Evaluating the Use of Exploratory Factor Analysis in Psychological Research

Leandre R. Fabrigar
Queen's University

Duane T. Wegener
Purdue University

Robert C. MacCallum
Ohio State University

Erin J. Strahan
Queen's University

Despite the widespread use of exploratory factor analysis in psychological research, researchers often make questionable decisions when conducting these analyses. This article reviews the major design and analytical decisions that must be made when conducting a factor analysis and notes that each of these decisions has important consequences for the obtained results. Recommendations that have been made in the methodological literature are discussed. Analyses of 3 existing empirical data sets are used to illustrate how questionable decisions in conducting factor analyses can yield problematic results. The article presents a survey of 2 prominent journals that suggests that researchers routinely conduct analyses using such questionable methods. The implications of these practices for psychological research are discussed, and the reasons for current practices are reviewed.

Since its initial development nearly a century ago (Spearman, 1904, 1927), exploratory factor analysis (EFA) has been one of the most widely used statistical procedures in psychological research. Despite this long history and wide application, the use of factor analysis in psychological research has often been criticized. Some critics have raised concerns about fundamental limitations of factor analysis for contributing to theory development (e.g., Gould, 1981; Hills, 1977; Overall, 1964). For instance, Armstrong (1967), in an article entitled "Derivation of theory by means

of factor analysis or Tom Swift and his electric factor analysis machine," argued that factor analysis had limited utility for aiding in the development of theory, because it could not be relied on to provide meaningful insights into data.[1] He attempted to demonstrate this point by creating artificial data with a known structure and then ostensibly showing that EFA failed to accurately represent the structure.[2] Other critics have not challenged the fundamental utility of EFA but have instead criticized the manner in which it is

[1] The name Tom Swift refers to a character in a popular series of juvenile science fiction novels published in the 1960s. In each novel, Tom Swift makes use of futuristic devices with near-miraculous powers. Armstrong's reference to this character in the title of his article highlighted what he regarded as the naive belief by researchers in fundamental utility of EFA.

[2] It is important to note that Armstrong (1967) was not the first person to examine the effectiveness of EFA procedures using data sets with a known underlying structure (e.g., see Thurstone, 1947; Cattell & Dickman, 1962; Cattell & Sullivan, 1962; Cattell & Jaspers, 1967). Interestingly, in these other cases, the authors concluded that appropriate EFA procedures were reasonably effective in revealing the known underlying structure of the data.

sometimes applied (e.g., Comrey, 1978; Ford, Mac-Callum, & Tait, 1986; Gorsuch, 1983; McNemar, 1951; Skinner, 1980). In this article, we primarily address the latter issue. That is, we explore the manner in which factor analysis is applied in psychological research and evaluate the soundness of current practices. However, we contend that these two issues are intertwined. The utility of factor analysis for theory development is dependent on the manner in which it is implemented (see Cattell, 1978; Comrey, 1978). Furthermore, we suggest that some critics who have questioned the fundamental value of factor analysis have not been sufficiently sensitive to this relationship.

We begin our discussion by reviewing some of the major methodological decisions that researchers must make when conducting a factor analysis. Next, we illustrate with published data sets how poor choices when making these decisions can substantially distort the results. We then turn our attention to the extent to which current use of factor analysis reflects sound practices. We conclude with discussions of the implications of current factor analytic practices for psychological theory and reasons for the prevalence of certain practices.

Methodological Issues in the Implementation of Factor Analysis

Perhaps more than any other commonly used statistical method, EFA requires a researcher to make a number of important decisions with respect to how the analysis is performed (see Finch & West, 1997). Specifically, there are at least five major methodological issues that a researcher should consider when conducting a factor analysis. First, he or she must decide what variables to include in the study and the size and nature of the sample on which the study will be based. Second, a researcher must determine if EFA is the most appropriate form of analysis given the goals of the research project. Third, assuming that EFA is appropriate, a specific procedure to fit the model to the data must be selected. Fourth, the researcher must decide how many factors should be included in the model. Finally, it is usually necessary for a researcher to select a method for rotating the initial factor analytic solution to a final solution that can be more readily interpreted. Each of these decisions can have important consequences for the results obtained (see Armstrong & Soelberg, 1968; Cattell, 1978; Comrey, 1978; Ford et al., 1986; MacCallum, 1983; MacCal-

lum, Widaman, Zhang, & Hong, 1999; Velicer & Fava, 1998; Weiss, 1976). To the extent that a researcher makes poor decisions, the analysis is more likely to provide misleading results. However, researchers often appear to be unaware of the issues involved in these decisions.

Decisions in Conducting an EFA

Study design. As with any statistical procedure, the utility of the results obtained in EFA is in large part determined by the soundness of the design of the study from which the data are collected. Within the context of EFA, one design issue that is especially important is what measured variables to include in the study (Cattell, 1978). If a researcher inadequately samples measured variables from the domain of interest, he or she may fail to uncover important common factors. Conversely, if measured variables irrelevant to the domain of interest are included, then spurious common factors might emerge or true common factors might be obscured. Therefore, researchers should carefully define their domain of interest and specify sound guidelines for the selection of measured variables.

Research suggests that EFA procedures provide more accurate results when each common factor is represented by multiple measured variables in the analysis (i.e., when common factors are "overdetermined"; MacCallum et al., 1999; see also Velicer & Fava, 1998). Methodologists have recommended that at least three to five measured variables representing each common factor be included in a study (MacCallum et al., 1999; Velicer & Fava, 1998). Thus, when designing studies for which EFA is likely to be used, a researcher should consider the nature and number of common factors he or she expects might emerge. The total number of measured variables included should be at least 3 to 5 times the number of expected common factors, and the selected variables should include multiple variables likely to be influenced by each of the common factors. Alternatively, in cases in which there is little or no basis to anticipate the number and nature of common factors, a researcher should attempt to delineate as comprehensively as possible the population of measured variables for the domain of interest. He or she should then include in the study a sample of these measured variables that is as large as feasible (see Cattell, 1978).

Sound selection of measured variables also requires consideration of psychometric properties of measures. When EFA is conducted on measured variables with

low communalities (i.e., variables for which the common factors explain little variance), substantial distortion in results can occur (MacCallum et al., 1999; Velicer & Fava, 1998). There are a number of reasons why communalities for measured variables might be low. One obvious reason is low reliability. As explained later, variance due to random error cannot, by definition, be explained by common factors. Because of this, variables with low reliability will have low communalities and thus should be avoided. A second reason why a variable might have a low communality is that the variable is unrelated to the domain of interest and thus shares little in common with other measured variables in that domain. Therefore, to the extent such information is available, a researcher should consider the validity (e.g., face validity, convergent validity) of measured variables when selecting items to include in the analysis.

A second important design decision is the selection of the sample. A researcher must determine how large the sample should be and how that sample will be selected from the population of interest. Methodologists have proposed a host of rough guidelines for estimating an adequate sample size for an EFA. Most of these guidelines involve determining sample size based on the number of measured variables included in the analysis—with more measured variables requiring larger sample sizes. Sometimes such guidelines also specify a minimum necessary sample size regardless of the number of measured variables.

Unfortunately, there are serious drawbacks to such guidelines. One problem is that these recommendations vary dramatically. For instance, Gorsuch (1983) suggested a ratio of 5 participants per measured variable and that the sample size never be less than 100. In contrast, Nunnally (1978) and Everitt (1975) proposed ratios of 10 to 1. More important, recent research has suggested that such guidelines are not sufficiently sensitive to a variety of important characteristics of the data (MacCallum et al., 1999; Velicer & Fava, 1998). The primary limitation of such guidelines is that adequate sample size is not a function of the number of measured variables per se but is instead influenced by the extent to which factors are overdetermined and the level of the communalities of the measured variables. When each common factor is overdetermined (i.e., at least three or four measured variables represent each common factor) and the communalities are high (i.e., an average of .70 or higher), accurate estimates of population parameters can be obtained with samples as small as 100 (MacCallum et

al., 1999). However, under more moderate conditions a sample size of at least 200 might be needed; when these conditions are poor it is possible that samples as large as 400 to 800 might not be sufficient.

It is worth noting that obtaining parameter estimates that closely approximate population values is only one criterion a researcher might consider when determining sample size. In some situations, additional concerns might also play a role. Most notably, when EFA involves the testing of formal hypotheses regarding model fit or parameter estimates (as is sometimes done in maximum likelihood [ML] EFA), statistical power might also be considered. A researcher could specify a hypothesis of interest, a desired level of power, and an assumed population value for model fit. The sample size necessary to achieve these objectives can then be calculated (see MacCallum, Browne, & Sugawara, 1996).

Researchers should also consider the nature of the sample on which the study is based. Psychologists often select samples based on convenience. In many cases, this practice will not pose a problem. However, if the sample is considerably more homogeneous than the population on the common factors, this can lead to restriction of range in the measures, thereby attenuating correlations among variables. This attenuation can result in falsely low estimates of factor loadings and correlations among factors (see Comrey & Lee, 1992; Gorsuch, 1983; Tucker & MacCallum, 1997). Additionally, selection biases related to a single measured variable in the analysis can also distort results (Tucker & MacCallum, 1997). For these reasons, researchers should consider the nature of the measured variables they are investigating and the manner in which their sample is selected. Overly homogeneous samples and samples whose selection is related to measured variables in the analysis should be avoided. Thus, when there is a substantial basis to expect that convenience samples will not be appropriate, a researcher should consider obtaining a sample representative of the population of interest. Alternatively, a researcher might wish to select a sample to maximize variance on measured variables relevant to the constructs of interest and minimize variance on measured variables irrelevant to the constructs of interest (see Cattell, 1978).

Determining whether EFA is appropriate. The primary purpose of EFA is to arrive at a more parsimonious conceptual understanding of a set of measured variables by determining the number and nature of common factors needed to account for the pattern

of correlations among the measured variables. That is, EFA is used when a researcher wishes to identify a set of latent constructs underlying a battery of measured variables. Before using EFA, a researcher should carefully consider if this is a goal of the research project.

In reaching this decision, it is important to recognize that the goal of identifying latent constructs (i.e., understanding the structure of correlations among measured variables) is different from that of data reduction. Data reduction involves taking scores on a large set of measured variables and reducing them to scores on a smaller set of composite variables that retain as much information from the original variables as possible. Data reduction does not attempt to model the structure of correlations among the original variables. This distinction is important, because different methods have been designed to achieve these two objectives. If the goal is to arrive at a parsimonious representation of the associations among measured variables, EFA can be an appropriate form of analysis. If the goal is data reduction, principal components analysis (PCA) is more appropriate. Many researchers mistakenly believe that PCA is a type of EFA when in fact these procedures are different statistical methods designed to achieve different objectives (for a discussion of the distinction, see Bentler & Kano, 1990; Bookstein, 1990; Gorsuch, 1990; Loehlin, 1990; McArdle, 1990; Mulaik, 1990; Rozeboom, 1990; Schonemann, 1990; Steiger, 1990; Velicer & Jackson, 1990a, 1990b; Widaman, 1990).

EFA is based on the common factor model (Thurstone, 1947). This model postulates that each measured variable in a battery of measured variables is a linear function of one or more common factors and one unique factor. Common factors are unobservable latent variables that influence more than one measured variable in a battery and are presumed to account for the correlations (covariances) among the measured variables (i.e., two measured variables are assumed to be correlated, because they are influenced by one or more of the same common factors). Unique factors are latent variables that influence only one measured variable in a battery and do not account for correlations among measured variables. Unique factors are assumed to have two components: a specific factor component (i.e., systematic latent factors that influence only one measured variable) and an error of measurement component (i.e., unreliability in a measured variable). The goal of the common factor model is to understand the structure of correlations among

measured variables by estimating the pattern of relations between the common factor(s) and each of the measured variables (i.e., as indexed by factor loadings).

In contrast, PCA does not differentiate between common and unique variance. Rather, this approach defines each measured variable as a linear function of principal components, with no separate representation of unique variance. Mathematically, these principal components can be defined as linear composites of the original measured variables and thus contain both common and unique variance. Therefore, principal components are not latent variables, and because of this it is not conceptually correct to equate them with common factors (although in practice researchers often do so). Furthermore, whereas the goal of common factor analysis is to explain correlations among measured variables, the goal of PCA is to account for variance in the measured variables. That is, the objective of PCA is to determine the linear combinations of the measured variables that retain as much information from the original measured variables as possible. Thus, although PCA is often referred to and used as a method of factor analysis, it is not factor analysis at all.

Nevertheless, some methodologists have argued that PCA is a reasonable substitute for analyses of common factors and might even be superior (e.g., see Velicer & Jackson, 1990a, 1990b). They noted that PCA is computationally simpler than common factor analysis and therefore requires less computer memory and processing time (e.g., Velicer & Jackson, 1990a). They also argued that the two approaches generally produce very similar results (e.g., Arrindell & van der Ende, 1985; Velicer, 1977; Velicer & Jackson, 1990a; Velicer, Peacock, & Jackson, 1982). Also, they noted that common factor analysis procedures sometimes produce "Heywood cases" (e.g., Velicer & Jackson, 1990a)—situations in which a communality for a measured variable (i.e., the proportion of variance in the measured variable accounted for by the common factors) is estimated to be at 1 or greater than 1. Because it is impossible to account for more than 100% of the variance in a variable, such an estimate is potentially problematic. Finally, advocates of PCA note that it is determinate, whereas the common factor model is not (e.g., Steiger, 1979, 1990; Velicer & Jackson, 1990a). That is, it is possible to compute an individual person's score on a principal component, whereas it is not possible to do so for a common factor.

There are, however, reasons to question these arguments. Advances in the speed and memory capabilities of computers have made the advantage of the computational simplicity of PCA trivial. Also, although PCA and EFA do often produce similar results, there are some contexts in which this is not the case (e.g., see Bentler & Kano, 1990; Borgatta, Kercher, & Stull, 1986; Gorsuch, 1988, 1990; Hubbard & Allen, 1987; McArdle, 1990; Snook & Gorsuch, 1989; Tucker, Koopman, & Linn, 1969; Widaman, 1990, 1993). Differences in results are most likely when communalities are low (e.g., .40) and there are a modest number of measured variables (e.g., three) per factor (Widaman, 1993). Furthermore, when the data correspond to assumptions of the common factor model, EFA produces more accurate results than PCA (e.g., McArdle, 1990; Snook & Gorsuch, 1989; Tucker et al., 1969; Widaman, 1990, 1993). In contrast, when the data are relatively consistent with the assumptions of PCA (e.g., little unique variance present in measured variables), extraction of common factors does as well as extraction of principal components (e.g., see Gorsuch, 1990; McArdle, 1990; Velicer et al., 1982).

The occasional occurrence of Heywood cases should also not be regarded as a flaw of common factor analysis. Heywood cases often indicate that a misspecified model has been fit to the data or that the data violate assumptions of the common factor model (van Driel, 1978). Thus, Heywood cases can have diagnostic value (Velicer & Jackson, 1990a). In contrast, because such cases do not occur in PCA, such problems are not solved but simply go unnoticed (McArdle, 1990).[3]

Indeterminacy of individual factor scores in the common factor model need not be a problem. In most applications of EFA, this issue is irrelevant in that the objective is to identify common factors that account for the structure of the correlations among the measured variables. This goal does not require the computation of factor scores but rather only factor loadings and factor intercorrelations (McArdle, 1990). In those situations where factor scores might be of interest, researchers usually obtain scores for the purpose of assessing correlations of factors with other variables, or using factors as independent variables or dependent variables in regression models. However, the development of structural equation modeling makes it unnecessary to estimate factor scores to obtain such information. One can use structural equation modeling to specify a model with factors as corre-

lates, predictors, or consequences of other variables and obtain estimates of the relevant parameters without ever estimating factor scores. Additionally, recent developments have provided useful nonstructural equation modeling procedures for estimating the correlations between factors and other variables without the need to compute factor scores (Gorsuch, 1997).

Finally, advocates of common factor analysis note that it has certain advantages over PCA when the goal is to identify latent constructs. They argue that most measures used in psychological research contain some random error. Because EFA procedures reflect a recognition of this fact, whereas PCA does not, the common factor model is a more realistic model of the structure or correlations (e.g., see Bentler & Kano, 1990; Gorsuch, 1973; Loehlin, 1990). Additionally, the common factor model is testable, whereas the PCA model is not (e.g., see Bentler & Kano, 1990; McArdle, 1990). The common factor model specifies certain hypotheses about the data. Thus, it can be fit to data and the model rejected if the fit is poor. In contrast, because PCA does not involve a specific hypothesis to be tested, it does not provide information on which one could base a decision to reject the "model."

Therefore, there are clear conceptual distinctions between PCA and EFA. Although these approaches often produce similar results, this is not true in certain contexts. When the goal of the analysis is to identify latent constructs underlying measured variables, it is more sensible to use EFA than PCA (see Cattell, 1978; Gorsuch, 1983; McDonald, 1985; Mulaik, 1972).

Assuming that identifying latent variables that account for the correlations among measured variables is the goal of the research project, a researcher must then decide if an exploratory or confirmatory approach will be used. Both EFA and confirmatory factor analysis (CFA) are based on the common factor model, and both seek to represent the structure of correlations among measured variables using a rela-

---

[3] Furthermore, there might be some cases in which a Heywood case would not represent a misspecified model and might not be problematic. For example, circumstances in which population parameters are close to logical boundaries (e.g., if true communalities are close to 1.0) could produce Heywood cases simply because of expected variability in parameter estimates resulting from sampling error (van Driel, 1978).

tively small set of latent variables. However, EFA is primarily a data-driven approach. No a priori number of common factors is specified and few restrictions are placed on the patterns of relations between the common factors and the measured variables (i.e., the factor loadings). EFA provides procedures for determining an appropriate number of factors and the pattern of factor loadings primarily from the data. In contrast, CFA requires a researcher to specify a specific number of factors as well as to specify the pattern of zero and nonzero loadings of the measured variables on the common factors. Alternatively, a researcher might specify a priori a small set of competing models postulating differing numbers of factors, different patterns of factor loadings, or both.

In situations in which a researcher has relatively little theoretical or empirical basis to make strong assumptions about how many common factors exist or what specific measured variables these common factors are likely to influence, EFA is probably a more sensible approach than CFA. EFA is likely to be more desirable in these situations, because the number of plausible alternative models might be so large that it would be impractical to specify and test each one in CFA. Additionally, when a strong basis does not exist for identifying a single model or a few specific competing models, it is quite possible that a researcher might fail to identify a number of plausible models. Therefore, in these contexts, the data-driven approach of EFA seems advisable.

However, when there is sufficient theoretical and empirical basis for a researcher to specify the model or small subset of models that is the most plausible, CFA is likely to be a better approach. This is because CFA allows for focused testing of specific hypotheses about the data (e.g., see Finch & West, 1997; Wegener & Fabrigar, in press). Also, the a priori nature of CFA makes it less likely that a researcher will capitalize on chance characteristics in the data.

It is also often useful to use EFA and CFA in conjunction with one another. An EFA can be conducted in an initial study to provide a basis for specifying a CFA model in a subsequent study. Alternatively, if the sample size in a single study is sufficiently large, the sample could be randomly split in half. An EFA could then be conducted on one half of the data providing the basis for specifying a CFA model that can be fit to the other half of the data.

Choice of model-fitting procedure. If EFA is the most appropriate form of analysis, it is then necessary to decide what procedure will be used to fit the com-

mon factor model to the data. A number of model-fitting methods (i.e., factor-extraction procedures) are available. Most widely used among these are ML, principal factors with prior estimation of communalities, and iterative principal factors. These are different methods for estimating the parameters (factor loadings and unique variances) of the same model, the common factor model.[4]

Although these procedures fit the same model, each method does have certain advantages and disadvantages. The primary advantage of ML is that it allows for the computation of a wide range of indexes of the goodness of fit of the model. ML also permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals for these parameters (Cudeck & O'Dell, 1994). The primary limitation of ML estimation is its assumption of multivariate normality. When this assumption is severely violated, this procedure can produce distorted results (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992). On the other hand, principal factors methods (both iterated and noniterated) have the advantage of entailing no distributional assumptions. Principal factors are also less likely than ML to produce improper solutions (i.e., a solution with a Heywood case or a solution that fails to converge on a final set of parameter estimates; Finch & West, 1997). The major drawback of the principal factor methods is that they provide a much more limited range of goodness-of-fit indexes and generally do not allow for computation of confidence intervals and significance tests. Regardless of these differences, when the common factor model holds reasonably well in the population and severe violations of distributional assumptions are not present, solutions provided by these methods are usually very similar.

Selecting the number of factors. Determining how many factors to include in the model requires the researcher to balance the need for parsimony (i.e., a model with relatively few common factors) against the need for plausibility (i.e., a model with a sufficient number of common factors to adequately account for

---

[4] Although the common factor model makes a further conceptual distinction by decomposing unique variance into specific variance and error variance, methods of fitting the common factor model allow for estimates of the amount of unique variance only and do not provide separate estimates of specific variance and error variance.

the correlations among measured variables). In other words, the goal of the researcher is to determine the number of "major" factors underlying a battery of measures. Importantly, errors in selection of the number of factors in a model can have a substantial effect on the results obtained (e.g., Comrey, 1978; Fava & Velicer, 1992; Levonian & Comrey, 1966; Wood, Tataryn, & Gorsuch, 1996).

Traditionally, methodologists have regarded specifying too few factors in a model (i.e., underfactoring) as a much more severe error than specifying too many factors (i.e., overfactoring; see Cattell, 1978; Rummel, 1970; Thurstone, 1947). Empirical research has generally supported this notion. When too few factors are included in a model, substantial error is likely (Fava & Velicer, 1992; Wood et al., 1996). Measured variables that load on factors not included in the model can falsely load on factors included in the model, and poor estimates of the factor loadings can be obtained for measured variables that do actually load on the factors included in the model (Wood et al., 1996). Such distortions can result in rotated solutions in which two common factors are combined into a single common factor (thereby obscuring the true factor structure) and in solutions with complex patterns of factor loadings that are difficult to interpret (see Comrey, 1978).

Empirical research suggests that overfactoring introduces much less error to factor loading estimates than underfactoring (Fava & Velicer, 1992; Wood et al., 1996). Such models often result in rotated solutions in which the major factors are accurately represented and the additional factors have no measured variables that load substantially on them or have only a single measured variable that loads substantially on each additional factor. Nonetheless, overfactoring should be avoided (see Comrey, 1978; Comrey & Lee, 1992). Solutions with too many factors might prompt a researcher to postulate the existence of constructs with little theoretical value and thereby develop unnecessarily complex theories. Additionally, overfactoring can accentuate poor decisions made at other steps in a factor analysis. For example, PCA tends to produce loadings that are larger than factor loadings (see Snook & Gorsuch, 1989; Widaman, 1993). Thus, solutions using this approach can sometimes make minor components appear to be major components (Wood et al., 1996).

Given these consequences, it is not surprising that an extensive methodological literature has developed exploring the issue of determining the optimal number

of factors. A number of procedures for answering this question have been proposed. Perhaps the best known of these procedures is the Kaiser criterion of computing the eigenvalues for the correlation matrix to determine how many of these eigenvalues are greater than 1 (for discussion of the nature of eigenvalues, see Gorsuch, 1983). This number is then used as the number of factors. Although this procedure is appealing for its simplicity and objectivity, the approach has significant problems. First, it is often misapplied by referring to the eigenvalues of the correlation matrix with communality estimates in the diagonal (i.e., the reduced correlation matrix) rather than eigenvalues of the correlation matrix with unities in the diagonal (i.e., the unreduced correlation matrix; see Guttman, 1954; Kaiser, 1960). Application of this rule to the eigenvalues of the reduced correlation matrix is an erroneous procedure (Gorsuch, 1980; Horn, 1969). Second, as with any mechanical rule, the procedure can to some extent be arbitrary. For instance, it is not really meaningful to claim that a common factor with an eigenvalue of 1.01 is a "major" factor whereas a common factor with an eigenvalue of 0.99 is not. Finally, in numerous studies involving both principal components and common factors, this procedure has been demonstrated to lead to substantial overfactoring and occasionally to underfactoring (Cattell & Jaspers, 1967; Cattell & Vogelmann, 1977; Hakstian, Rogers, & Cattell, 1982; Linn, 1968; Tucker et al., 1969; Zwick & Velicer, 1982, 1986). In fact, we know of no study of this rule that shows it to work well.

Another widely known approach for determining the number of factors is the "scree test" (Cattell, 1966; Cattell & Jaspers, 1967). In this procedure, the eigenvalues of the correlation matrix (or the reduced correlation matrix) are computed and then plotted in order of descending values. This graph is then examined to identify the last substantial drop in the magnitude of the eigenvalues. A model with the same number of common factors as the number of eigenvalues prior to this last substantial drop is then fit to the data. This procedure is often used with eigenvalues from the unreduced correlation matrix. However, when the goal is to identify common factors, it is more conceptually sensible to examine the plot of eigenvalues from the reduced correlation matrix, because this matrix more directly corresponds to the common factor model. Regardless of which matrix is used, this procedure has been criticized (e.g., Kaiser, 1970) because of its subjectivity (i.e., there is no clear objective definition of what constitutes a "substantial" drop in mag-

nitude). Additionally, sometimes the obtained pattern of eigenvalues is ambiguous in that no clear substantial drop may be present. However, when strong common factors are present in data, studies indicate that this procedure functions reasonably well (Cattell & Vogelmann, 1977; Hakstian et al., 1982; Tucker et al., 1969).

A third factor-number procedure that has been investigated in the methodological literature is parallel analysis (Horn, 1965; Humphreys & Ilgen, 1969; Humphreys & Montanelli, 1975; Montanelli & Humphreys, 1976). This approach is based on a comparison of eigenvalues obtained from sample data to eigenvalues one would expect to obtain from completely random data (i.e., the predicted means of eigenvalues produced by repeated sets of random data). Suppose a set of measured variables observed in a given sample depends on m major common factors. Parallel analysis is based on the notion that the m largest sample eigenvalues of the reduced correlation matrix should be larger than the m largest expected values of eigenvalues obtained from repeated corresponding sets of random data (based on the same sample size and number of variables). The eigenvalues that would be expected from random data are then compared with the eigenvalues actually produced by the data, and a model is specified with the same number of common factors as real eigenvalues that are greater than the eigenvalues expected from random data. Like other objective mechanical rules, this procedure can sometimes be arbitrary in that a factor just meeting the criterion is retained, whereas a factor falling just below the criterion is ignored. This procedure is also not available in major statistical programs. Nonetheless, simulation research suggests parallel analysis functions fairly well (Humphreys & Montanelli, 1975). Parallel analysis methods have also been developed for use in components analysis and found to function well in this context (Allen & Hubbard, 1986; Lautenschlager, 1989; Longman, Cota, Holden, & Fekken, 1989; Zwick & Velicer, 1986).

The methods described to this point for determining the number of factors to retain involve analysis of eigenvalues obtained in principal factors or PCA. Although such methods of factoring are widely used, the ML method of factor extraction is becoming increasingly popular. The ML method has a more formal statistical foundation than the principal factors methods and thus provides more capabilities for statistical inference, such as significance testing and determination of confidence intervals. These capabilities allow

a researcher to adopt a somewhat different approach to determining the optimal number of factors.

One can conceptualize the number-of-factors issue as choosing the most appropriate model from a series of alternative factor analysis models that differ in their complexity (i.e., the number of factors). The objective is to select a model that explains the data substantially better than simpler alternative models (i.e., models with fewer factors) but does as well or nearly as well as more complex alternative models (i.e., models with more factors). Because factor analysis is a special case of structural equation modeling, many of the procedures used in structural equation modeling for model selection can be applied to EFA. That is, the use of ML estimation for EFA introduces a vast array of goodness-of-fit information that can be used to determine the appropriate number of factors (see Browne & Cudeck, 1992; Hu & Bentler, 1998; Marsh, Balla, & McDonald, 1988; Mulaik et al., 1989). These fit measures, which are intended to assess the degree to which a given model provides an approximation of observed correlations or covariances, can be applied to the number-of-factors problem. The general procedure is to fit models with a range of numbers of factors, beginning with zero, and increasing through some maximally interesting number. Fit measures for each model can then be evaluated, and a decision can be made as to the appropriate number of factors to retain. The desired model is that which constitutes a substantial improvement in fit over a model with one fewer factor but for which a model with one more factor provides little if any improvement in fit.[5]

One common statistic for assessing fit in ML factor analysis solutions is the likelihood ratio statistic (Lawley, 1940). If N is sufficiently large and the distributional assumptions underlying ML estimation are adequately satisfied, the likelihood ratio statistic approximately follows a chi-square distribution if the specified number of factors is correct in the popula-

---

[5] Ideally, the preferred model should not just fit the data substantially better than simpler models and as well as more complex models. The preferred model should also fit the data reasonably well in an absolute sense. When this is not the case, a researcher should exercise some caution in interpreting the results. A preferred model that fits the data poorly might do so, because the data do not correspond to assumptions of the common factor model. Alternatively, it might suggest the existence of numerous minor common factors.

tion. Thus, this statistic can be used to test the null hypothesis that the model holds exactly in the population with a given number of common factors. This test can be applied to a series of numbers of factors, beginning with zero and continuing until a nonsignificant test statistic is obtained, indicating that the common factor model with the corresponding number of factors is not rejected as a correct model in the population. Although this procedure is intuitively appealing in that it is based on a clear statistical rationale, it is susceptible to serious problems. Most importantly, it is highly influenced by sample size. When N is large, even trivial discrepancies between the model and the data are likely to give rise to rejection of the model with any reasonable number of factors (Hakstian et al., 1982; M. L. Harris & Harris, 1971; Humphreys & Montanelli, 1975; MacCallum, 1990). In situations where N is small, even large discrepancies between the model and the data may not be statistically significant, thereby leading to underfactoring (Humphreys & Montanelli, 1975). In addition, the use of the likelihood ratio test is problematic, because the null hypothesis of perfect fit is an unrealistic standard in that all models are approximations of reality (Browne & Cudeck, 1992; Cudeck & Henly, 1991; MacCallum, 1990). The realistic goal in factor analysis is to obtain a parsimonious solution that provides a good approximation to the real world; thus, the hypothesis of perfect fit is not generally of empirical interest.

Because of these problems, methodologists have developed an array of alternative "descriptive" measures of fit. These developments began with the work of Tucker and Lewis (1973), who proposed a "reliability coefficient" for ML factor analysis solutions, and have continued through the present day. Our view is that developments reviewed and illustrated by Browne and Cudeck (1992) provide a promising approach for assessing fit of these models. Browne and Cudeck focused on the Root Mean Square Error of Approximation (RMSEA) fit index (Steiger & Lind, 1980) and the Expected Cross-Validation Index (ECVI; Browne & Cudeck, 1989) and illustrated their use for determining the number of factors in EFA. RMSEA is an estimate of the discrepancy between the model and the data per degree of freedom for the model. It has been suggested that values less than 0.05 constitute good fit, values in the 0.05 to 0.08 range acceptable fit, values in the 0.08 to 0.10 range marginal fit, and values greater than 0.10 poor fit (see Browne & Cudeck, 1992; Steiger, 1989).[6] ECVI, on the other hand, is an estimate of how well the solution

obtained from one sample will generalize to other samples. Although no guidelines exist for interpreting ECVI in absolute terms, it is useful for relative comparisons among alternative models. The smaller the value of ECVI, the better the expected cross-validation of the model. An important property of ECVI discussed by Cudeck and Henly (1991) and illustrated by Browne and Cudeck (1992) is that it is sensitive to sample size. ECVI will tend to support the retention of simpler models (fewer factors) when N is small and more complex models (more factors) when N is large. One can think of ECVI as addressing the following question: Based on the available data, for how many factors can we obtain accurate and generalizable parameter estimates?

A major advantage of both these indexes is the availability of confidence intervals. Thus, a researcher can compare the point estimates and corresponding confidence intervals of these fit indexes for a series of models with varying numbers of factors. A model can be selected that shows fit that is substantially better than simpler models and fit that is comparable to that of more complex models. Of course, the logic of the approach illustrated by Browne and Cudeck (1992) of using RMSEA and ECVI to determine the number of factors in EFA can be extended to other descriptive fit indexes. Thus, a researcher might choose to examine additional indexes of model fit when determining the number of factors (for a recent evaluation of fit indexes, see Hu & Bentler, 1998).

Although the use of RMSEA, ECVI, and other descriptive measures of model fit has yet to be extensively tested within the context of determining the number of factors in EFA, there is a compelling logic to this approach. Furthermore, many of these indexes have been tested within the context of more general covariance structure models. Thus, users of ML factor analysis would do well to consider the use of these fit indexes when determining the number of factors to retain.

In summary, a number of procedures exist to determine the appropriate number of common factors.[7]

---

[6] Although these guidelines for RMSEA are generally accepted, it is of course possible that subsequent research might suggest modifications.

[7] One method that has been found to perform well in determining the number of principal components to retain is the minimum average partial procedure (Velicer, 1976; Zwick & Velicer, 1982, 1986). We do not discuss this pro-

Some of these procedures are highly problematic (i.e., eigenvalues greater than 1, the likelihood ratio statistic), whereas others are likely to perform reasonably well at least under a range of conditions (e.g., scree test, parallel analysis, descriptive indexes of model fit such as RMSEA). However, even the best procedures are not infallible. Furthermore, it is important to remember that the decision of how many factors to include in a model is a substantive issue as well as a statistical issue. A model that fails to produce a rotated solution that is interpretable and theoretically sensible has little value. Therefore, a researcher should always consider relevant theory and previous research when determining the appropriate number of factors to retain. It should also be noted that because any individual data set is likely to have its own idiosyncrasies, decisions regarding the appropriate number of factors can be further improved by examining the replicability of the decision over multiple data sets (Cattell, 1978). When a given number of factors for a battery of measured variables is shown to be appropriate in more than one data set (e.g., sensible rules for determining the number of factors give similar results, rotated solutions produce similar interpretable patterns of factor loadings), a researcher can be more confident that the optimal number of factors has been extracted.

Factor rotation. For any given solution with two or more factors (or principal components), there exists an infinite number of alternative orientations of the factors in multidimensional space that will explain the data equally well. This means that EFA models with more than one factor do not have a unique solution. Therefore, a researcher must select a single solution from among the infinite number of equally fitting solutions.

The criterion most commonly used for selecting among solutions in EFA is the property of simple

structure (Thurstone, 1947). Thurstone proposed that for any given set of mathematically equivalent solutions, the solution with the best "simple structure" would generally be the most easily interpretable, psychologically meaningful, and replicable. Thurstone used the term *simple structure* to refer to solutions in which each factor was defined by a subset of measured variables that had large loadings relative to the other measured variables (i.e., high within-factor variability in loadings) and in which each measured variable loaded highly on only a subset of the common factors (i.e., low factorial complexity in defining variables). Therefore, Thurstone suggested that factors be rotated in multidimensional space to arrive at the solution with the best simple structure.

A number of analytic rotation methods have been developed to seek simple structure, and numerous articles have been published comparing the utility of these various rotation procedures (e.g., Crawford & Ferguson, 1970; Dielman, Cattell, & Wagner, 1972; Gorsuch, 1970; Hakstian, 1971; Hakstian & Boyd, 1972; Hofmann, 1978). Although these rotation methods differ in a number of respects, perhaps the most fundamental distinction that can be made is between orthogonal and oblique rotations. Orthogonal rotations constrain factors to be uncorrelated. Though a number have been developed, varimax (Kaiser, 1958) has generally been regarded as the best orthogonal rotation and is overwhelmingly the most widely used orthogonal rotation in psychological research.

In contrast to orthogonal rotations, oblique rotations permit correlations among factors. One common misconception among researchers is that oblique rotations require factors to be correlated. This is not the case. If the solution with the best simple structure involves orthogonal factors, a successful oblique rotation will provide estimates of the correlations among factors that are close to zero and produce a solution that is quite similar to that produced by a successful orthogonal rotation (Harman, 1976). However, in situations in which the best simple structure is a solution with correlated factors, successful oblique rotations will produce solutions with correlated factors. Unlike orthogonal rotation, there is no single method of oblique rotation that is clearly dominant in psychological research. Several oblique rotation procedures are commonly used and have been found to generally produce satisfactory solutions. These include direct quartimin rotation (Jennrich & Sampson,

---

cedure, because it is a method for use with principal components and has not been developed for use with common factors. Another approach to determining the number of components is to examine the comparability of component scores across split halves of a sample (Everett, 1983). Unfortunately, formal approaches for determining the number of common factors based on the stability of common factors have not been extensively developed and tested. Nonetheless, approaches based on this logic seem like a promising direction for future research. Cattell (1978) suggested that true factors should replicate across samples whereas spurious factors should be unstable over samples.

1966), promax rotation (Hendrickson & White, 1964), and Harris-Kaiser orthoblique rotation (Harris & Kaiser, 1964).[8]

Some researchers have indicated a preference for orthogonal rotation because of its simplicity and conceptual clarity (e.g., Nunnally, 1978). However, there are a number of reasons to question the wisdom of this view. First, for many constructs examined in psychology (e.g., mental abilities, personality traits, attitudes), there is substantial theoretical and empirical basis for expecting these constructs (or dimensions of these constructs) to be correlated with one another. Therefore, oblique rotations provide a more accurate and realistic representation of how constructs are likely to be related to one another. Thus, from a substantive perspective, the restriction of uncorrelated factors imposed by varimax and other orthogonal rotations is often unwarranted and can yield misleading results. Second, because orthogonal rotations require factors to be oriented at 90° angles from one another in multidimensional space (i.e., uncorrelated factors) whereas oblique rotations allow for orientations of less than 90° (i.e., correlated factors), orthogonal rotations are likely to produce solutions with poorer simple structure when clusters of variables are less than 90° from one another in multidimensional space (i.e., when the true underlying factor structure is based on correlated factors). Finally, oblique solutions provide more information than orthogonal rotations. Oblique rotations produce estimates of the correlations among common factors. Knowing the extent to which factors are correlated with one another can often be useful in interpreting the conceptual nature of the common factors. Indeed, the existence of substantial correlations among factors suggests that higher order factors may exist. Correlation matrices of factors can in turn be analyzed to gain insight into the number and nature of these higher order factors and thereby further refine a researcher's understanding of the data (see Gorsuch, 1983). Because orthogonal rotations do not provide correlations among factors, it is impossible to determine if one or more higher order factors are present in the data.

Summary and Recommendations

The implementation of EFA requires a researcher to address five major types of methodological decisions. There are a number of options available for each decision, and the issues involved in selecting among these options are sometimes complex. Nonetheless, these decisions are not arbitrary in that

some options are clearly more optimal than others. Thus, the utility of an EFA is likely to be a function of the decisions made in the design of the study and the implementation of the analysis. Indeed, it is interesting that some notable examples of the supposed failure of EFA to provide valid insights into data involved poor factor analytic methodology. For instance, Armstrong's (1967) classic demonstration of the alleged limitations of EFA used a PCA, the eigenvalues-greater-than-1 rule, and varimax rotation. In light of the questionable nature these choices, it is not surprising the analysis failed to uncover the true structure of the data.[9] Given these facts, some clear recommendations can be made regarding how EFA should be conducted.

1. Study design. Because EFA results are likely to be more accurate if sensible decisions are made in selecting measured variables and samples, it is essential that researchers carefully consider both of these issues. Obviously, because EFA is used in situations where there is relatively little prior theory and empirical evidence, variable selection can be difficult. Nonetheless, to the extent it is possible, researchers should try to anticipate the number and nature of the factors they expect to obtain and use this as a guide for selecting variables. We suggest researchers include at least four measured variables for each common factor they expect to emerge and perhaps as many as six given that there is usually considerable uncertainty about the nature of the common factors and their relations to the measured variables. Furthermore, although EFA should be based on comprehensive sam-

---

[8] Unfortunately, the selection of oblique rotations offered in major statistical packages is quite limited. SPSS offers promax and Harris-Kaiser orthoblique rotations. Direct oblimin rotation is a family of rotations defined by different values of the delta parameter, which governs the obliqueness of the solution. The default value for this parameter in SPSS is 0, which corresponds to a direct quartimin rotation. A new EFA program, CEFA (comprehensive exploratory factor analysis; Browne, Cudeck, Tateneni, & Mels, 1998) offers a much wider range of rotations.

[9] Additionally, because the data created by Armstrong have a number of properties that violate assumptions of the common factor model (e.g., nonlinear relations between factors and measured variables), his data are not really appropriate for EFA. Nonetheless, we found that if one uses more optimal EFA procedures, the analysis produces a solution that provides a good representation of the underlying structure of the data, thus invalidating Armstrong's arguments against the use of factor analysis.

pling of measured variables in the domain of interest, we recommend that researchers carefully consider the psychometric properties of variables. Variables with low reliabilities (e.g., below .70) should be avoided, and when validity information is available, it should be used as a basis for selecting items. With respect to determining adequate sample size, the properties of the measured variables should be taken into account. Under good conditions (communalities of .70 or higher, four to five variables for each factor), a sample size of 100 might well be adequate (although it is always best to have larger sample sizes if possible). Under conditions of moderate communalities (e.g., .40 to .70) and moderate overdetermination of factors, a sample of 200 or more seems advisable. Finally, under poor conditions, no sample size may be sufficient to produce accurate estimates of the population parameters. However, it seems likely that a sample less than 400 will lead to distorted results.

2. *Appropriateness of EFA.* We urge researchers to carefully consider if EFA is the most appropriate form of analysis to meet their research objectives. EFA should be used when the primary goal is to identify latent constructs and there is insufficient basis to specify an a priori model (or small subset of models). CFA should be used when the goal is to identify latent constructs and a substantial basis exists to specify an a priori model or small subset of models. PCA should not be used as a substitute for EFA.

3. *Model-fitting procedures.* With respect to selecting one of the major methods of fitting the common factor model in EFA (i.e., principal factors, iterated principal factors, maximum likelihood), all three are reasonable approaches with certain advantages and disadvantages. Nonetheless, the wide range of fit indexes available for ML EFA provides some basis for preferring this approach. However, researchers should recognize that this procedure can produce misleading results when assumptions of multivariate normality are severely violated (see Curran et al., 1996; Hu et al., 1992). Therefore, we recommend that the distributions of measured variables be examined prior to conducting ML EFA. If nonnormality is severe (e.g., skew > 2; kurtosis >7; West, Finch, & Curran, 1995), one of several remedies might be employed (see West et al., 1995). Measured variables could be transformed to normalize their distributions. Corrections to fit indexes and standard errors could be performed (Bentler & Dudgeon, 1996; Browne, 1984; Satorra & Bentler, 1994). Alternatively, one might wish to use a principal factors procedure.[10]

4. *Determining the number of factors.* We suggest that researchers rely on multiple criteria when deciding on the appropriate number of factors to include in a model. In the use of principal factors methods, we recommend the scree test and parallel analysis using eigenvalues from the reduced correlation matrix. In ML factor analysis, we encourage the use of descriptive fit indexes such as RMSEA and ECVI as discussed by Browne and Cudeck (1992) along with more traditional approaches such as the scree test and parallel analysis. A sensible strategy in most cases would be to use multiple methods to make this decision and then carefully examine the rotated solution for the suggested model to confirm that it is interpretable and theoretically plausible. In situations in which the sample size is sufficiently large, a researcher might also wish to randomly split the data and examine the stability of the solution across the two halves (for a discussion of assessing factor comparability across samples, see Gorsuch, 1983). In contexts in which procedures suggest different numbers of factors or in which the procedures produce somewhat ambiguous results, the researcher should examine the subset of models that these procedures suggest are most plausible. The rotated solutions for these models can then be examined to see which model produces the most readily interpretable and theoretically sensible pattern of results (Comrey, 1978; Ford et al., 1986; Hakstian & Muller, 1973; Hakstian et al., 1982; C. W. Harris, 1967) and when possible which of these solutions is most stable over different data sets or split halves of data sets.

5. *Rotation.* Given the advantages of oblique rotation over orthogonal rotation, we see little justification for using orthogonal rotation as a general approach to achieving solutions with simple structure (Gorsuch, 1983). Instead, it is most sensible to first examine the solutions produced by one or more of the common methods of oblique rotation (e.g., promax, orthoblique, or direct quartimin). If this solution in-

---

[10] Another approach to addressing violations of normality is to create "item parcels" (i.e., composites of several measured variables that tap on the same common factor). Creation of such item parcels often results in measured variables with more normal distributions than the individual items (West et al., 1995). However, such an approach will often not be feasible in EFA, because little is known regarding which measured variables are influenced by the same common factors, and thus it will be difficult to form item parcels.

dicates that the factors are uncorrelated, then it is reasonable for the researcher to conduct a varimax rotation and use this as the basis for interpretation (though this solution will likely be very similar to the oblique solution). On the other hand, if at least some of the factors are found to be correlated with one another, it is most defensible to use the oblique solution as the basis for interpretation.

Illustration of Consequences of Choices in Analysis

As noted earlier, some critics of EFA have been insufficiently sensitive to the extent to which particular decisions about *how* the analysis is conducted can influence the results. In order to illustrate some of the effects that these decisions can have, we reanalyzed three data sets from a well-developed psychological literature on the tripartite model of attitude structure.

According to the tripartite model (Rosenberg & Hovland, 1960; Smith, 1947), attitudes have affective, cognitive, and behavioral components. That is, feelings experienced when in the presence of an attitude object, beliefs about the attributes of the object, and past or present behaviors relevant to the object make up the structure underlying overall summary evaluations (see Eagly & Chaiken, 1998; Petty, Priester, & Wegener, 1994; Wegener & Gregg, in press, for more recent discussions). Numerous studies have investigated the hypothesized tripartite nature of attitudes (e.g., Breckler, 1984; Kothandapani, 1971; Ostrom, 1969), and in particular, the affective and cognitive components have received a great deal of attention as separable aspects of attitude structure. For example, assessments of the affect and cognition associated with particular attitude objects have independently predicted attitudes in a variety of settings (e.g., Abelson, Kinder, Peters, & Fiske, 1982; Crites, Fabrigar, & Petty, 1994), and confirmatory analyses have supported multifactor (tripartite) rather than single-factor models of these attitude structures (Breckler, 1984). Moreover, attitudes primarily based on affect versus cognition predict different classes of behaviors (e.g., Millar & Tesser, 1986, 1989) and are differentially influenced by affective versus cognitive persuasive appeals (e.g., Edwards, 1990; Fabrigar & Petty, 1999; Millar & Millar, 1990). Therefore, there are compelling theoretical and empirical rationales for separating affect and cognition in attitude structure. Because of this, there is a strong basis for expecting that EFA should ideally recover separable affect, cognition, and

(perhaps) behavior factors when examining relevant data sets.

We addressed this question by reanalyzing three data sets from this literature. First, we analyzed two 9-item matrices from Breckler (1984). In these matrices, three items each had been designed to assess affective, cognitive, and behavioral aspects of attitudes, respectively (see Breckler, 1984, for detail on the measures). We took the third data set from Crites et al. (1994), in which general measures were developed to assess the affective and cognitive bases of attitudes (three types of measures for each). In order to take advantage of the existence of an additional measure of affect and an additional measure of cognition specific to one of the attitude objects (snakes; Breckler, 1984), we analyzed 8 items from Study 1 of the Crites et al. (1994) article, in which people responded to the attitude object "snakes" (see Breckler, 1984; Crites et al., 1994, for additional detail on the measures).

We wished to use these data sets to illustrate how different procedures in conducting EFA analyses could lead one to either accurately or inaccurately identify the number and nature of the common factors underlying these data sets. We submitted the matrices to a PCA, examining both a varimax (orthogonal) and a direct quartimin (oblique) rotation. We also submitted the same matrices to a ML factor analysis, again examining both the varimax and direct quartimin rotations. In order to determine the number of factors, we examined the eigenvalues from the unreduced correlation matrices applying the eigenvalues-greater-than-1 rule. For the ML factor analysis models, we also examined the RMSEA and ECVI indexes of model fit, as well as the scree test and parallel analysis (Montanelli & Humphreys, 1976) for the eigenvalues from the reduced correlation matrix.

The sample sizes were 138, 105, and 164, for Breckler (1984) Studies 1 and 2 and Crites et al. (1994) Study 1, respectively. Though the sample sizes are a bit smaller than one might want, the sampling error that might occur in small samples would only work against the observation of consistent results across data sets. Therefore, consistency across the samples would increase confidence in the obtained results, despite any concerns about small samples. The initial communality estimates (squared multiple correlations) are somewhat variable in each study (see Table 1). For Breckler, Study 1, the estimates ranged from .61 to .25 with $M = .42$. For Breckler, Study 2, the estimates ranged from .81 to .37 with $M = .62$.

Table 1

Squared Multiple Correlations (SMCs) for Items in Breckler (1984; Studies 1 and 2) and Crites, Fabrigar, and Petty (1994; Study 1)

	SMCs	
Item	Study 1	Study 2
Breckler		
Affect		
Thurstone affect	0.43	0.54
Mood Checklist (+)	0.25	0.37
Mood Checklist (−)	0.33	0.51
Behavior		
Action sequence	0.53	0.73
Distance	0.37	0.81
Thurstone behavior	0.56	0.77
Cognition		
Thurstone cognition	0.38	0.50
Semantic differential	0.61	0.76
Listed thoughts	0.35	0.58
Crites et al.		
Affect		
Multiresponse Checklist	0.82	
Dichotomous Checklist	0.78	
Semantic differential	0.78	
Thurstone affect	0.34	
Cognition		
Multiresponse Checklist	0.85	
Dichotomous Checklist	0.79	
Semantic differential	0.85	
Thurstone cognition	0.56	

For Crites et al., Study 1, the estimated ranged from .85 to 34 with $M = .72$. However, these average communalities, especially for Study 2 of Breckler (1984) and Study 1 of Crites et al. (1994), are actually higher than the initial communalities that often occur

in applied research (see later discussion). Of course, most variables in real data will have substantial unique variance, and such conditions should be especially likely to lead to differences between EFA (which separately represents common and unique variance in the model) and PCA (which does not differentiate between common and unique variance). Unfortunately, information about the distributional properties of the items in the Breckler (1984) data sets was unavailable. For the Crites et al. (1994) data, however, the skew (−0.77 to 0.73, $|M| = 0.41$) and kurtosis (−1.02 to 3.28, $|M| = 0.79$) of the items were far smaller than the recommended thresholds for questioning the adequacy of ML estimation methods (skew > 2, kurtosis > 7; West et al., 1995). Therefore, across the data sets we analyzed, consistency in solutions could provide a compelling illustration of how the various decisions can influence the results for real data.

Number of Factors

Table 2 provides the eigenvalues for the reduced and unreduced correlation matrices for each data set. As can be seen from the eigenvalues for the unreduced correlation matrix, the eigenvalues-greater-than-1 rule suggested two-factor and one-factor models for the Breckler (1984) data sets (Studies 1 and 2, respectively). For the Crites et al. (1994) data, the rule suggested a one-factor model (see Table 1). Scree plots of these values, though a bit ambiguous, would perhaps suggest one-factor models for all three studies. However, given that the goal is to identify common factors, it is more sensible to examine the scree plots of the eigenvalues from the reduced correlation matrices. These plots would once again perhaps sug-

Table 2

Eigenvalues for Breckler (1984; Studies 1 and 2) and Crites, Fabrigar, and Petty (1994; Study 1), From the Unreduced and Reduced Correlation Matrices

	Factor number								
Correlation matrices	1	2	3	4	5	6	7	8	9
Breckler Study 1									
Unreduced	3.84	1.20	0.98	0.85	0.65	0.47	0.41	0.35	0.25
Reduced	3.30	0.60	0.39	0.20	0.02	−0.10	−0.12	−0.21	−0.27
Breckler Study 2									
Unreduced	5.59	0.91	0.70	0.50	0.42	0.34	0.23	0.19	0.13
Reduced	5.24	0.40	0.34	0.05	−0.04	−0.07	−0.09	−0.13	−0.14
Crites et al. Study 1									
Unreduced	5.72	0.88	0.46	0.38	0.19	0.17	0.12	−0.09	
Reduced	5.48	0.50	0.07	0.01	−0.05	−0.06	−0.08	−0.09	

gest a one-factor model for the three data sets. However, on the basis of these scree plots, one might also argue for a four-factor model in the Breckler Study 1 data set, a three-factor model in the Breckler Study 2 data set, and two-factor model in the Crites et al. data set.

When one examines the parallel analysis performed on the reduced correlation matrix (Montanelli & Humphreys, 1976), a one-factor solution appears less reasonable. For Breckler (1984) Study 1, the first four eigenvalues expected for random data (0.51, 0.34, 0.25, 0.17) fall below the observed eigenvalues from the reduced matrix (3.30, 0.60, 0.39, 0.20). For Breckler (1984) Study 2, three of the first four eigenvalues expected for random data (0.60, 0.40, 0.29, 0.20) fall at or below the observed eigenvalues from the reduced correlation matrix (5.24, 0.40, 0.34, 0.05) with

Table 3
Selected Fit Indexes for Analyses of Breckler (1984; Studies 1 and 2) and Crites, Fabrigar, and Petty (1994; Study 1): RMSEAs and ECVIs From ML and PACE Factor Analyses

Model-fitting procedure	Number of factors			
	1	2	3	4
Breckler Study 1 (*n* = 138)				
RMSEA				
ML	0.15	0.12[a]	0.03[a]	0.00[a]
PACE	0.13	0.14	0.08	0.25
ECVI				
ML	1.04	0.79	0.58	0.59
PACE	0.93	0.87	0.65	0.97
Breckler Study 2 (*n* = 105)				
RMSEA				
ML	0.14	0.09	0.00	0.00
PACE	0.14	0.09	0.00	0.00
ECVI				
ML	1.16	0.85	0.68	0.76
PACE	1.11	0.85	0.69	0.76
Crites et al. Study 1 (*n* = 164)				
RMSEA				
ML	0.23	0.06	0.02	
PACE	0.23	0.06	0.04	
ECVI				
ML	1.41	0.42	0.40	
PACE	1.38	0.41	0.41	

Note. Noniterated principal factoring and PACE factoring produce similar solutions with no boundary estimates. (See Footnote 12 for a description of this analysis.) RMSEA = root mean square error of approximation; ECVI = expected cross-validation index; ML = maximum likelihood; PACE = partitioned covariance estimator.
[a] The solutions noted included one or two boundary estimates (0) for unique variances.

the fourth observed eigenvalue falling far below the fourth value for random data. For Crites et al. (1994) Study 1, two of the four eigenvalues expected for random data (0.42, 0.27, 0.19, 0.11) fall below the observed eigenvalues (5.48, 0.50, 0.07, 0.01).[11] Thus, the parallel analyses question the one-factor decision that might come from the eigenvalues-greater-than-1 rule. They suggest a three-factor solution in one of the Breckler data sets and a two-factor solution in the Crites et al. data set (both as expected given past data and theory), but they suggest a four-factor solution for Study 1 of the Breckler (1984) data.

The goodness-of-fit indexes, however, strongly suggest the theoretically expected factor structures (see Table 3). For the Breckler (1984) data sets, RMSEA shows poor fit with the one-factor model and presents substantial improvement with a three-factor model. Moreover, overall fit becomes quite good (see Browne & Cudeck, 1992) when the third factor is added. There is also very little overlap in the 90% confidence intervals (CI) for the two-factor (RMSEA = 0.12, CI = 0.09–0.16 for Study 1; RMSEA = 0.09, CI = 0.04–0.14 for Study 2) and three-factor models (RMSEA = 0.03, CI = 0.00–0.10 for Study 1; RMSEA = 0.00, CI = 0.00–0.00 for Study 2) but complete overlap between the three- and four-factor models (RMSEA = 0.00, CI = 0.00–0.07 for Study 1; RMSEA = 0.00, CI = 0.00–0.00 for Study 2). The ECVI has it lowest value (i.e., the smallest point estimate for a predicted discrepancy in a new sample) for the three-factor model. This pattern clearly suggests the three-factor model as superior to the one- or two-factor models. For the Crites et al. (1994) data, the expected two-factor structure is strongly suggested. The model has quite poor fit with one factor

_____

[11] As noted earlier, there have also been parallel analyses developed for determining the number of components to retain (e.g., Longman et al., 1989). Given our interest in determining the number of factors underlying the data, however, it seems most reasonable to use the parallel analysis method designed for determining the number of common factors (i.e., parallel analysis with squared multiple correlations in the diagonal of the correlation matrix). Using parallel analyses for the unreduced matrix suggests extraction of a single principle component (though such a conclusion is questioned by the parallel analysis on the reduced matrix and by the indexes of EFA model fit). Moreover, one would want to be cautious about accepting a single factor model in this context given that underfactoring is often thought to be a more severe error than overfactoring.

and acceptable fit with two factors. Moreover, there is no overlap in the 90% CIs for the one-factor (RMSEA = 0.23, CI = 0.20–0.26) and two-factor models (RMSEA = 0.06, CI = 0.00–0.11) but nearly complete overlap between the two- and three-factor models (RMSEA = 0.02, CI = 0.00–0.10). Again, ECVI fails to improve beyond the level obtained with the theoretically expected number of factors.

When taken together, these analyses provide a useful illustration of how nonoptimal procedures can lead a researcher to misleading conclusions regarding the appropriate number of factors. The eigenvalues-greater-than-1 rule consistently suggested fewer factors than past research and theory would indicate was most plausible for each data set. The scree test provided somewhat ambiguous information. In contrast, parallel analysis suggested the most plausible number of factors for two of the three data sets, and model fit suggested the most plausible number of factors in all three data sets. Thus, consistent with the methodological literature, a researcher clearly would have been better served relying on parallel analysis and model fit than the eigenvalues-greater-than-1 rule when determining the number of factors for these data.

Orthogonal Versus Oblique Rotations

Researchers are often tempted to seek "conceptually distinct" factors by conducting varimax (orthogonal) rotations in factor analyses. It is sometimes thought that this retention of statistically independent factors "cleans up" and clarifies solutions, making them easier to interpret. Unfortunately, this intuition is exactly the opposite of what the methodological literature suggests, and the present results bear out the methodology-based conclusions. Tables 4, 5, and 6 present PCA and ML factor analysis results (the left and right columns) using varimax and direct quartimin rotations (the top and bottom portions of the tables). For ease of examination, loadings above 0.30 are presented in bold face. We have ordered factors similarly for each analysis (for the Breckler data sets, with the most "affective" factors first, most "behavioral" factors second, and most "cognitive" factors third; for the Crites et al. data, with the most "affective" factor first, and most "cognitive" factor second).

For each of the studies, it would have been unfortunate for a researcher to rely solely on a varimax rotation. Examining only those loadings above 0.30, there are many more cross-factor loadings (i.e., items that load above 0.30 for more than one factor) with the orthogonal rotation. Across the six varimax rota-

tions reported in Tables 4, 5, and 6, there are 28 times that an item is found to load on more than one factor (or component). However, the six direct quartimin rotations reported in Tables 4, 5, and 6 produce only eight cross-factor loadings. Examining the loadings below 0.30, the varimax rotation produced only 13% with magnitudes (ignoring sign) of less than 0.10; 42.6% had values between 0.10 and 0.20, and 44.4% had values greater than 0.20. Using the direct quartimin rotation, 60.8% of the loadings below 0.30 had magnitudes equal or less than 0.10; 25.3% had values between 0.10 and 0.20, and only 13.9% had values greater than 0.20 (despite including the 25 loadings that fell below 0.30 using direct quartimin but were above 0.30 using varimax). Therefore, it seems quite clear that direct quartimin (the oblique rotation) provided superior simple structure.

Besides the "cleaner" solutions provided by the oblique rotation, a researcher relying on an orthogonal rotation would also forfeit any knowledge of the existing correlations among factors. Although some of the varimax solutions for the current data (e.g., the ML factor analysis solution for Breckler, 1984, Study 1) might have led the researcher to think in terms of "affect, behavior, and cognition," he or she might have thought that there was little reason to think of the three factors as correlated (given the orthogonal rotation). Yet, when the direct quartimin rotation was used on the same data, it not only produced better simple structure clearly breaking along affective, behavioral, and cognitive lines, but it also revealed that the factors were correlated in the .4–.6 range (see Tables 4, 5, and 6).

Factor Analysis Versus Principal Components

Comparing the PCA and EFA columns also provides some useful information. As noted earlier, Widaman (1993) used simulation data to show that PCA inflates loadings when compared with factor analysis. Widaman found that "salient" loadings (i.e., the highest loadings on a given factor) are higher in PCA than in factor analysis and that such inflation is greater when the salient loadings are more moderate in value (e.g., 0.40 in the population) rather than high (e.g., 0.80 in the population). Because factor and component analyses lead to some differences in the patterns of loadings across factors in the current data, direct comparisons of loadings might not always be meaningful. Yet, an overall pattern consistent with Widaman's (1993) analyses seems to be present.

Table 4

Loadings for Principal Components and Common Factors Using Varimax and Direct Quartimin Rotations: Breckler (1984; Study 1)

Rotation Item	Principal components analysis (components)			Maximum likelihood factor analysis (factors)		
	1	2	3	1	2	3
Varimax						
Affect						
Thurstone affect	**0.83**	0.29	0.09	**0.99**	0.10	0.11
Mood Checklist (+)	**0.47**	**0.57**	−0.19	**0.43**	0.15	0.13
Mood Checklist (−)	**0.77**	−0.05	**0.34**	**0.46**	0.23	0.16
Behavior						
Action sequence	0.25	**0.66**	**0.37**	0.24	**0.86**	0.20
Distance	−0.05	**0.80**	0.25	0.15	**0.49**	**0.34**
Thurstone behavior	0.19	**0.66**	**0.46**	0.28	**0.60**	**0.40**
Cognition						
Thurstone cognition	0.03	0.11	**0.78**	0.10	0.17	**0.58**
Semantic differential	0.12	**0.39**	**0.77**	0.17	0.26	**0.95**
Listed thoughts	0.26	0.17	**0.68**	0.23	0.27	**0.47**
Direct quartimin						
Affect						
Thurstone affect	**0.82**	0.18	−0.06	**1.08**	−0.14	−0.05
Mood Checklist (+)	**0.40**	**0.57**	**−0.35**	**0.43**	0.04	0.03
Mood Checklist (−)	**0.81**	−0.20	0.26	**0.45**	0.13	0.03
Behavior						
Action sequence	0.14	**0.64**	0.23	0.03	**0.97**	−0.12
Distance	−0.20	**0.84**	0.12	−0.01	**0.47**	0.20
Thurstone behavior	0.08	**0.64**	**0.33**	0.10	**0.56**	0.21
Cognition						
Thurstone cognition	0.00	0.04	**0.78**	−0.01	−0.01	**0.62**
Semantic differential	0.04	**0.34**	**0.71**	−0.01	−0.04	**1.02**
Listed thoughts	0.23	0.08	**0.63**	0.12	0.13	**0.43**
Correlations among factors or components						
1	—			—		
2	.32	—		.51	—	
3	.18	.28	—	.42	.60	—

Note. Loadings in bold are values above 0.30.

If one takes all the loadings that are above 0.30 for at least one of the analyses (PCA or ML factor analysis) and compares the magnitude of those loadings across methods, the PCA loadings tend to be higher. For example, the salient loadings for Component 2 of the varimax rotation in Study 1 (in bold on Table 4) have values of 0.57, 0.66, 0.80, 0.66, and 0.39 (*M* = 0.62). The corresponding loadings for the ML factor analysis (Factor 2) are 0.15, 0.86, 0.49, 0.60, and 0.26 (*M* = 0.47). There are 16 such comparisons across the analyses presented, and 12 of the comparisons show higher salient loadings for components than for factors. Although existing work has generally exam-

ined possible inflation of salient loadings, some of the inflations in the current data are "nonsalient" loadings from factor analyses that become large enough for some researchers to consider them salient in defining a component. Such inflations would pose a potentially major problem, because they could change the presumed nature and interpretation of the constructs under investigation. One possible instance of such inflation is the creation of a sizeable loading of the Affect Dichotomous Checklist (0.57) in the PCA solution for Crites et al. (1994) Study 1 in which the direct quartimin rotation was used (see Table 4). The same loading (of an "affect" item on a primarily "cognitive"

Table 5

Loadings for Principal Components and Common Factors Using Varimax and Direct Quartimin Rotations: Breckler (1984; Study 2)

Rotation Item	Principal components analysis (components)			Maximum likelihood factor analysis (factors)		
	1	2	3	1	2	3
Varimax						
Affect						
Thurstone affect	**0.72**	**0.43**	0.18	**0.75**	**0.32**	0.20
Mood Checklist (+)	**0.89**	0.18	0.07	**0.63**	0.24	0.12
Mood Checklist (−)	**0.31**	**0.69**	0.19	**0.40**	**0.53**	0.28
Behavior						
Action sequence	0.27	**0.81**	0.27	**0.39**	**0.69**	**0.34**
Distance	0.25	**0.89**	0.19	**0.32**	**0.90**	0.24
Thurstone behavior	0.24	**0.84**	0.29	**0.36**	**0.73**	**0.36**
Cognition						
Thurstone cognition	0.08	0.22	**0.94**	0.13	0.22	**0.76**
Semantic differential	**0.36**	**0.55**	**0.65**	**0.43**	**0.44**	**0.70**
Listed thoughts	**0.53**	**0.49**	**0.39**	**0.57**	**0.38**	**0.39**
Direct Quartimin						
Affect						
Thurstone affect	**0.68**	0.24	0.03	**0.81**	0.02	0.01
Mood Checklist (+)	**0.98**	−0.11	−0.04	**0.71**	−0.01	−0.05
Mood Checklist (−)	0.11	**0.72**	−0.01	0.22	**0.49**	0.08
Behavior						
Action sequence	0.01	**0.87**	0.04	0.10	**0.72**	0.10
Distance	−0.05	**1.01**	−0.07	−0.09	**1.10**	−0.10
Thurstone behavior	−0.04	**0.90**	0.07	0.04	**0.79**	0.11
Cognition						
Thurstone cognition	−0.05	−0.07	**1.02**	−0.06	−0.01	**0.84**
Semantic differential	0.18	**0.36**	**0.56**	0.24	0.19	**0.63**
Listed thoughts	**0.42**	**0.32**	0.26	**0.51**	0.14	0.25
Correlations among factors or components						
1	—			—		
2	.59	—		.73	—	
3	.35	.56	—	.52	.64	—

Note. Loadings in bold are values above 0.30.

component) was 0.09 in the factor analysis. It is clear to see that the direct quartimin ML solutions are substantially "cleaner" than the same rotation of the PCA solutions.

Another notable difference between the PCA and ML solutions concerns the identified correlations among factors. Consistent with Widaman's (1993) simulations, the correlations identified by oblique rotation of the PCA solution were substantially lower than the correlations identified by the same rotation of the ML factor analysis. It makes sense that PCAs should generally underestimate relations among the constructs, because random error is included in the components. Because factor analyses remove random

error from the factors, the relations among factors are more likely to approach the population values. Such differences are especially evident in analyses of Breckler (1984) Study 1. The correlations among components (i.e., .32, .18, and .28) were roughly half the magnitude of the correlations among factors (i.e., .51, .42, and .60). If a researcher had used a PCA, he or she would have been tempted to conclude that the components were largely independent. However, using an EFA, the same researcher would have realized that the smallest of the interfactor correlations was actually greater than 0.4. These would certainly be rather different conclusions. Similar deflations of correlations occur for the PCAs of Breckler (1984) Study

Table 6

Loadings for Principal Components and Common Factors Using Varimax and Direct Quartimin Rotations: Crites, Fabrigar, and Petty (1994; Study 1)

Rotation Item	Principal components analysis (components)		Maximum likelihood factor analysis (factors)	
	1	2	1	2
Varimax				
Affect				
Multiresponse Checklist	**0.63**	**0.66**	**0.81**	**0.48**
Dichotomous Checklist	**0.67**	**0.61**	**0.80**	**0.44**
Semantic differential	**0.57**	**0.70**	**0.74**	**0.52**
Thurstone affect	**0.91**	0.09	**0.56**	0.20
Cognition				
Multiresponse Checklist	0.27	**0.89**	**0.38**	**0.87**
Dichotomous Checklist	**0.31**	**0.86**	**0.40**	**0.81**
Semantic differential	0.28	**0.90**	**0.41**	**0.85**
Thurstone cognition	0.15	**0.83**	**0.38**	**0.64**
Direct Quartimin				
Affect				
Multiresponse Checklist	**0.42**	**0.64**	**0.82**	0.15
Dichotomous Checklist	**0.49**	**0.57**	**0.84**	0.09
Semantic differential	**0.35**	**0.69**	**0.70**	0.24
Thurstone affect	**0.94**	−0.04	**0.66**	−0.09
Cognition				
Multiresponse Checklist	−0.04	**0.95**	−0.04	**0.98**
Dichotomous Checklist	0.01	**0.90**	0.02	**0.89**
Semantic differential	−0.04	**0.95**	0.01	**0.94**
Thurstone cognition	−0.15	**0.90**	0.11	**0.66**
Correlations among factors or components				
1	—		—	
2	.46	—	.77	—

Note. Loadings in bold are values above 0.30.

2 (see Table 5) and Crites et al. (1994) Study 1 (see Table 6).[12]

The analyses of the Breckler (1984) and Crites et al. (1994) data provide examples of how questionable

---

[12] It is worth noting that some parameter estimation problems were encountered in the ML analyses of the Breckler (1984) Study 1 data set. Heywood cases were encountered for the two-, three-, and four-factor models. It is perhaps not surprising that difficulties were encountered for one of the four-factor models given that this model is almost certainly inappropriate for the present data (i.e., the model is overfactored). One common reason for encountering such estimation problems is when a misspecified model is fit to the data. In the case of the three-factor model for Study 1, the theoretical plausibility of the solution and the comparability of this solution with the three-factor model solution for Study 2 and the two-factor solution from the Crites et al. (1994) data suggest that the existence of Heywood cases does not constitute a serious problem. Nonetheless, to examine if the results of the ML analyses had produced unusual results as a function of severe nonnormality in the data

(which could not be tested for the Breckler data sets), we conducted noniterated principal factor analyses for these data. These analyses produced results very similar to those of the ML analyses. Additionally, the noniterated principal factor solutions for the four-factor models confirmed that a four-factor model was in fact inappropriate for these data. The four-factor models (as was the case for the ML analyses) failed to produce readily interpretable or plausible solutions. No estimation problems were encountered in any of the analyses of the Crites et al. (1994) data set, and not surprisingly the noniterated principal factor analyses of these data produced solutions very similar to the ML analyses. Finally, to further test the stability of the ML solutions across all three data sets, we conducted analyses using Partitioned Covariance Matrix Estimator Factor Analysis (PACE; Cudeck, 1991). This procedure is available in the program CEFA (Browne et al., 1998). It is a noniterated common factor model-fitting procedure that permits the computation of the same fit indexes available in ML. The PACE analyses produced patterns of fit indexes across the models that were comparable to those of the ML analyses. The solutions obtained from the PACE analyses were also quite similar to those of the ML analyses.

procedures can produce results that are rather confusing and misleading. For all three studies, an EFA with an oblique rotation provides much better simple structure, more interpretable results, and more theoretically plausible representations of the data than a PCA with an orthogonal rotation. Moreover, the oblique rotation (especially when paired with EFA rather than PCA) provides important information concerning the relations among extracted factors. Across the example data sets, we see that different procedures for determining the appropriate number of factors, different analytic methods (PCA vs. EFA), and different rotation procedures all have an effect on results and interpretations.

The Use of Factor Analysis in Current Psychological Research

One obvious question that arises from our review and illustration is the issue of the extent to which researchers actually use the procedures we have shown to produce misleading results. An answer to this question is not readily available. Although methodologists have often commented on the use of factor analysis in psychology (e.g., Comrey, 1978; Ford et al., 1986; McNemar, 1951; Skinner, 1980; see also Gorsuch, 1983), most of these criticisms have been based on subjective impressions rather than systematic reviews of published applications. One exception is a review of EFA practices reported by Ford et al. (1986). They systematically examined applications of EFA published between 1975 and 1984 in the area of industrial–organizational psychology. They concluded that factor analytic practices within this area were generally inconsistent with the methodological literature. However, their review did not fully address study design issues, was confined to a single area of psychology, and covered applications published 14 or more years ago.

Assessing Current Factor Analytic Practices

To explore the question of how factor analysis is currently used in psychological research, we conducted a systematic review of articles published from 1991 through 1995 in the *Journal of Personality and Social Psychology* (*JPSP*) and the *Journal of Applied Psychology* (*JAP*). We selected these journals, because they represent two areas of psychology (personality–social psychology and industrial–organizational

psychology) in which EFA has been widely used. We also chose these particular journals, because both are among the most prestigious journals in their respective areas, so the articles found in these journals should presumably reflect methodologically rigorous work.

In our review, we examined every article published in these journals during the specified time period to determine if any of the statistical analyses reported in the article addressed an exploratory factor analytic question. If one or more analyses of this type was conducted, we then examined the description of each analysis. Studies were coded for the ratio of measured variables to factors, average reliability (or communality) of the measured variables, sample size, examination of common factors versus principal components, model-fitting procedure, method for determining the number of factors, and rotation procedure. The results of this review are presented in Table 7. The first two columns of the table indicate the number of articles and corresponding percentage of total number of articles falling into each category for *JPSP*. The third and fourth columns present the same information for *JAP*.

There are several aspects of these results that merit commentary. First, our review indicates that EFA continues to be an extremely popular statistical procedure in psychological research. A total of 159 of the 883 articles published in 60 issues of *JPSP* over a 5-year period reported EFAs. A total of 58 of 455 articles published in 30 issues of *JAP* over the same 5-year period reported EFAs. Thus, the typical issue of these journals contained two or three articles using EFA.

The first section of Table 7 shows the distribution of variable to factor ratios. These results indicate that the great majority of analyses were based on ratios of at least 4:1. Thus, in most analyses, factors were probably adequately overdetermined. However, there was a nontrivial number of articles (i.e., about one in five) with ratios below 4:1. This finding is noteworthy, because one condition in which PCAs produce substantially inflated estimates of factor loadings is when the ratio is 3:1 or less (Widaman, 1993).

The next section of Table 7 shows the distribution of the average reliability of measured variables in analyses. These results show that when researchers reported the reliability of their measured variables, the values were generally high (i.e., .70 or greater). One might conclude from this finding that the psychometric properties of variables analyzed in EFA are typi-

Table 7

Summary Information of Current Practices in the Use of Factor Analysis

Variable	Journal of Personality and Social Psychology		Journal of Applied Psychology	
	N	%	N	%
Ratio of variable to factors				
Less than 3:1	1	0.6	1	1.7
3:1	28	17.6	9	15.5
4:1	26	16.4	10	17.2
5:1	14	8.8	10	17.2
6:1	13	8.2	6	10.3
More than 6:1	74	46.5	18	31.0
Unknown	2	1.3	4	6.9
Average reliability of variables				
Less than .60	3	1.9	2	3.4
.60–.69	6	3.8	5	8.6
.70–.79	33	20.8	9	15.5
.80–.89	33	20.8	11	19.0
.90–1.00	14	8.8	9	15.5
Unknown	70	44.0	22	37.9
Sample size				
100 or less	30	18.9	8	13.8
101–200	44	27.7	14	24.1
201–300	25	15.7	9	15.5
301–400	13	8.2	2	3.4
More than 400	47	29.6	25	43.1
Type of analysis				
Principal components	84	52.8	28	48.3
Common factors	31	19.5	13	22.4
Multiple methods	8	5.0	2	3.4
Other	1	0.6	0	0.0
Unknown	35	22.0	15	25.9
Factor–component number procedure				
Eigenvalue > 1.0	25	15.7	11	19.0
Scree Test	24	15.1	9	15.5
Parallel analysis	1	0.6	0	0.0
Model fit	0	0.0	0	0.0
Theory	2	1.3	4	6.9
Interpretability	4	2.5	0	0.0
Multiple methods	35	22.0	12	20.7
Other	2	1.3	0	0.0
Unknown	66	41.5	22	37.9
Factor–component rotation				
Varimax	87	54.7	28	48.3
Harris-Kaiser	1	0.6	1	1.7
Promax	2	1.3	2	3.4
Direct quartimin	21	13.2	9	15.5
No rotation	23	14.5	4	6.9
Multiple methods	3	1.9	2	3.4
Other	1	0.6	0	0.0
Unknown	21	13.2	12	20.7

cally quite good. However, in both journals about 40% of analyses did not include reports of the reliability of the measured variables. A large number of the analyses in which reliabilities were not reported involved single-item measures, whereas those for which reliabilities were reported involved multi-item measures. Because single-item variables are likely to be considerably less reliable than multi-item variables, it seems probable that many of the analyses reported in these journals were based on variables with less than optimal psychometric properties. Also of interest is the fact that we found virtually no cases in which authors reported the communalities of their measured variables. This practice is unfortunate given that, in some ways, the communalities are more informative than the reliabilities regarding the soundness of the EFA results. However, for 18 data sets we were able to obtain the correlation matrices on which the authors' factor analyses were based. This allowed us to index the communalities by examining the squared multiple correlations associated with each data set. The average communality associated with each data set ranged from .12 to .65 with the median of these averages being .425.

The third section of Table 7 presents the distribution of sample sizes across articles. More than a third of the articles across the two journals conducted EFA based on modest-to-small sample sizes (i.e., samples of 200 or less). A somewhat smaller number of articles used moderate sample sizes (i.e., 201 to 400), and about a third or more used large samples (i.e., greater than 400). Thus, there was a substantial number of articles based on sample sizes sufficiently small that results could have been distorted if the measured variables included in the analysis were less than optimal or the factors were underdetermined.

The next section of Table 7 indicates the greater popularity of PCA relative to EFA. Approximately half of the published applications reported using the PCA method. This method was used despite the fact that in the vast majority of these articles, the primary goal was to identify latent constructs underlying measured variables rather than data reduction per se. In contrast, only about 20% of analyses used some form of EFA (with some type of principal factors analysis accounting for three out of every four articles using common factor analysis). In approximately a fourth of the articles in both journals, it was impossible to determine what method was used.

The next section of Table 7 reports the distribution of articles across different procedures for determining

the number of factors. The eigenvalue-greater-than-1 rule was the most popular single procedure in both journals, followed closely by the scree test. Use of other methods of determining the number of factors was relatively rare. A fairly sizable number of articles (i.e., about 20%) reported using multiple methods. In the overwhelming majority of these cases, multiple methods involved some combination of the eigenvalues-greater-than-1 rule, scree test, or factor interpretability—a priori theory. However, it was extremely common (i.e., about 40% of the time) for authors to fail to clearly indicate how they arrived at their decisions as to the number of factors to include.

The final section of Table 7 provides the distribution of articles across different methods of rotation. Varimax rotation was clearly the most commonly used rotation with approximately half of all factor analyses in both journals using this procedure. The second most popular method of rotation was direct quartimin although this rotation was used in only about a third to a fourth as many articles as varimax rotation. A nontrivial number of articles (i.e., 13% to 21%) failed to indicate the rotation used.

When viewed in their entirety, the results reported in Table 7 are discouraging. A fair number of analyses were based on studies in which at least one design feature was marginal (i.e., low measured variable to factor ratios, modest sample sizes, or both). A sizable number of analyses provided no information regarding the psychometric properties of the measured variables. The most popular types of analysis and rotation (i.e., PCA and varimax rotation) were not optimal or even near-optimal choices. In the case of determining the number of factors, the largest proportion of published articles indicated that multiple methods were used. Such an approach seems sensible in light of the methodological literature. However, exclusive use of the largely discredited eigenvalues-greater-than-1 rule was the second most popular choice. Thus, it appears that the same poor choices found to be problematic in the methodological literature and demonstrated to produce the misleading results in our examples are popular among researchers. Equally disconcerting, the results in Table 7 indicated that researchers often failed to provide adequate information about the procedures used. In 15.7% of *JPSP* factor analyses and 22.4% of *JAP* factor analyses it was impossible to determine what procedures were used for at least two of the three major analytic decisions. In more than half of all factor analyses reviewed, information was not provided concerning at least one of the decisions.

Implications of Current Practices for Psychological Research

Given the prevalence of poor decision making by researchers in the use of EFA, many readers might wonder what the implications of such practices are for the interpretations and conclusions reached in these articles. A precise answer to this question is not possible. Although the methodological literature suggests (and the examples presented in this article illustrate) that different choices in designing studies and selecting factor analytic procedures can produce substantial differences in results, there are clearly many cases in which this does not occur. The methodological literature has delineated some of the conditions under which differences in procedures will produce substantial differences in results. However, understanding of these issues is far from complete. Furthermore, in many cases, it is impossible to know how often these conditions exist in actual data. One method for answering this question would be to conduct comprehensive reanalyses of published data sets. Unfortunately, such an approach is not practical. Beyond the fact that such a study would require the reanalysis of hundreds of data sets, access to data sets is very difficult. It is relatively rare for researchers to report the correlation matrix on which their factor analyses are based, and requesting access to data is often unsuccessful.[13]

Nonetheless, we were able to address this question in a somewhat informal manner by reanalyzing those data sets for which we had access to the correlation matrices (a total of 18 data sets). These analyses revealed a number of cases in which extracting factors versus components, using different procedures to determine the number of factors, or using different rotation methods produced changes in results sufficiently substantial to have altered the conclusions a

---

[13] Interestingly, when we contacted authors to request their data, we found that a relatively small proportion of them agreed to our request. In the majority of cases, authors did not answer our request even after reminder letters were sent. In other cases, authors indicated that the data were no longer available. Finally, in a few cases, authors indicated that they would send the data but did not or directly refused to send the data at all. This last situation is particularly interesting given that current APA ethical guidelines specify that authors publishing in APA journals must make their data available to third parties for a period of 5 years following publication.

researcher might have reached concerning the data. For example, we compared the solutions produced by ML factor analysis and PCA for models with the same number of factors retained by the authors and using a rotation comparable to that used by the authors. We found that there were differences between PCA and ML sufficiently substantial that a researcher could have reached a different interpretation of the factors in 5 of the 18 data sets. That is, for 5 of these data sets, one or more measured variables changed the factors on which they had their primary loadings across the two types of analysis. Interestingly, for 4 of the data sets, it was impossible to compare the PCA and ML solutions. The solutions could not be compared, because the authors in their original analyses retained so many components relative to the number of measured variables that the corresponding ML factor analysis models would have had negative degrees of freedom.

Using the same models (i.e., common factor or principal component) and fitting procedures used by the authors and the same number of factors retained by the authors, we also compared varimax rotation and direct quartimin rotation. We found that direct quartimin rotations often produced slightly better simple structure than varimax rotation when the factors were correlated, but the pattern of loadings was substantially different (i.e., there were changes in the factors on which one or more of the measured variables had their primary loadings) in only 1 of the data sets.

Finally, using the eigenvalues-greater-than-1 rule, scree test, and model fit (as indexed by RMSEA), we also examined the issue of the appropriate number of factors. In 7 of the 18 data sets, the scree test, model fit, or both contradicted the eigenvalues-greater-than-1 rule. Interestingly, when we examined the results of the scree test and model fit (two of the better procedures for determining the number of factors), we found that the results often contradicted the decisions made by the authors in the original articles. In 3 of the data sets, both model fit and the scree test clearly suggested a different number of factors than that retained by the authors. In 5 of the data sets, model fit clearly suggested a different number of factors. In 1 data set, the scree test clearly suggested an alternative number of factors.[14]

Although this small sample of data sets cannot be regarded as fully representative of the entire applied EFA literature, it does suggest that differences in results due to changes in EFA procedures might be more common than many researchers realize. Addi-

tionally, our survey indicated that many researchers made more than one poor decision. The combination of several poor decisions is likely to produce even more serious distortions. For example, it was relatively common for researchers to conduct a PCA, retain as many factors as eigenvalues greater than 1, and conduct a varimax rotation. This particular "package" of decisions is especially likely to result in poor recovery of the underlying factors.[15]

General Discussion

We began our article by noting that researchers must consider five major methodological issues when conducting a factor analysis. We argued that the methodological literature suggests that not all options available to researchers for each of these decisions are equally sound and that poor choices in designing the study and conducting the factor analysis can produce poor results. To further illustrate this point, we reanalyzed previously published data sets and showed that the (mis)use of EFA can produce misleading results. We demonstrated that the misleading results obtained in these analyses did not really reflect a fundamental weakness in EFA but instead demonstrated the consequences of using questionable analytic procedures. We then went on to establish that the same questionable procedures used in our examples are in fact quite prevalent in current empirical research using factor analysis.

We believe that our review highlights two important points that should be brought to the attention of researchers using EFA and to readers of articles in which EFAs are reported. First, contrary to what many researchers probably believe, the decisions in the design of studies and in selecting factor analytic procedures are not arbitrary and inconsequential. There is reason to consider some design features and

---

[14] In those cases in which model fit clearly contradicted either the eigenvalues-greater-than-1 rule or the original decisions made by the authors but the scree test did not, it was almost never the case that the scree test supported the eigenvalues-greater-than-1 rule or the original decision. Instead, in these cases, the scree test was sufficiently ambiguous that it was difficult to argue that it clearly indicated an appropriate number of factors.

[15] It is perhaps not surprising that this particular choice of options is so popular. This set of decisions are the default options for the factor analysis procedure in SPSS.

EFA procedures to be markedly better than others. Second, this article suggests that the quality of EFAs reported in psychological research is routinely quite poor. Researchers sometimes base their analyses on studies with less than optimal features, commonly make questionable choices when selecting analytic procedures, and do not provide sufficient information for readers to make informed judgements of the soundness of the EFA being reported.

One obvious question that arises from these observations is why factor analysis is so poorly used. There are several possible reasons. One is that researchers are ill-informed regarding the use of EFA. Although there is a substantial methodological literature on EFA, much of this literature is relatively complex and published in quantitatively oriented journals that most psychologists are unlikely to read on a regular basis. With a few exceptions (e.g., Finch & West, 1997), nontechnical, concise, and up-to-date reviews of this literature have generally not been available to researchers. Furthermore, most researchers probably receive little formal training in EFA. Although it is common for graduate programs to require their students to take courses on the use of analysis of variance (ANOVA), it is much rarer for this to be the case for EFA (see Aiken, West, Sechrest, & Reno, 1990). Many graduate programs do not offer courses that cover factor analysis or devote only 1 week or 2 to this topic as part of a broad survey course covering different multivariate statistical procedures. Therefore, it is not surprising that many researchers are relatively uninformed regarding the implications of study design features and choosing different factor analytic procedures.

A second major reason why factor analyses are poorly conducted is simple tradition. There is a strong tendency for researchers to conduct analyses in a manner that is similar to what has been done before. Researchers do so because (a) they wish for their results to be directly comparable to past studies, (b) they naively believe that procedures must be reasonable if so many people have used them in the past, or (c) they feel (perhaps correctly) that the surest way to avoid difficulties in the peer review process is to do what has been done before.

Finally, another reason for the poor use of EFA has to do with the statistical software currently popular in psychological research. Such programs are likely to exert a tremendous influence on the way analyses are conducted. For example, many researchers probably follow the default options of their programs, because they believe that these options would not be the defaults unless they were the most acceptable methods currently available. Additionally, if a particular procedure is not offered in these programs, it is unlikely that a researcher will or can even be expected to use it. Unfortunately, the factor analytic procedures offered in the major statistical programs are far from ideal (e.g., see MacCallum, 1983; Wood et al., 1996). Given the inadequacies of these programs, it is not surprising that factor analyses conducted in psychological research are often far from optimal.

Although it is relatively easy to understand why EFA is often misused, it is more difficult to formulate how such practices might be changed. It seems unlikely that psychology departments will be able or willing to invest the resources necessary to substantially upgrade their statistics courses. However, methodologists could accept a greater responsibility for educating the research community regarding the use of EFA and other statistical procedures. This would require them not only to write highly technical papers targeted at the quantitative methodology community (which are undeniably important) but also to write less technical papers clearly explaining the practical implications of this methodological research. Furthermore, editors must be willing to publish these articles in nonquantitative journals that are likely to be read by researchers. The use of EFA might also be improved by editors of journals adopting higher standards for the manner in which factor analyses are conducted and reported. At the very least, researchers should be required to report what procedures they used when conducting an EFA. Researchers should also be expected to offer a brief rationale for their design decisions and choices of EFA procedures. Finally, developers and users of EFA should more actively pressure the manufacturers of the major statistical programs to improve their products. Taken together, these initiatives would be a good start toward improving the use of EFA in psychological research.

References

Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality of Social Psychology, 42,* 619–630.

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. *American Psychologist, 45,* 721–734.

Allen, S. J., & Hubbard, R. (1986). Regression equations for the latent roots of random data correlation matrices with unities in the diagonal. *Multivariate Behavioral Research, 21,* 393–398.

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. *The American Statistician, 21,* 17–21.

Armstrong, J. S., & Soelberg, P. (1968). On the interpretation of factor analysis. *Psychological Bulletin, 70,* 361–364.

Arrindell, W. A., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement, 9,* 165–178.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology, 47,* 563–592.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25,* 67–74.

Bookstein, F. L. (1990). Least squares and latent variables. *Multivariate Behavioral Research, 25,* 75–80.

Borgatta, E. F., Kercher, K., & Stull, D. E. (1986). A cautionary note on the use of principal components analysis. *Sociological Methods and Research, 15,* 160–168.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47,* 1191–1205.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445–455.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21,* 230–258.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (1998). CEFA: Comprehensive exploratory factor analysis. Authors. [Computer program]. To obtain this program, see URL *http://quantrm2.psy.ohio-state.edu/browne/.*

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

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences.* New York: Plenum.

Cattell, R. B., & Dickman, K. (1962). A dynamic model of physical influence demonstrating the necessity of oblique simple structure. *Psychological Bulletin, 59,* 389–400.

Cattell, R. B., & Jaspers, J. (1967). A general plasmode (No. 30-10-5-2) for factor analytic exercises and research. *Multivariate Behavioral Research Monographs,* No. 67-3.

Cattell, R. B., & Sullivan, W. (1962). The scientific nature of factors: A demonstration by cups of coffee. *Behavioral Science, 7,* 184–193.

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

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

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika, 35, 321–332.*

Crites, S. L., Jr., Fabrigar, L. R., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin, 20,* 619–634.

Cudeck, R. (1991). Noniterative factor analysis estimators with algorithms for subset and instrumental variable selection. *Journal of Educational Statistics, 16,* 35–52.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures and the "problem" of sample size: A clarification. *Psychological Bulletin, 109,* 512–519.

Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin, 115,* 475–487.

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

Dielman, T. E., Cattell, R. B., & Wagner, A. (1972). Evidence on the simple structure and factor invariance achieved by five rotational methods on four types of data. *Multivariate Behavioral Research, 7,* 223–242.

Eagly, A., & Chaiken, S. (1998). Attitude structure. In D. Gilbert, S. Fiske, & G. Lindsay (Eds.), *Handbook of social psychology* (Vol. 1, 4th ed., pp. 269–322). New York: McGraw-Hill.

Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology, 59,* 202–216.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research, 18,* 197–218.

Everitt, B. S. (1975). Multivariate analysis: The need for

data. and other problems. *British Journal of Psychiatry*, *126*, 237–240.

Fabrigar, L. R., & Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Personality and Social Psychology Bulletin*, *25*, 91–109.

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, *27*, 387–415.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality*, *31*, 439–485.

Ford, J K., MacCallum, R. C., & Tait, M. (1986). The applications of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*, 291–314.

Gorsuch, R. L. (1970). A comparison of biquartimin, maxplane, promax, and varimax. *Educational and Psychological Measurement*, *30*, 861–872.

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. (1980). Factor score reliabilities and domain validities. *Educational and Psychological Measurement*, *40*, 895–897.

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

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 231–258). New York: Plenum.

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

Gorsuch, R. L. (1997). New procedures for extension analysis in exploratory factor analysis. *Educational and Psychological Measurement*, *57*, 725–740.

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

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

Hakstian, A. R. (1971). A comparative evaluation of several prominent methods of oblique factor transformation. *Psychometrika*, *36*, 175–193.

Hakstian, A. R., & Boyd, W. M. (1972). An empirical investigation of some special cases of the general "orthomax" criterion for orthogonal factor transformation. *Educational and Psychological Measurements*, *32*, 3–22.

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, W. T., & Cattell, R. B. (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: The University of Chicago Press.

Harris, C. W. (1967). On factors and factor scores. *Psychometrika*, *32*, 363–379.

Harris, C. W., & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, *29*, 347–362.

Harris, M. L., & Harris, C. W. (1971). A factor analytic interpretation strategy. *Educational and Psychological Measurement*, *31*, 589–606.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, *17*, 65–70.

Hills, M. (1977). A review of "An introduction to multivariate techniques for social and behavioral sciences." *Applied Statistics*, *26*, 339–340.

Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor solutions. *Multivariate Behavioral Research*, *13*, 247–250.

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

Horn, J. L. (1969). On the internal consistency reliability of factors. *Multivariate Behavioral Research*, *4*, 115–125.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.

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

Hubbard, R., & Allen, S. J. (1987). A cautionary note on the use of principal components analysis: Supportive empirical evidence. *Sociological Methods and Research*, *16*, 301–308.

Humphreys, L. G., & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurements*, *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.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika*, *31*, 313–323.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200.

Kaiser, H. F. (1960). The application of electronic comput-

ers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151.

Kaiser, H. F. (1970). A second generation Little-Jiffy. *Psychometrika, 35,* 401–415.

Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology, 19,* 321–333.

Lautenschlager, G. J. (1989). A comparison of alternatives to conducting monte carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research, 24,* 365–395.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinborough, 60A,* 64–82.

Levonian, E., & Comrey, A. L. (1966). Factorial stability as a function of the number of orthogonally-rotated factors. *Behavioral Science, 11,* 400–404.

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika, 33,* 37–72.

Loehlin, J. C. (1990). Component analysis versus common factor analysis: A case of disputed authorship. *Multivariate Behavioral Research, 25,* 29–31.

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,* 59–69.

MacCallum, R. C. (1983). A comparison of factor analysis programs in SPSS, BMDP, and SAS. *Psychometrika, 48,* 223–231.

MacCallum, R. C. (1990). The need for alternative measures of fit in covariance structure modeling. *Multivariate Behavioral Research, 25,* 157–162.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–89.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

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

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

McNemar, Q. (1951). The factors in factoring behavior. *Psychometrika, 16,* 353–359.

Millar, M. G., & Millar, K. U. (1990). Attitude change as a function of attitude type and argument type. *Journal of Personality and Social Psychology, 59,* 217–228.

Millar, M. G., & Tesser, A. (1986). Effects of affective and cognitive focus on the attitude–behavior relation. *Journal of Personality and Social Psychology, 51,* 270–276.

Millar, M. G., & Tesser, A. (1989). The effects of affective–cognitive consistency and thought on the attitude–behavior relation. *Journal of Experimental Social Psychology, 25,* 189–202.

Montanelli, R. G., Jr., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika, 41,* 341–348.

Mulaik, S. A. (1972). *The foundation of factor analysis.* New York: McGraw-Hill.

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

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430–445.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology, 5,* 12–30.

Overall, J. E. (1964). Note on the scientific status of factors. *Psychological Bulletin, 61,* 270–276.

Petty, R. E., Priester, J. R., & Wegener, D. T. (1994). Cognitive processes in attitude change. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, 2nd ed., pp. 69–142). Hillsdale, NJ: Erlbaum.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1–14). New Haven, CT: Yale University Press.

Rozeboom, W. W. (1990). Whatever happened to broad perspective? *Multivariate Behavioral Research, 25,* 1–28.

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

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Analysis of*

latent variables in developmental research (pp. 399–419). Newbury Park, CA: Sage.

Schonemann, P. H. (1990). Facts, fictions, and common sense about factors and components. *Multivariate Behavioral Research, 25,* 47–51.

Skinner, H. A. (1980). Factor analysis and studies of alcohol. *Journal of Studies on Alcohol, 41,* 1091–1101.

Smith, M. B. (1947). The personal setting of public opinions: A study of attitudes toward Russia. *Public Opinion Quarterly, 11,* 507–523.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carle study. *Psychological Bulletin, 106,* 148–154.

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

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's: Some interesting parallels. *Psychometrika, 44,* 157–167.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH.* Evanston, IL: SYSTAT.

Steiger, J. H. (1990). Some additional thoughts on components, factors, and factor indeterminacy. *Multivariate Behavioral Research, 25,* 41–45.

Steiger, J. H., & Lind, J. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City.

Thurstone, L. L. (1947). *Multiple factor analysis.* 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.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Tucker, L. R., & MacCallum, R. C. (1997). *Exploratory factor analysis.* Unpublished manuscript, Ohio State University, Columbus.

van Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika, 43,* 225–243.

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

Velicer, W. F. (1977). An empirical comparison of the similarity of principal component, image, and factor patterns. *Multivariate Behavioral Research, 12,* 3–22.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3,* 231–251.

Velicer, W. F., & Jackson, D. N. (1990a). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25,* 1–28.

Velicer, W. F., & Jackson, D. N. (1990b). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research, 25,* 97–114.

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research, 17,* 371–388.

Wegener, D. T., & Fabrigar, L. R. (in press). Analysis and design for nonexperimental data: Addressing causal and noncausal hypotheses. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology.* New York: Cambridge University Press.

Wegener, D. T., & Gregg, A. P. (in press). Attitude structure. In A. Kazdin (Ed.), *Encyclopedia of Psychology.* New York: American Psychological Association.

Weiss, D. (1976). *Multivariate procedures.* In M. D. Dunnete (Ed.), Handbook of industrial/organizational psychology. Chicago, IL: Rand McNally.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Newbury Park, CA: Sage.

Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research, 25,* 89–95.

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

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354–365.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research, 17,* 253–269.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 17,* 253–269.