

A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research

JAMES M. CONWAY
Central Connecticut State University

ALLEN I. HUFFCUTT
Bradley University

The authors surveyed exploratory factor analysis (EFA) practices in three organizational journals from 1985 to 1999 to investigate purposes for conducting EFA and to update and extend Ford, MacCallum, and Tait's (1986) review. Ford et al. surveyed the same journals from 1975 to 1984, concluding that researchers often applied EFA poorly (e.g., relying too heavily on principal components analysis [PCA], eigenvalues greater than 1 to choose the number of factors, and orthogonal rotations). Fabrigar, Wegener, MacCallum, and Strahan (1999) reached a similar conclusion based on a much smaller sample of studies. This review of 371 studies shows reason for greater optimism. The tendency to use multiple number-of-factors criteria and oblique rotations has increased somewhat. Most important, the authors find that researchers tend to make better decisions when EFA plays a more consequential role in the research. They stress the importance of careful and thoughtful analysis, including decisions about whether and how EFA should be used.

Keywords: *exploratory factor analysis; factor extraction; number of factors; factor rotation*

Exploratory factor analysis (EFA) is an important tool for organizational researchers. It can be useful for refining measures, evaluating construct validity, and in some cases testing hypotheses. Not surprisingly, EFA has been found to be quite common in organizational research (Ford, MacCallum, & Tait, 1986). A casual reading of organizational journals shows that EFA is still in wide use. Unfortunately, Ford et al. conducted a comprehensive review of EFA practices and issued a gloomy assessment, concluding

Authors' Note: Thanks to Lisa Brzostek for her help in data preparation and to Filip Lievens for his comments on an earlier version of this article. Correspondence concerning this article should be addressed to James M. Conway, Department of Psychology, Central Connecticut State University, New Britain, CT 06050-4010; e-mail: conwayj@ccsu.edu; telephone: 860-832-3107; fax: 860-832-3123.

Organizational Research Methods, Vol. 6 No. 2, April 2003 147-168
DOI: 10.1177/1094428103251541
© 2003 Sage Publications

that organizational researchers have tended to make poor decisions about what factor extraction model to use (i.e., principal components rather than common factors), what criteria to use for deciding on the number of factors to retain (e.g., eigenvalues greater than 1), what type of rotation to use (i.e., orthogonal rather than oblique), and other critical aspects of the analysis. For example, they argued that oblique rotations are superior to orthogonal rotations but found that researchers overwhelmingly (about 80%) reported using orthogonal rotations. Ford et al.'s concern about EFA decisions is validated by evidence that these decisions can have an important impact on the quality of EFA results (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Gorsuch, 1997).

Ford et al. (1986) made recommendations for researchers regarding higher quality EFA decisions in hopes that practices would improve (e.g., greater use of a common factor model rather than principal components, multiple number-of-factors criteria, and oblique rather than orthogonal rotations). Their recommendations provide one reason for anticipating a change in EFA practices. A second reason is the advances in EFA technology, such as fit statistics for maximum likelihood analysis. A third reason is the increases in computing power since the 1980s. The approach of PCA, eigenvalues greater than 1, and varimax rotation (dubbed the "Little Jiffy") was proposed by Kaiser in 1956 when computing power was negligible. Common factors approaches and oblique rotations require greater computing resources but are easily handled by modern personal computers.

A more recent review by Fabrigar et al. (1999) is relevant to the question of whether practices have changed. Later, we review Fabrigar et al.'s results along with Ford et al.'s (1986). Fabrigar et al.'s overall conclusions were negative, but their results do suggest some changes. However, we view Fabrigar et al.'s results as merely suggestive because they did not focus squarely on organizational research (they reviewed one organizational journal along with one social psychology journal) and they covered only the 5-year period from 1991 to 1995.

We believe it is time for a comprehensive reevaluation of the use of EFA in organizational research. Our first goal in reevaluating EFA use, one that has not been addressed in previous research, was to compile information on the purposes for which EFA is being used (e.g., hypothesis testing, checking unidimensionality). We believe that it is difficult to evaluate usage and promote changes in EFA practices without knowing exactly how EFA is being used. Our second goal was to assess EFA practices in organizational research since Ford et al.'s (1986) review to see if practices have changed. Our third goal was to assess whether the quality of EFA decisions is related to the purpose of the EFA. It is possible that researchers using EFA for more important applications (e.g., hypothesis testing) may make higher quality decisions than researchers using EFA for less consequential applications such as verifying unidimensionality.

Purpose for Conducting EFA

EFA can be conducted for a variety of research purposes. One fundamental distinction is that of simple data reduction versus understanding latent constructs. In the former case, the research goal is simply to take a fairly large set of variables and reduce them to a smaller, more manageable number while retaining as much of the original variance as possible. In such uses, there is no attempt to interpret the resulting vari-

ables in terms of latent constructs—the use is more pragmatic than theoretical. However, it is likely that in most uses of EFA, organizational researchers do make interpretations regarding constructs rather than purely reducing data. When EFA is used with the intention of understanding latent constructs, we can distinguish purposes from each other based on how consequential they are in the scheme of the research. We will group purposes as either less consequential or more consequential in the scheme of a research project.

Less consequential purposes involve use of EFA for a preliminary evaluation of variables. In other words, the EFA serves a subsidiary role, merely helping in preparation for the hypothesis testing that is the central purpose of the study. One such use of EFA is for a well-established multi-item instrument, simply to verify the scale's unidimensionality. For example, Chan, Drasgow, and Sawin (1999) intended to conduct item response theory analyses on each of the subtests of the Armed Services Vocational Aptitude Battery, but they first applied EFA to each subtest to confirm that the subtest was unidimensional. Another preliminary-evaluation purpose involves the use of EFA with new or ad hoc instruments to find out what the dimensionality is (so composite scores can be created for use in hypothesis testing). For instance, Rousseau and Tijoriwala (1999) intended to study factors (e.g., trust in management) related to workers' beliefs about organizational change. As a preliminary step, they developed a questionnaire to measure perceptions of organizational change and subjected the questionnaire responses to EFA. They developed scales based on the EFA and used the scales in subsequent hypothesis tests. (Note that EFA can also be applied to existing instruments to assess dimensionality; we will argue later that researchers should think carefully about use of EFA with existing instruments and that sometimes confirmatory factor analysis [CFA] may be preferable.) Another example of preliminary evaluation is applying EFA to a set of self-report instruments, using the pervasiveness of the first factor to evaluate the possibility of same-method bias before testing hypotheses (e.g., C. Tinsley, 1998).

On the other hand, EFA sometimes plays a very consequential role in a study. This is the case when EFA is not just for preliminary evaluation of variables but is critically involved in accomplishing the major goals of the study. One example is when a study is conducted on the development and validation of an instrument, and EFA serves a function such as developing and refining the instrument's scales. McCauley, Ruderman, Ohlott, and Morrow (1994) developed an instrument for assessing developmental components of managerial jobs. They used EFA for creating and refining the instrument's scales. Other examples of very consequential EFA uses involve hypothesis testing. EFA might be used to develop a hypothesized model that is then tested with new data using CFA. Conway (2000) used EFA for supervisor ratings of managers' performance and then applied the supervisor results to peers and subordinates using CFA. Here, the EFA is used for more than preliminary evaluation—it is integrally involved in the development and testing of hypotheses. Another hypothesis-testing example is when EFA is conducted under different conditions to see if the number of factors changes. For example, Ellingson, Sackett, and Hough (1999) used EFA to see whether personality factor structures differed when respondents were instructed to respond honestly versus when they were instructed to "fake good."

Our second goal was to update Ford et al.'s (1986) review of EFA practices, to evaluate the quality of researchers' decision making. Our third goal was to see if quality of decision making is related to purpose of the EFA. We believed we would find relatively

higher quality decisions when the EFA directly addressed a major goal of the study than when EFA did not directly address a major goal. We expected this because researchers using EFA in a major role are probably more likely to take the time to give careful consideration to the decisions that must be made. This greater care should lead researchers to be better informed about the options available and the consequences of different choices. If our hypothesis is correct, then this would suggest an important qualification to Ford et al.'s discouraging conclusions: The picture would seem much more encouraging because the most important and consequential uses of EFA would be characterized by relatively good decisions. Achieving our second and third goals (evaluating the quality of decision making and seeing if quality is related to purpose of the EFA) depended on being able to specify what the higher quality decisions are, the issue to which we now turn.

What Are High-Quality EFA Decisions?

There are many decisions confronting users of EFA. We will focus primarily on three EFA decisions we believe to be both important for the outcome and easily under the researcher's control: (a) the factor extraction model used, (b) the number of factors retained, and (c) the method used to rotate factors (assuming more than one factor is retained). In addition to these three decisions, we will discuss reporting of information about the EFA. For each of the decisions, we will first discuss options available and recommendations for high-quality EFA practices. Our brief discussion will rely heavily on Ford et al. (1986) and on Fabrigar et al. (1999) but will also draw from other sources such as Floyd and Widaman (1995) and Gorsuch (1997); readers may consult these sources for more detail. Second, we will describe Ford et al.'s and Fabrigar et al.'s findings regarding what decisions organizational researchers have actually tended to make.

Selection of the factor extraction model. Although a variety of factor extraction models are available, most can be categorized as either a common factor model or a components model (Gorsuch, 1983). Of components models, by far the most popular is PCA, so hereafter we will refer only to PCA. Among common factor models, maximum likelihood and principal axis factoring with estimated communalities are popular. The main difference between common factor and PCA models is in their purposes. The purpose of common factor models is to understand the latent (unobserved) variables that account for relationships among measured variables; the goal of PCA is simply to reduce the number of variables by creating linear combinations that retain as much of the original measures' variance as possible (without interpretation in terms of constructs).

Due to their different purposes, common factor and PCA models differ in how they conceptualize sources of variance in measured variables. Common factor models assume the factors are imperfectly reflected by the measured variables and differentiate between variance in measures due to the common factors (factors that influence more than one measure) and variance due to unique factors (factors that influence only one measure). PCA makes no such distinction, and the components therefore contain a mixture of common and unique variance.

If a researcher's purpose is to understand the latent structure of a set of variables (which will usually be the case), then use of a common factor model such as principal

axis or maximum likelihood factoring represents a high-quality decision. If a researcher's purpose is pure reduction of variables without interpreting the resulting variables in terms of latent constructs, then use of PCA represents a high-quality decision. Given that most researchers do attach meaning beyond the observed variables, the common factor model will generally be the better choice.

Even if the research goal involves latent constructs, it has been claimed that PCA is a very good substitute for common factor analysis and gives almost identical results (e.g., Goldberg & Digman, 1994; Velicer & Jackson, 1990). But although PCA results may very closely resemble common factor results in some cases, there are other cases in which noticeable differences emerge. For example, Widaman (1993) showed that when the data fit the assumptions of the common factor model, PCA loadings tend to be too high whereas common factor loadings are very accurate. Gorsuch (1997) pointed out that when conducting item analysis, such inflation can make items look better than they really are. Fabrigar et al. (1999) analyzed a number of data sets and found that although maximum likelihood and PCA solutions were often similar, there were cases in which interpretations were different (at least one variable's largest loading was on different factors for PCA versus maximum likelihood analyses). And Gorsuch (1990) provided a particularly compelling example in which a researcher found quite high and apparently significant PCA loadings for a correlation matrix in which none of the correlation coefficients were statistically significant. This situation is very counterintuitive, but common factor analysis of the same data showed no high loadings for any variables (an intuitive result).

It clearly can make a meaningful difference whether a researcher uses common factor analysis or PCA, and both theory and empirical evidence favor common factor analysis as the more appropriate. Gorsuch (1990) pointed out that if common factor analysis produces more sensible results some of the time and basically equivalent results the rest of the time, there is little reason to use PCA. The evidence therefore shows common factor analysis to be the better decision (assuming a focus on latent constructs).

But what decisions have organizational researchers actually made? Ford et al. (1986) surveyed EFAs in articles published in the *Journal of Applied Psychology (JAP)*, *Personnel Psychology (PP)*, and *Organizational Behavior and Human Performance* from 1975 through 1984. They found that PCA predominated over common factor analysis, with PCA being reported 42.1% of the time and common factors only 34.2% of the time; 23.7% of the time, the factor extraction model was not stated. Ford et al. therefore concluded that researchers have frequently made questionable decisions. More recently, Fabrigar et al. (1999) found that for *JAP* only, from 1991 to 1995, PCA was even more common at 48.3% versus 22.4% for common factors; 25.9% of the time, the factor extraction model was unknown. These results suggest a trend toward increasing use of PCA over common factor analysis.

Number of factors. A second important decision is the criterion for the number of factors (or components) to retain. Available options include Kaiser's (1956) "eigenvalues greater than one" rule, the scree test, parallel analysis, a priori theory, and retaining the number of factors that gives a high proportion of variance accounted for or that gives the most interpretable solution. If PCA is used, then the minimum average partial method can be used (Velicer, 1976). If maximum likelihood analysis is used, then fit indices can be used as described by Fabrigar et al. (1999) and Browne and

Cudeck (1992). See Fabrigar et al. and Zwick and Velicer (1986) for more comprehensive reviews of options.

Research and experience show that the choice of number-of-factors criteria is an extremely important one—studies clearly show that different techniques often lead to different numbers of factors being retained (Fabrigar et al., 1999; Zwick & Velicer, 1986). Furthermore, ample research shows that the commonly used eigenvalues-greater-than-1 rule does not consistently give an accurate number of factors (it tends to produce too many factors) (Gorsuch, 1997), so it probably should not be relied on (e.g., Hakstian, Rogers, & Cattell, 1982; Tucker, Koopman, & Linn, 1969; Zwick & Velicer, 1986).

If different criteria lead to different numbers of factors, which should researchers use? Zwick and Velicer (1986) concluded that for PCA, parallel analysis was generally most accurate, followed by the minimum average partial procedure (though a limitation of minimum average partial is that its use has not been extended to common factor analysis). However, the recommendation by both Ford et al. (1986) and Fabrigar et al. (1999) was to use a combination of techniques. This is sensible because no single technique has been shown to be highly accurate over a wide range of conditions in pinpointing the number of factors. Furthermore, Fabrigar et al. noted that choice of number of factors is a substantive issue as well as a statistical one because an uninterpretable solution will not be helpful. Therefore, the combination of techniques should probably include examination of multiple solutions with different numbers of factors, for interpretability.

Regarding what decisions researchers actually make, Ford et al. (1986) found that the most common technique reported was retaining factors with eigenvalues greater than 1, at 21.7%. Only 13.8% reported using multiple techniques. This is an important finding because the eigenvalues-greater-than-1 technique has generally been found to work poorly relative to other techniques (Fabrigar et al., 1999). Fabrigar et al. found slightly more promising results, with 19.0% reporting the eigenvalues-greater-than-1 rule and 20.7% reporting use of multiple techniques. One final note is that Ford et al. found that fully 30.9% of researchers did not report how they decided on the number of factors, and the corresponding figure was 37.9% for Fabrigar et al. As with the selection of a factor extraction model, Ford et al. concluded that researchers have often made questionable decisions, although Fabrigar et al. showed some reason for optimism that using multiple techniques has become more common.

Rotation. Given a number of factors greater than one, the factors are usually rotated to find a more interpretable solution. An important criterion for interpretability is what Thurstone (1947) referred to as “simple structure.” Thurstone’s criteria are somewhat complex, but Fabrigar et al. (1999) gave a relatively straightforward description: Simple structure means that each factor has a subset of variables with high loadings, and the rest with low loadings, and that each variable has high loadings on only some of the factors and low loadings on the rest.

Two basic types of analytical rotations can be used to reach a more interpretable solution: orthogonal rotations, forcing uncorrelated factors, and oblique rotations, allowing correlated factors. By far, the most popular orthogonal rotation is varimax, which attempts to maximize the variance of squared loadings on a factor (i.e., to produce some high loadings and some low loadings for each factor) (Kim & Mueller,

1978). A number of oblique rotations are used, such as direct oblimin or Promax. Direct oblimin finds the oblique solution balancing the criteria that (a) each variable be relatively unifactorial (ideally one high loading and other loadings near zero) and (b) the covariance between elements on factors be minimized (Kim & Mueller, 1978). Promax, on the other hand, begins with an oblique rotation and uses it to compute a target matrix. The final solution is the oblique solution that most closely matches the target matrix.

Regarding a high-quality rotation decision, we agree with Ford et al. (1986), Fabrigar et al. (1999), and Gorsuch (1997) that an oblique rotation is preferred. If factors really are correlated (a likely situation), then orthogonal rotation forces an unrealistic solution that will probably distort loadings away from simple structure, whereas an oblique rotation will better represent reality and produce better simple structure. If factors really are uncorrelated or show a very low correlation, then an orthogonal rotation will be appropriate, but so will an oblique rotation, which will give a factor correlation of about zero and loadings that are very similar to those from an orthogonal rotation (Floyd & Widaman, 1995).

Consistent with the logical argument in the preceding paragraph, evidence supports the use of oblique rotations (Fabrigar et al., 1999; Gorsuch, 1970, 1997). For example, Gorsuch (1997) showed that varimax rotation is biased against finding a general factor when one really exists. And Fabrigar et al. (1999) showed that an oblique rotation, direct oblimin, produced considerably fewer "cross loadings" than did varimax rotation for the same data. That is, the oblique rotation resulted in superior simple structure. Although an orthogonal rotation may seem conceptually simpler due to the lack of factor correlations, in fact, it is the oblique rotation that is most likely to give a simple, interpretable solution. We therefore have a situation similar to that with factor extraction models. If oblique rotations sometimes produce better solutions (i.e., when the constructs are in reality correlated) and the rest of the time produce essentially equivalent solutions (i.e., when constructs are really uncorrelated or nearly so), then there seems no reason to use an orthogonal rotation.

Regarding actual decisions in organizational research, again Ford et al. (1986) reached a pessimistic conclusion. They found that almost 80% of EFAs involved orthogonal rotations, whereas 12.1% either used oblique rotations or did not rotate the solution (about 8% did not state the rotation). Fabrigar et al.'s (1999) results were somewhat more encouraging—they found that 48.3% used varimax (the only orthogonal rotation they listed) and 20.6% used oblique rotations.

Reporting EFA information. As Ford et al. (1986) noted (along with others, e.g., Floyd & Widaman, 1995; H. Tinsley & Tinsley, 1987), it is important for readers to be able to evaluate researchers' EFA practices and results. This requires that researchers report important decisions regarding their analysis, including factor extraction model, number of factors, rotation, how factors were interpreted, and how factor scores were computed. Researchers should also report results including descriptive statistics and the correlation matrix (whenever possible), eigenvalues, communalities, percentage of variance accounted for, the full factor loading matrix, and interfactor correlations if an oblique solution is used (Floyd & Widaman, 1995; Ford et al., 1986; H. Tinsley & Tinsley, 1987). Both Fabrigar et al. (1999) and Ford et al. found that substantial percentages of researchers omitted important information about their EFAs. For example,

both reviews found that more than 30% of the time, the decision criterion for number of factors was not specified. For certain results, the percentage not reporting was much higher (e.g., in Ford et al., communalities were reported only 16.4% of the time). Fabrigar et al.'s results even suggest an increase in nonreporting for rotation (20.7% vs. Ford et al.'s 8.3%).

Sample size, sample-to-variable ratio, and variable-to-factor ratio. MacCallum, Widaman, Zhang, and Hong (1999) discussed sample size in EFA, concluding that adequate sample size is a relatively complex issue (not well addressed by general rules about sample-to-variable ratios) and that often samples need to be quite large (e.g., 400 or greater) to produce undistorted results. Fabrigar et al. (1999) discussed variable-to-factor ratios, suggesting there should be at least four variables per factor.

Ford et al. (1986) did not address sample size per se or variable-to-factor ratios. They did report on sample-to-variable ratios using two categories: less than 5:1 and greater than 5:1. They found that 70% of studies had ratios greater than 5:1, 27% had ratios less than 5:1, and 3% did not report. Fabrigar et al. (1999) reported that 43.1% of studies had sample sizes exceeding 400 and that a large majority, 75.7%, had variable-to-factor ratios of at least 4:1.

The Present Review

There is evidence that which decisions are made can have an important impact on the interpretation of EFA results. There is also evidence that organizational researchers have tended to make questionable decisions. However, there has been no review of the purposes for which EFA is used. And the last comprehensive review of organizational researchers' decisions was published in 1986 (although Fabrigar et al., 1999, suggested the possibility of some changes). It is therefore important to evaluate whether the situation has improved since that time. Even if EFA decisions are relatively similar to Ford et al.'s (1986) and Fabrigar et al.'s (1999) findings, the pessimistic conclusions might be qualified if it turns out that researchers make relatively high-quality EFA decisions, and report more of the relevant information, in more consequential uses of EFA.

In our review, we examined EFAs published in *JAP*, *PP*, and *Organizational Behavior and Human Decision Processes (OBHDP)*, the successor to *Organizational Behavior and Human Performance* for the period from 1985 to 1999. Our review therefore begins the year after the period covered by Ford et al. (1986), and we used the same three journals in an attempt to produce comparable results. For each EFA, we categorized the purpose for which it was conducted. To evaluate whether EFA practices have changed since Ford et al.'s review, we examined the frequency of various decisions regarding selection of the factor extraction model, number of factors to retain, and rotation as well as other study characteristics (e.g., ratio of variables to factors, sample size). We also examined the frequency with which researchers reported their decisions and important information such as eigenvalues. To test our hypothesis regarding the purpose of the EFA, we investigated whether decisions differed according to more consequential versus less consequential purposes.

Method

Literature Review

We examined each article from *JAP*, *PP*, and *OBHDP* for the years 1985 to 1999, inclusive, to locate studies using EFA. The only exception was that we did not examine articles that clearly dealt with nonorganizational topics (e.g., eyewitness testimony). Like Ford et al. (1986), we did not include any article that referred to a previously conducted EFA. Unlike Ford et al., we did not eliminate articles in which EFA played only a very minor part in the overall analysis. This is because one focus of our review was to examine purposes for which EFA has been used, and we did not wish to restrict the range of purposes. However, we conducted additional analyses without studies in which EFA played a very minor part, to compare our results with Ford et al.'s. These analyses showed overall frequencies very similar to those of the total data set, so we report all analyses for the total data set.

We found 371 articles that reported using EFA. Of the 371, 226 (61%) were published in *JAP*, 92 (25%) were published in *PP*, and 53 (14%) were published in *OBHDP*. The percentage for *JAP* is somewhat higher than the 55% found by Ford et al. (1986), and the percentage for *OBHDP* is somewhat lower than Ford et al.'s 19% (the percentage for *PP* is roughly the same as Ford et al.'s 26%).

Coding

Each study was coded on a number of variables, including purpose for the EFA, several EFA decisions (factor extraction model used, decision criteria for number of factors, and type of rotation), reporting of information (correlation matrix, eigenvalues, percentage of variance accounted for, communalities of variables, and factor loading matrix), and several study design variables (sample size, number of variables, and number of factors).

Purpose for EFA. There were 12 categories of purpose for the EFA, but these purposes were subgrouped as either playing a relatively minor role in the study or addressing an important goal of the study. In the less consequential or relatively minor role category were reducing the number of observed variables with no attempt to interpret factors/components as latent constructs, assessing unidimensionality of existing measures, assessing unidimensionality of new or ad hoc measures, preliminary evaluation of existing measure for use in hypothesis testing (not directly addressing a goal of the study), preliminary evaluation of new or ad hoc measure for use in hypothesis testing (not directly addressing a goal of the study), post hoc exploration of correlations, deriving a measure of general intelligence, assessing monomethod bias, and other. In the more consequential or addressing an important goal category were development of a new measure or scale (important role in addressing goals of the study), hypothesis testing, and addressing a goal of the study in some other way.

Factor extraction model. Factor extraction model was coded as either PCA, principal axis factoring (this category included both iterated and noniterated factoring as well as studies simply stating that principal factoring was done), maximum likelihood,

multiple models, or no information (e.g., the study only stated that factor analysis was conducted). Overall frequency results are presented using these categories. But to simplify comparisons across different EFA purposes, we collapsed principal axis and maximum likelihood categories into a common factor analysis category. Studies using multiple models were categorized as common factor analysis if one of the multiple models was a common factor model. This was done because the concern expressed in the introduction was that results may differ for PCA versus common factor analysis, but if both are conducted, then this issue is not of concern—using both models allows the researcher to evaluate whether the PCA results differ appreciably. In summary, we collapsed factor models into three categories: common factors versus PCA versus no information.

Number-of-factors decision criteria. Number-of-factors decision criteria included eigenvalues greater than 1, large eigenvalues (without specifying a cutoff), scree test, examining multiple solutions/interpretability of the solution (including simple structure), a priori number of factors, percentage of variance accounted for, parsimony, parallel analysis, chi-square test (for maximum likelihood factoring), other (e.g., maximum likelihood fit statistics), or no information. If a study used more than one of these criteria, it was classified as using multiple criteria. To simplify comparisons across different EFA purposes, we collapsed techniques into three categories: single criterion versus multiple criteria versus no information.

Rotation. Type of rotation was coded as either varimax, Harris-Kaiser orthoblique, direct oblimin, promax, orthogonal (without specifying which orthogonal rotation), oblique (without specifying which oblique rotation), both orthogonal and oblique rotations used, not rotated (e.g., only one factor), or no information. To simplify comparisons across different EFA purposes, we excluded unrotated solutions and used just three categories: orthogonal rotation versus oblique rotation versus no information. Studies using both orthogonal and oblique rotations were put into the oblique category. This was because the concern expressed in the introduction was that results may differ for orthogonal versus oblique rotations, but if both are conducted, then this is not an issue.

Reporting of information. We simply recorded whether a given piece of information (e.g., eigenvalues) was reported.

Study design variables. We recorded the sample size for the EFA, the number of variables, and the number of factors extracted. We then computed the ratio of sample size to number of variables.

Coding process. The first author coded all 371 studies. Next, to check on the reliability of coding, the second author coded a random sample of 25 studies. Coding agreement was high for most variables. For numerical codes (sample size, number of variables, and number of factors), intercoder correlations were all .98 or higher. However, for number of factors, there were several studies for which one coder recorded “no information” and the other coder recorded a specific number of factors. We were therefore not satisfied with agreement on that variable.

For categorical codes, most variables had near perfect agreement. Exceptions were decision criterion for determining the number of factors (80% agreement), type of

rotation (84%), and purpose for the EFA (56% agreement when considering all coding categories listed in Table 1, although when codes were collapsed into the two major categories, addressing a goal of the study or not, agreement was 84%).

Due to our concern with agreement on several variables, we reviewed and discussed coding instructions and then coded another 10 randomly chosen studies (the first author recoded those studies on the variables in question) on number of factors, decision criterion for number of factors, type of rotation, and purpose for the EFA. Agreement on these 10 studies was perfect for number of factors and rotation, there was one disagreement for purpose, and there were three disagreements for number-of-factors decision criterion. Coding on these 10 studies showed few disagreements, and we were confident that any systematic differences between coders had been worked out. The first author therefore recoded all remaining studies on the four variables. Any discrepancies between the initial codes and recodes were scrutinized to determine the appropriate code.

Results

First, we present frequencies for each EFA characteristic coded. The frequencies address our first two goals, including (a) examining the purposes for which EFA has been used in published organizational research and (b) assessing EFA decision making in organizational research since Ford et al.'s (1986) review. Next, we present data on the relationship between purpose and EFA decisions, addressing our third goal of seeing whether EFA decisions tend to be of higher quality in studies where EFA serves a more important purpose.

Descriptive Statistics

Table 1 presents frequencies and percentage frequencies for the following variables: purpose for the EFA, factor extraction model, number-of-factors criteria, rotation, reporting of information (e.g., correlation matrix), sample size, ratio of sample size to number of variables, and ratio of variables to factors. For comparison, Ford et al.'s (1986) and Fabrigar et al.'s (1999) percentage frequencies are also shown wherever possible. For some variables, Ford et al. and Fabrigar et al. did not use exactly the same categories, so there are not corresponding percentage frequencies for all of our categories.

Purpose of the EFA. Table 1 shows that the most common purpose was preliminary evaluation of a new or ad hoc measure, at 46.1%. Next most common was preliminary evaluation of existing measures, at 21.3%, and another 4.0% assessed unidimensionality of existing measures. Also of interest are the two superordinate categories: purposes not addressing an important goal of the study versus purposes addressing an important goal. More frequent were the less important purposes (not addressing a goal) at 78.2%, with 21.8% addressing an important goal.

Factor extraction model. Fabrigar et al.'s (1999) results suggested an increase in the use of PCA and a decrease in use of common factor models. However, Table 1 shows that our results are very similar to those of Ford et al. (1986). Specifically, PCA was the most popular factor extraction model at about 40% (Ford et al.'s value was 42%). Com-

Table 1
Frequencies for Exploratory Factor Analysis Variables

	n	%f	Ford, MacCallum, and Tait (1986) %f	Fabrigar, Wegener, MacCallum, and Strahan (1999) %f ^a
Purpose of exploratory factor analysis				
Reduce number of observed variables	1	0.3		
Assess unidimensionality of existing measures	15	4.0		
Assess unidimensionality of new or ad hoc measures	12	3.2		
Preliminary evaluation of existing measure	79	21.3		
Preliminary evaluation of new or ad hoc measure	171	46.1		
Post hoc exploration of correlations	1	0.3		
Derive a measure of "g"	6	1.6		
Assess monomethod bias	3	0.8		
Development of new measure/scale (addressing goals of the study)	27	7.3		
Hypothesis testing	16	4.3		
Address a goal of the study in some other way	38	10.2		
Other	2	0.5		
Total of purposes not addressing a goal of the study	290	78.2		
Total of purposes addressing a goal of the study	81	21.8		
Factor extraction model				
Principal components	147	39.6	42.1	48.3
Principal axis	83	22.4		
Maximum likelihood	14	3.8		
Common factor (unspecified)	9	2.4		
Multiple	14	3.8		3.4
No information	104	28.0	23.7	25.9
Total common factor (including principal axis, maximum likelihood, unspecified, and use of multiple models)	120	32.3	34.2	22.4
Number-of-factors criteria				
Eigenvalue greater than 1 (Kaiser criterion)	57	15.4	21.7	19.0
Large eigenvalues (without specifying a cutoff)	10	2.7		
Scree test	22	5.9	11.2 ^b	15.5
Interpretability/examining multiple solutions/simple structure	10	2.7		0.0
A priori number of factors retained	34	9.2	11.2	6.9
Percentage variance accounted for	7	1.9		
Parsimony	0	0.0		
Principal components	147	39.6	42.1	48.3
Parallel analysis	4	1.1		0.0
Chi-square test	2	0.5		0.0
Other	5	1.3		0.0
No information	140	37.7	30.9	37.9
Multiple criteria	80	21.6	13.8	20.7
Rotation				
Varimax	140	37.7		48.3
Harris-Kaiser orthoblique	6	1.6		1.7

Table 1 (continued)

	n	%f	Ford, MacCallum, and Tait (1986) %f	Fabrigar, Wegener, MacCallum, and Strahan (1999) %f ^a
Direct oblimin	14	3.8		
Promax	8	2.2		3.4
Other orthogonal (or unspecified)	13	3.5		
Other oblique (or unspecified)	33	8.9		
Both orthogonal and oblique	6	1.6		
Not rotated (e.g., only one factor)	85	22.9		6.9
No information	66	17.8	8.3	20.7
Total orthogonal	153	41.2	79.6	48.3
Total oblique	67	18.1	12.1 ^c	20.6
Reporting of results				
Correlation matrix	17	4.6	5.3	
Eigenvalues	66	17.8	27.6	
Communalities	14	3.8	16.4	
Percentage of variance accounted for	163	43.9	71.7	
Full factor loading matrix	92	24.8	45.4	
Sample size				
1-100	74	19.9		13.8
101-200	97	26.1		24.1
201-300	51	13.7		15.5
301-400	33	8.9		3.4
401-500	21	5.7		
More than 500	86	23.2		
Unknown	9	2.4		
Ratio of sample size to number of variables				
5:1 or less	65	17.5		
6:1 to 10:1	80	21.6		
11:1 to 15:1	31	8.4		
16:1 to 20:1	31	8.4		
Greater than 20:1	135	36.4		
Unknown	29	7.8		
Ratio of number of variables to number of factors				
Less than 3:1	14	3.8		1.7
3:1	41	11.1		15.5
4:1	42	11.3		17.2
5:1	53	14.3		17.2
6:1	36	9.7		10.3
More than 6:1	150	40.4		31.0
Unknown	35	9.4		6.9

a. Fabrigar et al.'s (1999) results are presented only for the *Journal of Applied Psychology*.

b. Ford et al.'s (1986) value of 11.2% was actually a combination of scree test and other.

c. Ford et al.'s (1986) value of 12.1% was actually a combination of oblique rotation and no rotation at all.

mon factor models were used 32% of the time (34% from Ford et al.), and in 28% of the cases, the factor model was not reported.

Number-of-factors decision criteria. Fabrigar et al.'s (1999) results suggested essentially no change in use of the eigenvalues-greater-than-1 criterion but an increase in use of multiple methods. Our results show a decrease in the eigenvalues-greater-than-1 rule at 15.4% (compared to Ford et al.'s [1986] 21.7%) but confirm the increase in multiple methods to 21.6% from Ford et al.'s 13.8%.

Our findings differ a bit from Ford et al.'s (1986) regarding number-of-factors criteria (see Table 1). Unfortunately, nonreporting was more common in our study at 37.7% compared to 30.9% for Ford et al. (Fabrigar et al., 1999, showed a value of 37.9%).

Rotation. Fabrigar et al.'s (1999) rotation results showed a marked decrease in orthogonal rotations to 48.3% from 79.6% for Ford et al. (1986). Our rotation results show even more of a decrease at 41.2% (although our no information category was more common at 18% versus 8% for Ford et al.). And Fabrigar et al. found oblique rotations to be more common than did Ford et al. at 20.6%; our results show a similar value at 18%. Ford et al. reported a value of 12.1%, but this was a combination of oblique rotation and no rotation, so their actual percentage for oblique rotations was even lower.

Reporting of information. The pieces of information listed in Table 1 (e.g., correlation matrix) were not frequently reported; only percentage of variance accounted for approached 50%. Furthermore, all of our percentages were lower than those reported by Ford et al. (1986).

Sample size, sample-to-variable ratio, and variable-to-factor ratio. Table 1 shows that almost half the studies had samples of modest size (less than 200), and 29% used large samples (more than 400) (Fabrigar et al., 1999). Almost 40% of the time, the sample-to-variable ratio was 10:1 or less, which could be considered relatively low (although MacCallum et al., 1999, questioned the usefulness of general guidelines regarding sample-to-variable ratios). Regarding variable-to-factor ratios, Fabrigar et al. (1999) recommended a ratio of at least 4:1. Of the studies reviewed here, 15% fell below that standard, suggesting that although the large majority had adequate variables for the number of factors, there were some that did not.

Relationships of EFA Decisions With EFA Purpose

To test our hypothesis that we would find relatively higher quality decisions when the EFA directly addressed a major goal of the study than when EFA did not directly address a major goal, we constructed cross-tabulation tables (see Table 2) and conducted chi-square tests. Remember that we defined higher quality decisions as (a) using a common factor model instead of PCA (except when PCA is used purely for variable reduction; the one study in this category was excluded from this analysis), (b) using multiple number-of-factors criteria, and (c) using an oblique rotation. To test for relationships, we collapsed each variable into two or three categories. Factor extraction model was collapsed to common factor versus PCA versus no information, number-of-factors criteria was collapsed to multiple criteria versus a single criterion versus no information, rotation was collapsed into orthogonal versus oblique versus no information, and purpose was collapsed into less consequential (not directly addressing a goal of the study) versus more consequential (directly addressing a goal of the study).

Table 2
 Cross-Tabulations for Exploratory Factor Analysis Decisions
 and Exploratory Factor Analysis Purpose

<i>Purpose</i>	<i>Factor Extraction Model</i>		
	<i>Common Factor</i>	<i>Principal Components Analysis</i>	<i>No Information</i>
Not directly addressing a goal (<i>n</i> = 289)			
<i>n</i>	72	124	93
Row %	24.9	42.9	32.2
Directly addressing a goal (<i>n</i> = 81)			
<i>n</i>	48	22	11
Row %	59.3	27.2	13.6
<i>Purpose</i>	<i>Number-of-Factors Criterion</i>		
	<i>Multiple Criteria</i>	<i>Single Criterion</i>	<i>No Information</i>
Not directly addressing a goal (<i>n</i> = 290)			
<i>n</i>	48	120	122
Row %	16.6	41.4	42.1
Directly addressing a goal (<i>n</i> = 81)			
<i>n</i>	32	31	18
Row %	39.5	38.3	22.2
<i>Purpose</i>	<i>Rotation (excluding unrotated solutions)</i>		
	<i>Oblique</i>	<i>Orthogonal</i>	<i>No Information</i>
Not directly addressing a goal (<i>n</i> = 211)			
<i>n</i>	39	114	58
Row %	18.5	54.0	27.5
Directly addressing a goal (<i>n</i> = 75)			
<i>n</i>	29	38	8
Row %	38.7	50.7	10.7

For each of the three cross-tabulations shown in Table 2, chi-square statistics (each with 2 *df*) showed significant ($p < .05$) relationships between EFA purpose and EFA decisions. In all three cases, as expected, higher quality decisions were more likely in more consequential uses of EFA. In addition, we conducted chi-square tests (Fisher's exact tests) to determine whether each piece of information listed in Table 1 was reported more frequently in studies where the EFA was more consequential. Finally, we tested whether sample size, sample-to-variable ratio, and variable-to-factor ratio differed by purpose.

Factor extraction model. Table 2 shows that the common factor model was much more common in studies using EFA to address an important goal (59% vs. 25%), whereas PCA and nonreporting were much less common.

Number-of-factors criteria. As shown in Table 2, use of multiple criteria was more than twice as common in studies using EFA to address an important goal (40% vs. 17%), whereas nonreporting was much less common.

Rotation. For the 286 studies in which EFA solutions were rotated (excluding single-factor studies and a handful of multiple-factor studies that did not use rotations), oblique rotations were more than twice as common when EFA was used to address an important goal (39% vs. 19%), and nonreporting was much less common.

Reporting of information. Except for communalities, tests indicated that all pieces of information were presented more frequently ($p < .05$) in more consequential uses of EFA. However, information was infrequently reported even in consequential uses. Percentages for more versus less consequential uses, respectively, were 10% versus 3% for the correlation matrix, 7% versus 3% for communalities (*ns*), 33% versus 13% for eigenvalues, 54% versus 41% for percentage of variance accounted for, and 42% versus 20% for the full loading matrix.

Sample size, sample-to-variable ratio, and variable-to-factor ratio. There were no significant chi-squares for sample size, variables, or factors.

Discussion

Ford et al. (1986) offered sobering conclusions about the high frequency of poor use of EFA in organizational research. Our review was intended to investigate purposes for which researchers use EFA, to update our knowledge of EFA practices, and to see if these practices differed depending on the purpose of the EFA. Our results are more encouraging than those of Ford et al., suggesting that high-quality decisions have become somewhat more common (at least regarding the number of factors and choice of rotation). Furthermore, we found that high-quality EFA decisions are considerably more common in more consequential uses of EFA.

Purpose of EFA

One finding regarding purpose of the EFA was the higher frequency of use for relatively minor purposes (78.2% not directly addressing an important goal of the study vs. 21.8% addressing an important goal). We suspect that our results underestimate the use of EFA to address important goals such as instrument development. These uses are commonly described in instruments' technical manuals or publications for users (e.g., Leslie & Fleenor, 1998, describe the development of 24 managerial performance feedback instruments, many of which involved EFA) but are not as often seen as fit to publish in journal articles. Our sampling method may therefore not adequately represent such uses.

Another finding concerns the use of EFA for new or ad hoc versus existing measures, which we believe can have implications for the appropriateness of EFA in the first place. The most frequent purpose of EFA was for preliminary evaluation of a new or ad hoc measure (46.1%). A substantial percentage of studies (about one quarter) used EFA with an existing measure. Using EFA for new or ad hoc measures seems appropriate, given that these measures may not provide clear expectations about the factor pattern. However, we suggest that use of EFA with existing measures may not

always be appropriate. Existing instruments will at least sometimes provide a clear hypothesis about the factor pattern. In these situations, it may be better to consider using CFA, or a combination of EFA and CFA, as we now explain.

EFA, because it is more likely than CFA to capitalize on chance factors in the data (Fabrigar et al., 1999) may produce an extra, nonreplicable factor or show unexpected chance loadings for a variable. In support of this idea, repeated EFAs of the same instrument have shown considerably different results. Examples include (a) differing numbers of factors found for the Multifactor Leadership Questionnaire by Bass (1985), Hater and Bass (1988), and Den Hartog, Van Muijen, and Koopman (1997) and (b) differing numbers of factors and (for solutions with the same number of factors) different variables loading on the factors for Rosenberg's self-esteem scale by Hensley and Roberts (1976), Hensley (1977), and Dobson, Goudy, Keith, and Powers (1979).

These examples demonstrate that EFA results for an instrument can vary. The variation between studies may be due to real differences, to sampling error (e.g., fluctuations in eigenvalues), or to use of different rules or procedures. In some cases, these results can be illuminating if the differences across studies are scrutinized (e.g., Avolio, Bass, & Jung's, 1999, study of the Multifactor Leadership Questionnaire). But as we discuss later, use of EFA is often characterized by a general lack of care. Although it is not addressed in the Ford et al. (1986) or Fabrigar et al. (1999) reviews, this lack of care includes failure to discuss the rationale for choosing EFA over CFA and failure to carefully consider previous EFA results (if they exist). It is therefore possible that due to sampling error or careless procedures, EFAs in organizational research produce too many factors (or too few) or assign some variables to the wrong factors.

CFA probably provides a better approach for many existing instruments. It takes sampling error into account more effectively than EFA does, so it is less likely to produce the wrong number of factors or to assign variables to the wrong factors. If there are real population differences from expected results, CFA can detect this by lack of fit. We therefore urge organizational researchers to carefully consider whether EFA is appropriate (e.g., whether there is a clear expectation about the factor pattern) and whether another technique such as CFA might be better used. Further discussions of the relative merits of EFA and CFA can be found in Floyd and Widaman (1995) and Hurley et al. (1997).

Updating Ford et al.'s (1986) Findings

We found that decisions in EFAs conducted from 1985 to 1999 (the period of the present review) were different in some important ways from decisions in EFAs conducted from 1975 to 1984 (the period for Ford et al.'s, 1986, review). Ford et al. (1986) concluded that applied researchers generally had not given adequate thought to their EFA decisions and often applied EFA poorly. They further argued that the poor EFA practices likely had lead to distorted results. (Studies such as Fabrigar et al., 1999, and Gorsuch, 1997, provide further concrete evidence that it can make a considerable difference which EFA practices are used.) It is therefore important to ask whether the situation has changed.

Fabrigar et al. (1999) provided data suggestive of changes, some of which were confirmed by our more comprehensive results and some of which were not. Our find-

ings regarding the number-of-factors decision and choice of rotation are encouraging. In both cases, we found that what we define as high-quality decisions were more common in our review than in the earlier Ford et al. (1986) review. Specifically, regarding the number of factors, researchers were somewhat more likely to use multiple criteria, and regarding rotation, researchers were more likely to choose oblique rotations over orthogonal rotations. The percentages of researchers making high-quality decisions were still low (around 20% for both multiple number-of-factors criteria and oblique rotations), but the increase is encouraging. We hope that in the future more researchers will use these approaches.

Findings for the factor extraction model were not encouraging; rather, they were very similar to Ford et al.'s (1986) findings (although Fabrigar et al., 1999, had suggested a change for the worse). It is interesting to note that of the EFA decisions we considered, the factor extraction model has generated the most research literature. In this literature, some authors have argued strongly that PCA is an excellent substitute for the common factor model (e.g., Velicer & Jackson, 1990). But as we argued in the introduction, other researchers have shown important differences between PCA and common factor solutions (e.g., Gorsuch, 1990), and in such cases, the evidence favors the common factor model as the more accurate. We therefore urge researchers to make greater use of common factor model approaches such as principal axis and maximum likelihood factoring (maximum likelihood is particularly attractive due to the fit indices that can be used to help determine the number of factors) (Browne & Cudeck, 1992).

One disturbing change from Ford et al.'s (1986) results to ours is the relative increase in nonreporting of information (Fabrigar et al.'s [1999] results suggested this increase as well). This finding holds for EFA decisions and also for information such as eigenvalues. Furthermore, as we noted earlier, it also extends to the rationale for choosing to use EFA in the first place. We can only speculate on the reasons for increased nonreporting. Possibly, it is partially due to increased pressure to conserve journal space. This rationale is more likely for information requiring substantial space, such as tables for factor loadings and so forth, but describing the EFA decisions such as method for choosing the number of factors can be done in a short paragraph. Therefore, journal space cannot be the only reason for nonreporting. Another likely reason is a decreased sensitivity regarding the importance of providing details of the analysis. Ford et al. articulately described the importance of reporting EFA information (as have other authors in nonorganizational research journals—e.g., Floyd & Widaman, 1995; H. Tinsley & Tinsley, 1987), but researchers have not responded. We suspect this is due to a persistent belief that it does not matter very much which decisions are made, so there is no reason to report this information. We again refer researchers to demonstrations that EFA decisions can be quite consequential (Fabrigar et al., 1999; Gorsuch, 1990, 1997). Information about an EFA is critical in evaluating the results, and we strongly urge researchers to provide relevant details as outlined by Ford et al. and Fabrigar et al. (1999).

The lack of attention to providing information on EFA, a well-known and long-used analysis technique, may have implications for newer analysis techniques. For example, conducting CFA properly requires understanding its strengths and limitations (e.g., that post hoc modifications render a model somewhat exploratory rather than strictly confirmatory). Other analyses, such as hierarchical linear modeling, also require that decisions be made in the course of analysis. We urge researchers to be

informed about the various decisions required in whatever analysis they use and to be sensitive to the consequences of those decisions.

Regarding sample size, variables, and factors, a direct comparison with Ford et al. (1986) is possible only for sample-to-variable ratio. Ford et al. reported that 27% of their studies had a sample-to-variable ratio of less than 5:1. Our results show the corresponding figure was down to 18%, most likely indicating larger samples. Still, the samples we reviewed were generally not large by MacCallum et al.'s (1999) standards; they suggested that $N = 400$ or greater is a large sample, and less than 30% of studies we reviewed met that standard. We refer readers to MacCallum et al. for a detailed discussion of sample size. Finally, the variable-to-factor ratios were relatively large, indicating factors generally had an adequate number of variables to be well defined.

Purpose of the EFA and the Quality of Decisions

The most encouraging news from our results is that in more consequential uses of EFA, researchers were more than twice as likely to make high-quality decisions about factor extraction model and number of factors criteria, and about twice as likely with regard to rotation. More specifically, when researchers used EFA to address an important study goal such as development of a new measure (when the development was the focus of the study) or hypothesis testing, they were more likely to use a common factor model, to use multiple number-of-factors criteria, and to use oblique rotations. Also important is our finding that nonreporting of information was less common in these more consequential EFA applications.

The more consequential applications of EFA surely have a greater influence on research outcomes and on the cumulative knowledge resulting from organizational research. Although less consequential applications of EFA are certainly important as well, they do not have as much influence on our field. It is encouraging to know that when EFA has had the biggest impact on our research literature, it has been conducted relatively well. Having said this, there is still room for improvement in EFA practices when addressing an important goal of the study. For example, in these more consequential uses, only about 40% of researchers reported using oblique rotations, and important information was often not reported.

What Can Be Done to Improve EFA Usage and Practices?

We have argued that EFA decisions are generally characterized by a lack of care. Before discussing improvement of EFA usage and practices, it makes sense to talk about why researchers make questionable decisions in the first place. We believe that the tendency to use EFA when CFA may be more appropriate is due to a lack of education on both the limits of EFA and the theory and use of CFA. Regarding EFA practices, Ford et al. (1986) speculated that EFA users with relatively little training rely too heavily on statistical software package default settings, which are often (a) PCA, (b) number of factors with eigenvalues greater than 1, and (c) varimax (orthogonal) rotation, that is, the Little Jiffy analysis. We would further note that some software manuals and books appear to urge readers toward these options. For example, regarding factor extraction models, a SAS manual (SAS Institute, 1990) states that "the most important type of analysis performed by the FACTOR procedure is principal components analysis" (p. 777), implying that SAS's default option is preferable. Another

example regarding oblique rotations is from George and Mallery's (1999) book on the use of SPSS for Windows: "Don't even think of attempting an oblique rotation unless you've had a course in factor analysis" (p. 289). Although on a different page, these authors cautioned against doing EFA at all without adequate training, the quote regarding oblique rotations could be interpreted to mean inexperienced EFA users should avoid oblique rotations in favor of orthogonal rotations (the first author had an experience in which another researcher drew this inference). Fabrigar et al. (1999) also discussed possible reasons for poor decisions, including the rarity of EFA training in graduate programs, the fact that EFA literature tends to be highly technical and difficult for many researchers to read, and simply tradition.

The ideal solutions to the problem of poor EFA decisions would be better education for researchers regarding EFA (including appropriate use and guidance about decision making), such as greater emphasis in graduate training. A second suggestion is for editors and reviewers to recognize the importance of EFA decisions and reporting of information and to encourage authors toward high-quality practices. A third suggestion is for methodologists to write less technical papers aimed at users of EFA (Fabrigar et al., 1999). We would add that it is important to have well-written books of the type that researchers are likely to turn to when conducting EFA (e.g., books on using specific software packages). These articles and books need to clearly spell out the appropriate use of EFA as well as different EFA choices and their implications and urge readers to think carefully about their decisions rather than accepting default options.

References

- Avolio, B. J., Bass, B. M., & Jung, D. I. (1999). Re-examining the components of transformational and transactional leadership using the Multifactor Leadership Questionnaire. *Journal of Occupational and Organizational Psychology, 72*, 441-462.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230-258.
- Chan, K.-Y., Drasgow, F., & Sawin, L. L. (1999). What is the shelf life of a test? The effect of time on the psychometrics of a cognitive ability test battery. *Journal of Applied Psychology, 84*, 610-619.
- Conway, J. M. (2000). Managerial performance development constructs and personality correlates. *Human Performance, 13*, 23-46.
- Den Hartog, D. N., Van Muijen, J. J., & Koopman, P. L. (1997). Transactional vs. transformational leadership: An analysis of the MLQ. *Journal of Occupational and Organizational Psychology, 70*, 19-34.
- Dobson, C., Goudy, W. J., Keith, P. M., & Powers, E. (1979). Further analysis of Rosenberg's self-esteem scale. *Psychological Reports, 44*, 639-641.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155-166.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291-314.

- George, D., & Mallery, P. (1999). *SPSS for Windows step by step: A simple guide and reference*. Boston: Allyn and Bacon.
- Goldberg, L. R., & Digman, J. M. (1994). Revealing structure in the data: Principles of exploratory factor analysis. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 216-242). New York: Springer.
- Gorsuch, R. L. (1970). A comparison of biqurtimin, maxplane, promax, and varimax. *Educational and Psychological Measurement*, *30*, 861-872.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*, 33-39.
- Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*, 532-560.
- 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.
- Hater, J. J., & Bass, B. M. (1988). Superiors' evaluations and subordinates' perceptions of transformational and transactional leadership. *Journal of Applied Psychology*, *73*, 695-702.
- Hensley, W. E. (1977). Differences between males and females on the Rosenberg scale of self-esteem. *Psychological Reports*, *41*, 829-830.
- Hensley, W. E., & Roberts, M. K. (1976). Dimensions of Rosenberg's self-esteem scale. *Psychological Reports*, *38*, 583-584.
- Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, *18*, 667-683.
- Kaiser, H. F. (1956). *The varimax method of factor analysis*. Unpublished doctoral dissertation, University of California, Berkeley.
- Kim, J.-O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. Beverly Hills, CA: Sage.
- Leslie, J. B., & Fleenor, J. W. (1998). *Feedback to managers: A review and comparison of multi-rater instruments for management development* (3rd ed.). Greensboro, NC: Center for Creative Leadership.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.
- McCauley, C. D., Ruderman, M. N., Ohlott, P. J., & Morrow, J. E. (1994). Assessing the developmental components of managerial jobs. *Journal of Applied Psychology*, *79*, 544-560.
- Rousseau, D. M., & Tijoriwala, S. A. (1999). What's a good reason to change? Motivated reasoning and social accounts in promoting organizational change. *Journal of Applied Psychology*, *84*, 514-528.
- SAS Institute. (1990). *SAS/STAT user's guide* (Version 6, 4th ed., Vol. 1). Cary, NC: Author.
- Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. Chicago: University of Chicago Press.
- Tinsley, C. (1998). Models of conflict resolution in Japanese, German, and American cultures. *Journal of Applied Psychology*, *83*, 316-323.
- Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*, 414-424.
- 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.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327.
- Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*, 1-28.

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

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

James M. Conway is an associate professor in the Department of Psychology at Central Connecticut State University. His research on confirmatory factor analysis of multi-trait-multimethod data, employment interview ratings, and job performance ratings has appeared in Organizational Research Methods, Journal of Applied Psychology, Journal of Management, and Human Performance.

Allen I. Huffcutt is an associate professor in the Department of Psychology at Bradley University. His research on employment interviews, job performance ratings, and research methods has appeared in Organizational Research Methods, Journal of Applied Psychology, Journal of Organizational Behavior, and Human Performance.

