences in the performance of the SKMO-N50 and the performances of PA-RP, PA-UM, and PA-UP were never excessive (both overall and conditional).

There are numerous noteworthy exceptions to this profile. SKMO-N60 and SKMO-I60 performed just as well as all versions of parallel analysis at the single-factor condition. SKMO-N50 performed very nearly as well as (and in some instances, better than) all four versions of parallel analysis in the presence of three or four factors in the population. There are only slight differences in the performances of SKMO-N50 (*HR* = 85%) and PA-RP (*HR* = 89%) when the saturation range is wide. With high saturation, the performance of SKMO-N50 was virtually equivalent to the performances of all version of parallel analysis (all more than 95% accurate). The accuracy of

SKMO-N50 was very close to that of PA-RP at the moderate (81% and 86%, respectively) and large (86% and 90%, respectively) sample sizes.

The performance outcomes for parallel analysis observed in the present study are generally consistent with prior research. Similar to the present study, many other Monte Carlo simulations have shown that the various forms of parallel analysis are among the most accurate number-of-factors tests (e.g., Mumford et al., 2003; Velicer et al., 2000; Zwick & Velicer, 1986). The present study is also consistent with the simulation by Humphreys and Montanelli (1975) which demonstrated that parallel analysis (specifically, PA-RM) is accurate when applied to CFA, and that PA-RM tends to overextract. The results from this study generally agreed with those of the Zwick and Velicer (1986) Monte Carlo in that the accuracy and precision of PA-UM improved with increased sample size, saturation, and variable-to-factor ratio. However, Zwick and Velicer noted that PA-UM had a mild tendency to overextract, which is contradictory to the observations in the present study. As a final validity check, the overall outcomes for PA-UP and PA-RP in the present study were compared to those produced in the Monte Carlo by Piccone (2009). Both studies showed very similar overall hit rates for PA-UP (89% in the current study and 92% in Piccone), and both reported a very slight tendency for PA-UP to underextract. Likewise, the overall hit rates for PA-RP in these two simulations were comparable (86% and 83%), and both showed PA-RP to have a small upward bias.

Although not directly related to the research question, it is nevertheless interesting to note that the versions of parallel analysis which used unreduced correlation matrices (PA-UM and PA-UP) performed slightly yet consistently better than the versions using reduced correlation matrices (PA-RM and PA-RP). This certainly lends support to the argument that Horn and Kaiser based their rationales on the common-factor model.

Minimum average partial test. Another highly recommended test of dimensionality, the MAP procedure performed fairly well in this simulation. However, the SKMO-N50 and SKMO-I50

were both shown to be superior to the MAP in overall accuracy, precision, and margin of bias. Further, SKMO-N50 performed better at almost all of the simulation design settings. Especially conspicuous is that the SKMO-N50 was more robust to correlated factors and to smaller variable-to-factor ratios than the MAP. These two approaches have very similar rationales, but the results of this Monte Carlo clearly suggest that the SKMO is better suited for use with CFA than the MAP. These results are particularly promising given that the SKMO is still in the preliminary stages of its development.

There were only a few exceptions to this general comparison of MAP and SKMO. First, SKMO-N50 was not as accurate as the MAP when the population has only one factor; regardless, SKMO-N60 and SKMO-I60 were both slightly better than the MAP at this same condition. Next, the MAP is just barely more accurate than SKMO-N50 when there are 10 factors (52% and 51%, respectively). Lastly, the MAP was 90% accurate when the variable-to-factor ratio is 10:1, but SKMO-N50 was only 82% accurate at that same condition.

Like the results from the present simulation, numerous other Monte Carlo studies showed the MAP to have a strong inclination for underextraction (e.g., Mumford et al., 2003; Velicer et al., 2000; Zwick & Velicer, 1986). The Zwick and Velicer study further confirmed that the accuracy of the MAP test is relatively unaffected by sample size, and its accuracy does increase with greater saturation and with more variables per factor. However, the present study conflicts with Zwick and Velicer regarding the accuracy of the MAP when the factor saturation is high (84% and 97%, respectively). This is most likely because Zwick and Velicer used .8 as the value for all pattern coefficients in this condition, whereas the current study used a range of values from .6 to .8 for this condition. Finally, the Piccone (2009) study provides strong corroborating evidence as it reported an overall hit rate of 66% and mean discrepancy of -1.43 for the MAP, which is very similar to the corresponding results in the present study (*HR* = 63%, *M* = -1.79).

Conclusions regarding the second research question. An inspection of the various head-to-head comparisons of the dimensionality tests in this Monte Carlo simulation showed the SKMO (particularly SKMO-N50) to be consistently better than many other well-established number-of-factors procedures, including KG, BCS, PVE, and CNG. Most notably, the SKMO-N50 outperformed the MAP test and PA-RM. Yet the remaining three versions of parallel analysis (PA-UM, PA-UP, and PA-RP) were generally better than SKMO-N50. However, PA-UM, PA-UP, or PA-RP never surpassed SKMO-N50 by an extreme margin in any aspect of performance. The continued research suggested in the discussion of the first research question may also have implications for this question as well. In particular, if the iterative communality estimation procedure can be corrected, then the SKMO-I50 may have the potential to close the performance gap on these three variants of parallel analysis. Regardless of the potential improvements, the current incarnation of the SKMO is better suited for use in CFA than the MAP test and should be a candidate for general use alongside parallel analysis.

Limitations

The major findings of a study must also be tempered with the acknowledged limitations of its design. Although this was an extensive Monte Carlo simulation, some aspects are nevertheless simplistic, which potentially make this simulation less realistic—that is, it diminishes external validity. Each population pattern matrices were designed to have variable complexity of one, and the variable-to-factor ratio was homogeneous (i.e., each factor in a given population had the same number of salient loadings). Similarly, the pattern-magnitude intervals and inter-factor correlations were strictly homogeneous for any given population. It should be noted that all data generated for this simulation were continuous; the basic type of data (discrete or continuous) may also be influential in factor analysis (Turner, 1998). Finally, it must be acknowledged that these were pristine data generated from a standard-normal distribution with no missing observations.

For reasons such as this, generated data cannot be assumed to be completely representative of real data (Tucker et al., 1969).

Future Research

As previously suggested in the discussions for the research questions, one of the most interesting topics to arise from these results focuses on the fact that the noniterated versions of SKMO were consistently more accurate than the iterated forms. Numerous possible explanations have been offered, each of which is worthy of further exploration. Also mentioned in the previous discussion was the potential for future research regarding the unusual behavior of the SKMO when a single factor was present in the population. Specifically, a theoretical rationale is needed explain why SKMO-N60 and SKMO-I60 perform so well (and SKMO-N50 so poorly) at this particular condition. Another possible avenue for future research involves the manipulation of the SKMO cutoff criterion value to correct for underextraction (e.g., use .45 rather than .50). All of the issues mentioned in the limitations are certainly worthy topics for future research. An important next step is to explore how these more complex (and ostensibly more realistic) conditions will impact the performance of the SKMO.

Finally, it should be mentioned that during the course of this study, an overarching research theme has emerged regarding the joint use of numerous dimensionality procedures. It has often been suggested that multiple dimensionality tests should be used (Humphreys & Montanelli, 1975; Rummel, 1970; Stellefson & Hanik, 2008). However, these joint approaches rely on the rather subjective process of establishing a preponderance of evidence based upon an array of fragmented information. A systematic examination of the many different number-of-factors methods has revealed that each focuses on different theoretical aspects and therefore presents unique information when applied to observed data. Although there will undoubtedly be some common information among the various indicators, each of these tests would certainly have some unique information to contribute as well. Thus, future research in this area should focus on the

development of a method which assimilates such information from multiple number-of-factors predictors in order to produce a single integrated and unified indicator of dimensionality.

Another path to explore within this theme is the conditional application of these number-of-factors procedures based upon the characteristics of the sample data being analyzed. Although this study attempted to establish parsimonious profiles for the various dimensionality tests, there were nevertheless many exceptions to these generalities. Clearly, no one procedure is the best at every condition. Thus, an approach that takes the conditional performance characteristics of each number-of-factors test into consideration is a potentially powerful tool.

Conclusions

The SKMO is a new dimensionality procedure for aiding in determining the number-of-factors that should be rotated to a final solution for interpretation in common-factor analysis. A Monte Carlo study varying several key dimensions of the factor analysis problem found that the SKMO (particularly the SKMO-N50 variant) may prove to be a useful addition to currently used criteria. In fact, under most conditions the SKMO performed better than all of the existing criteria except some variants of Horn's parallel analysis. However, under some conditions the SKMO was found to be more effective in recovering dimensionality than parallel analysis. Another key finding of this study is that no one method is generally superior. This would seem to add credibility to the common recommendation that multiple criteria should be considered when performing common-factor analysis. Perhaps there is an opportunity to extrapolate information quantitatively from each criterion in a more robust procedure, although this approach will require a significant level of innovation relative to traditional approaches. The other general suggestion that emerged from this study is that when individual criteria are considered, they should be considered in light of the known performance characteristics and limitations of each criterion.

Footnotes

[1] The results of the Piccone (2009) study showed an overall hit rate of 23% and a mean discrepancy of 5.87 for KG, which is vastly different from the corresponding results of the present study. However, this disagreement would appear to be due to the inclusion of unique indicator variables (noise variables that are not correlated with any other variables in the population) in the Piccone study, which showed KG to be quite sensitive to unique indicators. Since the present study did not include unique indicators as a design element, the results from the Piccone simulation that had no unique indicator were comparable to the results from the present study and are therefore presented here.

REFERENCES

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Bartholomew, D. J. (2007). Three faces of factor analysis. In Robert Cudeck & Robert C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 9-21). Mahwah, NJ: Lawrence Erlbaum Associates.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *The British Journal of Psychology*, *3*, 77-85.

Bock, R. D. (2007). Rethinking Thurstone. In Robert Cudeck & Robert C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 35-45). Mahwah, NJ: Lawrence Erlbaum Associates.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*(4), 509-540.

Cattell, R. B. (1958). Extracting the correct number of factors in factor analysis. *Educational and Psychological Measurement*, *18*(4), 791-838.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *The Journal of Multivariate Behavioral Research*, *12*, 289-325.

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Reserarch*, *12*(1), 43-47.

Cliff, N. (1988). The eigenvalue-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*(2), 27-279.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology*, *46*(4), 648-659.

Coovert, M. D. & McNelis, K. (1988). Determining the number of common factors in factor analysis: A review and program. *Educational and Psychological Measurement*, *48*, 687-692.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research and Evaluation*, *10*(7). Retrieved from http://pareonline.net/pdf/v10n7.pdf

Cota, A. A., Longman, R. S., Holden, R. R., Fekken, G. C., & Xinaris, S. (1993). Interpolating 95th percentile eigenvalues from random data: An empirical example. *Educational and Psychological Measurement*, *53*, 585-596.

Crawford, C. B. (1975). Determining the number of interpretable factors. *Psychological Bulletin*, *82*(2), 226-237.

Crawford, C. B., & Koopman, P. (1973). A note on Horn's test for the number of factors. *Multivariate Behavioral Research*, *8*, 117-125.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Group.

Cudeck, R. (2007). Factor analysis in the year 2004: Still spry at 100. In Robert Cudeck & Robert C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 1-7). Mahwah, NJ: Lawrence Erlbaum Associates.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358-361.

Everitt, B. S. (2006). *The Cambridge dictionary of statistics* (3rd ed.). New York: Cambridge University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Educational and Psychological Measurement*, *27*(3), 387-415.

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement*, *56*(6), 907-929.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Fraleigh, J., & Beauregard, R. (1995). *Linear algebra* (3rd ed.). Reading, MA: Addison-Wesley.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*(3), 377-393.

Gorsuch, R. L. (1973). Using Bartlett's significance test to determine the number of factors to extract. *Educational and Psychological Measurement*, 33, 361-364.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gorsuch, R. L. (1990). Common factor analysis versus principal components analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*, 33-39.

Gorsuch, R. L. (2003). Factor analysis. In J. A Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (pp. 143-164). Hoboken, NJ: John Wiley.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, *19*, 149-161.

Hakstian, A. R., & Muller, V. J. (1973). Some notes on the number of factors problem. *Multivariate Behavioral Research*, *8*, 461-475.

Hakstian, A. R., Rogers, T., & Cattell, R. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, *17*, 193-219.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Harris, R. J. (1975). *A primer of multivariate statistics*. New York: Academic Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191-205.

Hong, S., Mitchell, S. K., & Harshman, R. A. (2006). Bootstrap scree tests: A Monte Carlo simulation and applications to published data. *British Journal of Mathematical and Statistical Psychology*, *59*, 35-57.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.

Horn, J. L., & Engstrom, R. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, *14*, 283-300.

Humphreys, L. G., & Illgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, *29*, 571-578.

Humphreys, L. G., & Montanelli, R. G., Jr. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*, 193-205.

Jones, L. V. (2007). Remembering L. L. Thurstone. In Robert Cudeck & Robert C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 23-33). Mahwah, NJ: Lawrence Erlbaum Associates.

Jöreskog, K. G. (2007). Factor analysis and its extensions. In Robert Cudeck & Robert C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 47-77). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401-415.

Kaiser, H. F. (1981). A revised measure of sampling adequacy for factor-analytic data matrices. *Educational and Psychological Measurement*, *41*(2), 379-381.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, *34*, 111-117.

Kendall, M. G., & Babington-Smith, B. (1950). Factor analysis. *Journal of the Royal Statistical Society*, *12*, 60-94.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1990). Assessing sampling variation relative to number-of-factors criteria. *Educational and Psychological Measurement*, *50*(1), 33-48.

Linn, R. L. (1968). A Monte Carlo approach to the number of factor problem. *Psychometrika*, *33*(1), 37-71.

Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G. C. (1989). A regression equation for the parallel analysis criterion in principal components analysis: Mean and 95th percentile eigenvalues. *Multivariate Behavioral Research*, *24*(1), 59-69.

MacCallum, R. C., Brown, M. W., & Cai, L. (2007). Factor analysis models as approximations. In Robert Cudeck & Robert C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 153-175). Mahwah, NJ: Lawrence Erlbaum Associates.

McArdle, J. J. (1990). Principles versus principals of structural factor analysis. *Multivariate Behavioral Research*, *25*, 81-87.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, *22*, 267-305.

Mulaik, S. A. (1990). Blurring the distinction between component analysis and common factor analysis. *Multivariate Behavioral Research*, *25*, 53-59.

Mulaik, S. A. (2010). *The foundations of factor analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Mumford, K. R., Ferron, J. M., Hines, C. V., Hogarty, K. Y., & Kromrey, J. D. (2003, April). *Factor retention in exploratory factor analysis: A comparison of alternative methods*. Paper presented at the annual conference of the American Educational Research Association, Chicago, IL.

Mundfrom, D., Shaw, D., & Ke, T. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*(2), 159-168.

Nasser, F., Benson, J., & Wisenbaker, J. (2002). The performance of regression-based variations of the visual scree for determining the number of common factors. *Educational and Psychological Measurements*, *62*(3), 397-419.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, *32*, 396-402.

Piccone, A. V. (2009). *A comparison of three computational procedures for solving the number of factors problem in exploratory factor analysis* (Unpublished doctoral dissertation). University of Northern Colorado, Greeley, CO.

Pohlmann, J. T. (2004). Use and interpretation of factor analysis in *The Journal of Educational Research*: 1992-2002. *Journal of Educational Research*, *98*(1), 14-22.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*. *Personality and Social Psychology Bulletin*, *28*(12), 1629-1646.

Shirkey, E. C., & Dziuban, C. D. (1976). A note on some sampling characteristics of the measure of sampling adequacy (MSA). *Multivariate Behavioral Research*, *11*(1), 125-128.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

SPSS, Inc. (2010). *IBM SPSS statistics 19 command syntax reference*. Chicago: Author.

Stellefson, M. & Hanik, B (2008, February). *Strategies for determining the number of factors to retain in exploratory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.).Mahwah, NJ: Lawrence Erlbaum Associates.

Suhr, D. D. (2005). Principal components analysis vs. exploratory factor analysis. *Proceedings of the Annual SAS Users Group International Conference*. Retrieved from http://www2.sas.com/proceedings/sugi30/203-30.pdf

Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? *Proceedings of the Annual SAS Users Group International Conference*. Retrieved from http://www2.sas.com/proceedings/sugi31/200-31.pdf

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Thurstone, L. L. (1935). *Vectors of the mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. Chicago: University of Chicago Press.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421-459.

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, *58*(4), 541-568.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 47–71). Norwell, MA: Kluwer Academic Publishers.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*, 263-311.

Widaman, K. F., & Herringer, L. G. (1985). Iterative least squares estimates of communality: Initial estimate need not affect stabilized value. *Psychometrika*, *50*(4), 469-477.

Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. *The Statistician*, *31*(3), 221-228.

Zoski, K., & Jurs, S. (1993). Using multiple regression to determine the number of factors to retain in factor analysis. *Journal of Multiple Linear Regression Viewpoints*, *20*(1), 5-9.

Zoski, K. W., & Jurs, S. (1996). An objective counterpart to the visual scree test for factor analysis: The standard error scree. *Educational and Psychological Measurement*, *56*(3), 443-451.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432-442.

APPENDIX

MONTE CARLO PROGRAM

The Monte Carlo was conducted using version 19 of the IBM SPSS statistical software package on a Windows 7 (64-bit) operating system. This system uses an Intel Core i7, which is a 2.66-GHz, quad-core processor with simultaneous multithreading capabilities. The simulation program was written using a combination of the native SPSS command syntax, macro, and scripting languages. This program was designed in a modular fashion so that a user would need to modify the parameter values in only one file (viz., Commands.sps) to conduct a complete Monte Carlo simulation.

In the base directory (folder), there are four files (Commands.sps, MC.sps, MC.sav, and Clear_Viewer.wwd) and three subdirectories (Procedures, Data, and Results). Note that the data file MC.sav contains no relevant data. Since SPSS requires for there to be at least one data window open at all times, this file serves as the default data window. The command program for each dimensionality procedure was stored in a separate file, and all procedure files are placed in the Procedures folder. For the Monte Carlo program to run properly, the naming and arrangement of the files and subdirectories within the main directory is crucial.

Commands.sps

This is the only code file that needs to be modified in order to perform the Monte Carlo simulation. Other than the condition parameters, three key pieces of information are needed: the name of the base directory (<DIR_NAME>), a run command (<CMD_NAME>), and a procedure name (<PROC_NAME>). The base directory is simply the location of the folder containing the Monte Carlo program. For example, C:\Dissertation\MC\ was the base directory used for this simulation. There are two run commands: gen, which generates the random datasets; and pro, which processes the datasets with a given dimensionality test. The procedure is specified using any of the names called by the Load_Procs macro.

```
define base_dir () !quote('<DIR_NAME>') !enddefine.
set miterate=1000000000 mxloops=1000000000 rng=mt mtindex=random.
insert file = base_dir + 'MC.sps'.

Load_Procs
SKMO_N50 SKMO_I50 SKMO_N60 SKMO_I60
PVE CNG BCS KG
PA_UM PA_UP PA_RM PA_RP
MAP.

Monte_Carlo
nof =   [1 2 3 4 5 8 10]
vfr =   [3 5 8 10]
pmi =   [(.3,.5) (.6,.8) (.3,.8)]
ssize = [250 500 750]
ifc =   [0 .2 .4]
reps =  100
cmd =   <CMD_NAME>
proc =  <PROC_NAME>.
```

MC.sps

```
*-----------------------------------------------------------------------.
* GENERATE DATASETS MACRO                                               .
*-----------------------------------------------------------------------.
define Generate_Datasets (
    nof=!tokens(1)/vfr=!tokens(1)/pmi=!enclose('[',']')
    /ifc=!tokens(1)/ssize=!tokens(1)/reps=!tokens(1)
)
!let !base_dir = !unquote(!eval(base_dir))
!let !minpat = !head(!pmi)
!if (!tail(!pmi) = !null) !then
  !let !maxpat = !minpat
!else
  !let !maxpat = !eval(!tail(!pmi))
!ifend
!let !p = !null
!do !i = 1 !to !nof
  !let !p = !concat(!p, !blank(!vfr))
!doend
!let !p = !length(!p)
!let !cols = !length(!concat(!blanks(!nof), !blanks(!p)))
!let !params = !concat(!nof, '_', !vfr, '_', !minpat, '_', !maxpat, '_',
  !ifc, '_', !ssize, '_', !reps)

input program.
loop #j = 1 to !reps*!ssize.
+  vector factor(!cols, f8.6).
+  loop #i = 1 to !cols.
+  +  compute factor(#i) = rv.normal(0,1).
+  end loop.
+  end case.
end loop.
end file.
end input program.
save outfile = !quote(!concat(!base_dir, 'Data\_random.sav')).

matrix.
compute pattern = {!maxpat}.
loop #i = 1 to (!vfr – 1).
  Compute pattern = {pattern; !maxpat - #i*(!maxpat - !minpat)/(!vfr – 1)}.
end loop.
compute Beta = pattern.
loop #j = 1 to (!nof – 1).
  Compute Beta = block(Beta, pattern).
end loop.
compute Phi = make(!nof, !nof, !ifc).
call setdiag(Phi, 1).
compute Psi = sqrt(ident(!p) – mdiag(diag(Beta*Phi*T(Beta)))).
compute Rho_star = Beta*Phi*T(Beta).
call eigen(Rho_star, V, d).
compute Alpha = {V*mdiag(sqrt(d))}(:,1:!nof).
get random /file = !quote(!concat(!base_dir, 'Data\_random.sav')).
compute X = random(:,1:!nof).
compute Y = random(:,(!nof+1):(!nof+!p)).
compute Z = X*T(Alpha) + Y*Psi.
```

```
compute R_aug = make(!reps*!p, !p, -1).
loop #s = 1 to !reps.
  compute #Z = Z(((#s – 1)*!ssize + 1):(#s*!ssize),:).
  compute #CSSCP = T(#Z)*#Z – (1/!ssize)*T(csum(#Z))*csum(#Z).
  compute #invRCSS = inv(sqrt(mdiag(diag(#CSSCP)))).
  compute #R = #invRCSS*#CSSCP*#invRCSS.
  compute #R = (10**-8)*rnd((10**8)*#R).
  compute R_aug(((#s-1)*!p+1):(#s*!p),:) = #R.
end loop.
save R_aug
  /outfile = !quote(!concat(!base_dir, 'Data\CORR_', !params, '.sav'))
  /vars = CORR1 to !concat('CORR', !p).
end matrix.

!enddefine.


*------------------------------------------------------------------------.
* PROCESS DATASETS MACRO                                                 .
*------------------------------------------------------------------------.
define Process_Datasets (
     nof=!tokens(1)/vfr=!tokens(1)/pmi=!enclose('[',']')
    /ifc=!tokens(1)/ssize=!tokens(1)/reps=!tokens(1)
    /proc=!tokens(1)
)
!let !minpat = !head(!pmi)
!if (!tail(!pmi) = !null) !then
  !let !maxpat = !minpat
!else
  !let !maxpat = !eval(!tail(!pmi))
!ifend

!let !params = !concat(!nof, '_', !vfr, '_', !minpat, '_', !maxpat, '_',
  !ifc, '_', !ssize, '_', !reps)
!let !data_file = !concat(!unquote(!eval(base_dir)), 'Data\CORR_', !params,
  '.sav')
!let !results_file = !quote(!concat(!unquote(!eval(base_dir)), 'Results\',
  !proc, '.sav'))

matrix.
get R_aug /file = !quote(!data_file).
compute p = ncol(R_aug).
compute n = !ssize.
compute results = make(!reps, 8, -1).

loop #s = 1 to !reps.
  compute R = R_aug(((#s-1)*p+1):(#s*p),:).
  !concat('Run_', !proc).
  compute results(#s,:) = {#s, !nof, !vfr, !ssize, !minpat, !maxpat, !ifc,
!proc}.
end loop.

save results
  /outfile = !quote(!concat(!unquote(!eval(base_dir)), 'Results\_temp.sav'))
  /vars = s, nof, vfr, ssize, b_min, b_max, ifc, !proc.
end matrix.
```

```
get file = !results_file.
add files file = !results_file /file =
  !quote(!concat(!unquote(!eval(base_dir)), 'Results\_temp.sav')).
save outfile = !results_file.

!enddefine.


*------------------------------------------------------------------------.
* LOAD PROCEDURES MACRO                                                  .
*------------------------------------------------------------------------.
define Load_Procs (!positional !cmdend)
!do !proc !in (!1)
  insert file = !quote(!concat(!unquote(!eval(base_dir)), 'Procedures\',
   !proc, '.sps')) syntax=batch.
!doend
!enddefine.


*------------------------------------------------------------------------.
* COMMAND MACRO                                                          .
*------------------------------------------------------------------------.
define Monte_Carlo (
   nof=!enclose('[',']')/vfr=!enclose('[',']')/pmi=!enclose('[',']')
  /ifc=!enclose('[',']')/ssize=!enclose('[',']')/reps=!tokens(1)
  /proc=!tokens(1)/cmd=!tokens(1)
)
!let !pat = !null
!do !str !in (!pmi)
  !if (!str='(') !then
    !let !pat = !concat(!pat, "'")
  !else
    !if (!str=')') !then
      !let !pat = !concat(!pat, "'", !blanks(1))
    !else
      !if ((!str=',') !or (!str=';')) !then
        !let !pat = !concat(!pat, !blanks(1))
      !else
        !let !pat = !concat(!pat, !str)
      !ifend
    !ifend
  !ifend
!doend

!if (!cmd='pro') !then
  new file.
  input program.
    numeric s nof (f5.0) /vfr (f4.2) /ssize (f5.0)
      /b_min b_max ifc (f3.2) /!proc (f3.0).
    end file.
  end input program.
  execute.
  save outfile = !quote(!concat(!unquote(!eval(base_dir)),
    'Results\',!proc,'.sav')).
!ifend

!let !hash = !blanks(1)
```

```
!do !a !in (!nof)
  !do !b !in (!vfr)
    !do !c !in (!pat)
      !do !d !in (!ifc)
        !do !e !in (!ssize)
          !let !params = !concat('nof=',!a,' vfr=',!b,'
            pmi=[',!unquote(!c),']',' ifc=',!d,' ssize=',!e,' reps=',!reps)
          echo !quote(!params).
          echo !quote(!blanks(1)).
          echo !quote(!concat('COUNT: ', !length(!hash))).
          !let !hash = !concat(!hash, !blanks(1))
          !if (!cmd='gen') !then
            Generate_Datasets !params
          !ifend
          !if (!cmd='pro') !then
            Process_Datasets !params proc=!proc
          !ifend
          script base_dir + 'Clear_Viewer.wwd'.
        !doend
      !doend
    !doend
  !doend
!doend

!enddefine.
```

Clear_Viewer.wwd

```
Sub Main
   Dim objOutputDoc As IspssOutputDoc
   Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc
   objOutputDoc.SelectAll
   objOutputDoc.Remove
End Sub
```

SKMO_N50.sps

```
define Run_SKMO_N50 ()
* SETTINGS ========================.
* - CUTOFF CRITERION.
compute crit = .50.
* ================================.

* ANTI-IMAGE CORRELATION MATRIX (Q).
compute W = sqrt(inv(mdiag(diag(inv(R))))).
compute Q = W*inv(R)*W.

* KAISER-MEYER-OLKIN (KMO).
compute sum_qSqr = trace(Q*Q) - p.
compute sum_rSqr = trace(R*R) - p.
compute KMO = sum_rSqr/(sum_rSqr + sum_qSqr).
compute K = {KMO}.

* COMMUNALITY ESTIMATION (SMCs).
compute SMC = 1 - (1/diag(inv(R))).

loop #t = 1 to trunc(p/2).

* REDUCED CORR MATRIX.
+  compute Rred = R.
+  call setdiag(Rred,SMC).
+  call eigen(Rred,V,d).
+  compute V = V(:,1:#t).
+  do if (#t > 1).
+  +  compute D = abs(mdiag(d(1:#t))).
+  else.
+  +  compute D = abs(d(1)).
+  end if.
+  compute A = V*sqrt(D).
+  compute Rrep = A*T(A).

* PARTIAL (RESIDUALIZED) CORRELATION MATRIX (Rt).
+  compute COVres = R - A*T(A).
+  compute invSDres = mdiag(1/sqrt(diag(COVres))).
+  compute Rt = invSDres*COVres*invSDres.

* SEQUENTIAL KMO (Kt).
+  compute Wt = sqrt(inv(mdiag(diag(inv(Rt))))).
+  compute Qt = Wt*inv(Rt)*Wt.
+  compute sum_qSqr = trace(Qt**2)-p.
+  compute sum_rSqr = trace(Rt**2)-p.
+  compute Kt = sum_rSqr/(sum_rSqr+sum_qSqr).
+  compute K = {K; Kt}.

end loop.

loop #j = 1 to trunc(p/2).
+  compute SKMO_N50 = #j-1.
end loop if (K(#j)<crit).

!enddefine.
```

SKMO_N60.sps

This program is the same as SKMO_N50.sps with the following changes:

```
define Run_SKMO_N60 ()
* SETTINGS ========================.
* - CUTOFF CRITERION.
compute crit = .60.
* ================================.
      ⋮
+  compute SKMO_N60 = #j-1.
      ⋮
!enddefine.
```

SKMO_I50.sps

```
define Run_SKMO_I50 ()
* SETTINGS ========================.
* - MAXIMUM NUMBER OF ITERATIONS.
compute max_iter = 25.
* - CONVERGENCE CRITERION.
compute conv = .001.
* - CUTOFF CRITERION.
compute crit = .50.
* ================================.

* ANTI-IMAGE CORRELATION MATRIX (Q).
compute W = inv(mdiag(diag(inv(R))))&**.5.
compute Q = W*inv(R)*W.

* KAISER-MEYER-OLKIN (KMO).
compute sum_qSqr = trace(Q*Q) - p.
compute sum_rSqr = trace(R*R) - p.
compute KMO = sum_rSqr/(sum_rSqr + sum_qSqr).
compute K = {KMO}.

* INITIAL COMMUNALITY ESTIMATION (SMCs).
compute SMC = 1 - (1/diag(inv(R))).

loop #t = 1 to trunc(p/2).

* INITIAL REDUCED CORR. MATRIX.
+  compute Rred = R.
+  call setdiag(Rred,SMC).
+  call eigen(Rred,V,d).
+  compute V = V(:,1:#t).
+  do if (#t > 1).
+  +  compute D = abs(mdiag(d(1:#t))).
+  else.
+  +  compute D = abs(d(1)).
+  end if.
+  compute A = V*sqrt(D).
+  compute Rrep = A*T(A).
```

```
* ITERATION BY REFACTORING.
+  loop #i = 2 to max_iter.
+    compute h = diag(Rrep).
+    call setdiag(Rred, h).
+    call eigen(Rred,V,d).
+    compute V = V(:,1:#t).
+    do if (#t > 1).
+      compute D = abs(mdiag(d(1:#t))).
+    else.
+      compute D = abs(d(1)).
+    end if.
+    compute A = V*sqrt(D).
+    compute Rrep = A*T(A).
+    compute maxdiff = cmax(abs(diag(Rred-Rrep))).
+  end loop if (maxdiff<=conv).

* PARTIAL (RESIDUALIZED) CORRELATION MATRIX (Rt).
+  compute COVres = R - A*T(A).
+  compute invSDres = mdiag(1/sqrt(abs(diag(COVres)))).
+  compute Rt = invSDres*COVres*invSDres.

* SEQUENTIAL KMO (Kt).
+  compute Wt = sqrt(abs(inv(mdiag(diag(inv(Rt)))))).
+  compute Qt = Wt*inv(Rt)*Wt.
+  compute sum_qSqr = trace(Qt**2)-p.
+  compute sum_rSqr = trace(Rt**2)-p.
+  compute Kt = sum_rSqr/(sum_rSqr+sum_qSqr).
+  compute K = {K; Kt}.

end loop.

loop #j = 1 to trunc(p/2).
+  compute SKMO_I50 = #j-1.
end loop if (K(#j)<crit).

!enddefine.
```

SKMO_I60.sps

This program is the same as SKMO_I50.sps with the following changes:

```
define Run_SKMO_I60 ()
* SETTINGS =======================.
      ⋮
* - CUTOFF CRITERION.
compute crit = .60.
* ================================.
      ⋮
+  compute SKMO_I60 = #j-1.
      ⋮
!enddefine.
```

PVE.sps

```
define Run_PVE ()
```

```
* SETTINGS ========================.
compute crit = .75.
* ================================.
compute d = eval(R).
compute Q = 0.
loop #t = 1 to (p-1).
+  compute Q = csum(d(1:#t,1))/p.
+  compute PVE = #t.
end loop if (Q>=crit).
!enddefine.
```

CNG.sps

```
define Run_CNG ()
do if (p>=7).
+  compute d = eval(R).
+  compute diff = make((p-5),1,0).
+  loop #t = 3 to (p-3).
+  +  compute diff((#t-2)) = d(#t+3)-d(#t+1)-d(#t)+d(#t-2).
+  end loop.
+  compute CNG = 3.
+  compute max_diff = diff(1).
+  loop #t = 4 to (p-3).
+  +  do if (diff((#t-2)) > max_diff).
+  +    compute CNG = #t.
+  +    compute max_diff = diff((#t-2)).
+  +  end if.
+  end loop.
else.
+  compute CNG = -99.
end if.
!enddefine.
```

BCS.sps

```
define Run_BCS ()
* SETTINGS ========================.
compute NSL = .01.
* ================================.
compute d = eval(R).
compute Qnum = det(R).
compute Qden = (csum(d)/p)**p.
compute Q = Qnum/Qden.
compute K = 1 - n + (2*p + 5)/6.
compute df = (p - 1)*(p + 2)/2.
compute OSL = 1 - chicdf(K*ln(Q), df).
compute BCS = 0.
loop #t = 1 to p-2 if (OSL<NSL).
+  compute BCS = #t.
+  compute Qnum = Qnum/d(#t).
+  compute Qden = (csum(d((#t+1):p))/(p-#t))**(p-#t).
+  compute Q = Qnum/Qden.
+  compute K = 1 - n + (2*p + 5)/6 + 2*#t/3.
+  compute df = (p-#t-1)*(p-#t+2)/2.
+  compute OSL = 1 - chicdf(K*ln(Q), df).
end loop.
```

```
!enddefine.
```

KG.sps

```
define Run_KG ()
compute d = eval(R).
compute KG = 0.
loop #t = 1 to p if (d(#t)>=1).
+  compute KG = #t.
end loop.
!enddefine.
```

PA_UM.sps

```
define Run_PA_UM ()
* SETTINGS ========================.
* - NUMBER OF PARALLEL DATASETS.
compute reps = 50.
* ================================.
compute d_obs = eval(R).
compute D_par = make(p,reps,-999).
loop #k = 1 to reps.
+  compute Z = uniform(n,p).
+  compute COV = (1/(n-1))*(sscp(Z) - ((T(csum(Z))*csum(Z))/n)).
+  compute SD = inv(mdiag(sqrt(diag(COV)))).
+  compute D_par(:,#k) = eval(SD*COV*SD).
end loop.
compute d_crit = rsum(D_par)/reps.
compute diff = d_obs - d_crit.
loop #j = 1 to p.
+  compute PA_UM = #j - 1.
end loop if (diff(#j)<=0).
!enddefine.
```

PA_UP.sps

```
define Run_PA_UP ()
* SETTINGS ========================.
* - NUMBER OF PARALLEL DATASETS.
compute reps = 50.
* - CRITERION PERCENT.
compute pct = 95.
* ================================.
compute d_obs = eval(R).
compute D_par = make(p,reps,-999).
loop #k = 1 to reps.
+  compute Z = uniform(n,p).
+  compute COV = (1/(n-1))*(sscp(Z) - ((T(csum(Z))*csum(Z))/n)).
+  compute SD = inv(mdiag(sqrt(diag(COV)))).
+  compute D_par(:,#k) = eval(SD*COV*SD).
end loop.
compute order = make(p,reps,-1).
compute D_ord = make(p,reps,-1).
loop #i = 1 to p.
+  compute order(#i,:) = grade(D_par(#i,:)).
```

```
end loop.
loop #i = 1 to p.
+  loop #j = 1 to reps.
+  +  compute D_ord(#i,order(#i,#j)) = D_par(#i,#j).
+  end loop.
end loop.
compute q = trunc(pct*reps/100).
compute d_crit = make(p,1,-1).
loop #i = 1 to p.
+  do if (q=pct*reps/100).
+  +  compute d_crit(#i) = .5*(D_ord(#i,q) + D_ord(#i,(q+1))).
+  else.
+  +  compute d_crit(#i) = D_ord(#i,(q+1)).
+  end if.
end loop.
compute diff = d_obs - d_crit.
loop #j = 1 to p.
+  compute PA_UP = #j - 1.
end loop if (diff(#j)<=0).
!enddefine.
```

PA_RM.sps

```
define Run_PA_RM ()
* SETTINGS ========================.
* - NUMBER OF PARALLEL DATASETS.
compute reps = 50.
* ================================.
compute SMC = 1 - (1 &/ diag(inv(R))).
call setdiag(R,SMC).
compute d_obs = eval(R).
compute D_par = make(p,reps,-999).
loop #k = 1 to reps.
+  compute Z = uniform(n,p).
+  compute CSSCP = sscp(Z) - ((T(csum(Z))*csum(Z))/n).
+  compute RRCSS = inv(mdiag(sqrt(diag(CSSCP)))).
+  compute Rstar = RRCSS*CSSCP*RRCSS.
+  compute SMC = 1 - (1 &/ diag(inv(Rstar))).
+  call setdiag(Rstar, SMC).
+  compute D_par(:,#k) = eval(Rstar).
end loop.
compute d_crit = rsum(D_par)/reps.
compute diff = d_obs - d_crit.
loop #j = 1 to p.
+  compute PA_RM = #j - 1.
end loop if (diff(#j)<=0).
!enddefine.
```

PA_RP.sps

```
define Run_PA_RP ()
* SETTINGS ========================.
* - NUMBER OF PARALLEL DATASETS.
compute reps = 50.
* - CRITERION PERCENT.
compute pct = 95.
```

```
* ================================.
compute SMC = 1 - (1 &/ diag(inv(R))).
call setdiag(R,SMC).
compute d_obs = eval(R).
compute D_par = make(p,reps,-999).
loop #k = 1 to reps.
+  compute Z = uniform(n,p).
+  compute CSSCP = sscp(Z) - ((T(csum(Z))*csum(Z))/n).
+  compute invRCSS = inv(mdiag(sqrt(diag(CSSCP)))).
+  compute Rstar = invRCSS*CSSCP*invRCSS.
+  compute SMC = 1 - (1 &/ diag(inv(Rstar))).
+  call setdiag(Rstar, SMC).
+  compute D_par(:,#k) = eval(Rstar).
end loop.
compute order = make(p,reps,-1).
compute D_ord = make(p,reps,-1).
loop #i = 1 to p.
+  compute order(#i,:) = grade(D_par(#i,:)).
end loop.
loop #i = 1 to p.
+  loop #j = 1 to reps.
+  +  compute D_ord(#i,order(#i,#j)) = D_par(#i,#j).
+  end loop.
end loop.
compute q = trunc(pct*reps/100).
compute d_crit = make(p,1,-1).
loop #i = 1 to p.
+  do if (q=pct*reps/100).
+  +  compute d_crit(#i) = .5*(D_ord(#i,q) + D_ord(#i,(q+1))).
+  else.
+  +  compute d_crit(#i) = D_ord(#i,(q+1)).
+  end if.
end loop.
compute diff = d_obs - d_crit.
loop #j = 1 to p.
+  compute PA_RP = #j - 1.
end loop if (diff(#j)<=0).
!enddefine.
```

MAP.sps

```
define Run_MAP ()
call eigen (R, V, d).
compute loadings = V*sqrt(mdiag(d)).
compute Vt = make(p, 1, 0).
compute Vt(1) = trace(R**2).
loop #t = 1 to p-1.
+ compute At = loadings(:,1:#t).
+ compute partcov = R - (At * T(At)).
+ compute sd = mdiag(1/(sqrt(diag(partcov)))).
+ compute Rt = sd*partcov*sd.
+ compute Vt(#t+1) = trace(Rt**2).
end loop.
compute gmin = mmin(Vt).
loop #j = 1 to p-1.
+  compute MAP = #j - 1.
```

```
end loop if (Vt(#j)=gmin).
!enddefine.
```

VITA

Brent Dale Hill

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE SEQUENTIAL KAISER-MEYER-OLKIN PROCEDURE AS AN ALTERNATIVE FOR DETERMINING THE NUMBER OF FACTORS IN COMMON-FACTOR ANALYSIS: A MONTE CARLO SIMULATION

Major Field: Educational Psychology

Biographical:

Personal Data: Born in Shawnee, Oklahoma, on October 9, 1973, the son of Bryce Dale and Wilma Dean Hill (née Carter), younger brother of Bryce Anthony Hill; married to Kimberly Dawn Hill (née Winkelman).

Education: Graduated from New Lima High School (New Lima, Oklahoma) in May, 1991; received Bachelor of Science degree in Psychology from the University of Tulsa (Tulsa, Oklahoma) in May 1997; received Master of Science degree in Educational Psychology (option in Research, Evaluation, Measurement, and Statistics) from Oklahoma State University (Stillwater, Oklahoma) in July, 2007; completed the requirements for the Doctor of Philosophy degree in Educational Psychology (option in Research, Evaluation, Measurement, and Statistics) from Oklahoma State University (Stillwater, Oklahoma) in July, 2011.

Name: Brent Dale Hill Date of Degree: July, 2011

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THE SEQUENTIAL KAISER-MEYER-OLKIN PROCEDURE AS AN ALTERNATIVE FOR DETERMINING THE NUMBER OF FACTORS IN COMMON-FACTOR ANALYSIS: A MONTE CARLO SIMULATION

Pages in Study: 144 Candidate for the Degree of Doctor of Philosophy

Major Field: Educational Psychology

Scope and Method of Study:

Widely utilized in the behavioral and social sciences, common-factor analysis (CFA) is a statistical technique which is used to investigate the latent traits (factors) that underlie a set of observed variables. The proper number of factors to extract is a fundamental question in exploratory CFA, and many methods to answer that question have been devised. This study examines the performance characteristics (accuracy, precision, and bias) of four variants of the sequential Kaiser-Meyer-Olkin (SKMO), a new method for determining dimensionality in CFA. This study also compares the SKMO to various other well-known dimensionality tests, such as the Kaiser-Guttman criterion, Horn's parallel analysis, and Velicer's MAP test. This study was conducted using an extensive Monte Carlo simulation which manipulated the actual number of factors, the variable-to-factor ratio, the pattern-magnitude interval, sample size, and inter-factor correlations.

Findings and Conclusions:

The simulation revealed that the best-performing SKMO variant was that which incorporated noniterated communality estimation and a .50 cutoff. The simulation also showed that the SKMO performed better than most other number-of-factors tests, including the Kaiser-Guttman criterion and Velicer's MAP test. The SKMO was better than one version of parallel analysis and a close second to the remaining forms. These results suggest that the SKMO is a viable candidate for general use with CFA. However, this suggestion is tentative as further research is needed to determine the performance characteristics of the SKMO under increasingly complex conditions.

ADVISOR'S APPROVAL: Dale R. Fuqua