Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis

Article *in* Organizational Research Methods · April 2004

DOI: 10.1177/1094428104263675

3 authors, including:

James C. Hayton
Rutgers, The State University of New Jersey
40 PUBLICATIONS **3,942** CITATIONS
SEE PROFILE

David Allen
The University of Memphis
63 PUBLICATIONS **4,539** CITATIONS
SEE PROFILE

Factor Retention Decisions in Exploratory Factor Analysis: a Tutorial on Parallel Analysis

James C. Hayton, David G. Allen and Vida Scarpello
Organizational Research Methods 2004 7: 191
DOI: 10.1177/1094428104263675

Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis

JAMES C. HAYTON
Utah State University

DAVID G. ALLEN
University of Memphis

VIDA SCARPELLO
University of Florida

The decision of how many factors to retain is a critical component of exploratory factor analysis. Evidence is presented that parallel analysis is one of the most accurate factor retention methods while also being one of the most underutilized in management and organizational research. Therefore, a step-by-step guide to performing parallel analysis is described, and an example is provided using data from the Minnesota Satisfaction Questionnaire. Recommendations for making factor retention decisions are discussed.

Keywords: *exploratory factor analysis; parallel analysis; factor retention decisions*

Construct validity is a central issue when inferences must be made concerning unobservable or latent variables, and factor analysis is an important tool for questions of validity and the measurement of psychological constructs (Nunnally, 1978). The close association between factor analysis and construct validation has been noted repeatedly in the literature (Gorsuch, 1983; Guilford, 1946; Nunnally & Bernstein, 1994; see Thompson & Daniel, 1996, for a discussion). For example, Gorsuch (1983) argued that a "prime use of factor analysis has been in the development of both the operational constructs for an area and the operational representatives for the theoretical constructs" (p. 350).

Authors' Note: The W. T. Beebe Institute of Personnel and Employment Relations at Georgia State University provided financial support for this research.

Organizational Research Methods, Vol. 7 No. 2, April 2004 191-205
DOI: 10.1177/1094428104263675

Factor analysis can be broadly characterized as a set of multivariate statistical methods for data reduction and for reaching a more parsimonious understanding of measured variables by determining the number and nature of common factors needed to account for the patterns of observed correlations (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Although both exploratory and confirmatory approaches seek to account for as much variance as possible in a set of observed variables with a smaller set of latent variables, components, or common factors, exploratory factor analysis (EFA) is particularly appropriate for scale development or when there is little theoretical basis for specifying a priori the number and patterns of common factors (Hurley et al., 1997). Thus, one of the most critical methodological decisions for researchers using EFA is the number of factors to retain.

The decision regarding the number of factors to retain is important for a number of reasons. One, factor retention decisions may be more important than other relevant decisions (e.g., choice of factor analytic method, type of rotation) because there is evidence of robustness across alternatives for these other decisions (Zwick & Velicer, 1986). Two, EFA needs to balance parsimony with adequately representing underlying correlations, so its utility depends on being able to differentiate major factors from minor ones (Fabrigar et al., 1999). Three, there is conceptual and empirical evidence that both specifying too few factors and specifying too many factors are substantial errors that affect results, although specifying too few is traditionally considered more severe. Both types of misspecifications have been empirically demonstrated to lead to poor factor-loading pattern reproduction and interpretation (Velicer, Eaton, & Fava, 2000).

Specifying too few factors results in the loss of important information by ignoring a factor or combining it with another (Zwick & Velicer, 1986). This can result in measured variables that actually load on factors not included in the model, falsely loading on the factors that are included, and distorted loadings for measured variables that do load on included factors. Furthermore, these errors can obscure the true factor structure and result in complex solutions that are difficult to interpret (Fabrigar et al., 1999; Wood, Tataryn, & Gorsuch, 1996). Although potentially less severe, specifying too many factors can lead to focusing on minor factors at the expense of major ones, the creation of factors with only one high loading, factors that are difficult to interpret, and factors that are unlikely to replicate (Zwick & Velicer, 1986). Therefore, selecting both too few or too many factors have significant consequences for the reduction and interpretation of information in a data set.

Despite the importance of factor retention decisions and extensive research on methods for making retention decisions, there is no consensus on the appropriate criteria to use. A number of criteria are available to assist these decisions, but they do not always lead to the same or even similar results (Carraher & Buckley, 1991; Thompson & Daniel, 1996; Zwick & Velicer, 1986). There is evidence, however, that parallel analysis (PA) (Horn, 1965) is one of the most accurate methods for determining the number of factors to retain (e.g., Velicer et al., 2000; Zwick & Velicer, 1986), while also being one of the most underutilized methods (e.g., Fabrigar et al., 1999; Ford, MacCallum, & Tait, 1986). Possible reasons for the lack of widespread use of PA include a lack of training in graduate school, the lack of inclusion of the method in most textbook discussions of the topic, lack of awareness by researchers because much of the factor analysis literature is complex and heavily quantitative, difficulty in performing PA, and simply tradition (Fabrigar et al., 1999). Velicer et al. (2000) rated

the ease of implementation of PA as difficult. To the extent that lack of training, lack of awareness, and perceptions of difficulty account for the limited use of PA, it may be useful to present a clear, user-friendly description of the technique and its use. Therefore, we offer a brief description of the major factor retention methods, the evidence regarding their accuracy, the evidence regarding their use in published articles, and an example of how to perform PA.

Factor Retention Criteria

In general, factor analysts should retain factors until additional factors account for trivial variance; however, different methods of specifying the number of factors to retain often lead to different solutions. One of the most commonly used methods is the Kaiser or mineigen greater than 1 criterion (K1), which retains factors with eigenvalues greater than 1 (Kaiser, 1960). Guttman (1954) showed that the lower bound for the rank of a population correlation matrix is obtained by extracting only those factors with eigenvalues greater than 1. The rank of a correlation matrix is equal to the smallest number of components than can account exactly for the off-diagonal correlations in the matrix. Kaiser (1970) proposed the same rule of thumb, with the rationale that the reliability of a component must always be nonnegative when its eigenvalue is greater than 1. The K1 rule is the default retention criterion for a number of commonly used statistical packages (e.g., SPSS, SAS).

The theoretical basis and ease of use of the K1 rule have led to its widespread adoption (Gorsuch, 1983, p. 149). However, there are at least three problems associated with this approach. One, the rule is intended as a lower bound for the rank of the correlation matrix and thus an upper bound for the number of factors to be retained. However, in practice, it is often used as the criterion to determine the exact number of factors (Gorsuch, 1983). Further, Guttman's proof applies only to the population correlation matrix. For finite samples, sampling error adds to the rank of a given correlation matrix, and thus the K1 criterion would tend to overestimate the number of factors (Horn, 1965). Finally, the rule is somewhat arbitrary in that it draws distinctions between factors with eigenvalues just above and just below 1 (Fabrigar et al., 1999).

Another commonly used method for determining the number of factors to retain is Cattell's (1966) scree test, which involves an examination of a plot of the eigenvalues for breaks or discontinuities. The rationale for this test is that a few major factors account for the most variance, resulting in a steep "cliff" as these factors are identified first, followed by a shallow "scree" describing the small and relatively consistent variance accounted for by the numerous minor factors. The criterion for retention is relatively straightforward: Identify the break point at which the scree begins and retain only factors that do not belong to the scree (Cattell & Jaspers, 1967).[1]

Although the scree test may work well with strong factors, it suffers from subjectivity and ambiguity, especially when there are either no clear breaks or two or more apparent breaks. Definite breaks are less likely with smaller sample sizes and when the ratio of variables to factors is lower (Cliff & Hamburger, 1967; Linn, 1968). Such subjectivity and ambiguity may lead to problems with low interrater reliabilities (e.g., Crawford & Koopman, 1979), especially with complex factor structures; however, several studies of scree test interrater reliabilities have reported satisfactory results (e.g., Cliff, 1970; Tucker, Koopman, & Linn, 1969; Zwick & Velicer, 1982).

A third factor retention method is PA (Horn, 1965). PA attempts to overcome a primary limitation of the K1 criterion: the overestimation of matrix rank due to sampling error. K1 is based on an assumed population correlation matrix and is appropriate only as the sample size approaches infinity (Glorfeld, 1995). In a population matrix, the eigenvalues for random or mutually uncorrelated variables would equal 1. However, in a finite sample, sampling error and least-squares bias lead initial eigenvalues to be greater than 1 and later eigenvalues to be less than 1 (Horn, 1965; Turner, 1998). This means that for finite samples, some factors with eigenvalues greater than 1 may occur purely as a result of sampling error. PA adjusts for the effect of sampling error and therefore is a sample-based alternative to the population-based K1 criterion (Carraher & Buckley, 1991; Zwick & Velicer, 1986).

The rationale underlying PA is that nontrivial components from real data with a valid underlying factor structure should have larger eigenvalues than parallel components derived from random data having the same sample size and number of variables (Ford et al., 1986; Lautenschlager, 1989). Thus, PA involves the construction of a number of correlation matrices of random variables based on the same sample size and number of variables in the real data set. The average eigenvalues from the random correlation matrices are then compared to the eigenvalues from the real data correlation matrix, such that the first observed eigenvalue is compared to the first random eigenvalue, the second observed eigenvalue is compared to the second random eigenvalue, and so on. Factors corresponding to actual eigenvalues that are greater than the parallel average random eigenvalues should be retained. Actual eigenvalues less than or equal to the parallel average random eigenvalues are considered due to sampling error (Glorfeld, 1995; Horn, 1965; Zwick & Velicer, 1986). Thus, researchers would not be interested in a factor that does not account for more variance than the parallel factor obtained from random numbers because meaningful components extracted from actual data should have larger eigenvalues than parallel eigenvalues obtained from random data (Montanelli & Humphreys, 1976; Turner, 1998).

There are other methods for deciding how many factors to retain as well. Bartlett's (1950) chi-square test tests the hypothesis that remaining eigenvalues are equal. Each eigenvalue is assessed sequentially until the chi-square test of the null fails to be rejected. Velicer's (1976) minimum average partial method (MAP) calculates the average of squared partial correlations after each component is partialled out. When the minimum average squared partial correlation is reached, the residual matrix resembles an identity matrix, and no further components are extracted. Other methods include maximum likelihood estimation and less quantitative approaches such as choosing the most interpretable solution and relying on theoretical expectations. In practice, many researchers use multiple decision criteria. There is a great deal of evidence, though, that all of the above methods differ in their ability to accurately identify factor structure and that researchers frequently rely on the K1 test and the scree test. We next review evidence for the relative accuracy of these factor retention methods. This is followed by evidence of the frequency of their use based on a review of the recent management literature.

Evaluating the Accuracy of Factor Retention Criteria

Numerous studies have investigated the relative merits and accuracy of various factor retention criteria, typically with simulated data (e.g., Humphreys & Montanelli,

1975; Linn, 1968; Zwick & Velicer, 1986), although a few have used actual sample data (e.g., Carraher & Buckley, 1991, 1995; Lee & Comrey, 1979; Revelle & Rocklin, 1979). A number of studies have shown that the K1 rule is inaccurate and tends to overfactor. Specifically, Horn (1965) showed that the K1 criterion overfactored compared to PA. Linn (1968) demonstrated that K1 overestimated the correct number of factors by 66%. Silverstein (1987) compared the K1 rule with PA on 24 data sets and concluded that PA consistently outperformed the K1 rule. Zwick and Velicer (1986), in a comprehensive comparison of the K1 rule, Bartlett's test, the scree test, MAP, and PA, concluded that K1 was correct only 22% of the time. Furthermore, the K1 rule overfactored in each case that it was found to be inaccurate (see Glorfeld, 1995, for a review). Zwick and Velicer (1986, p. 441) recommended that the K1 rule no longer be used as an exclusive or even primary method for determining the number of components to retain. In their review of factor retention methods, Fabrigar et al. (1999) concluded that not only is there substantial evidence that K1 is inaccurate, but they could find no study in which this rule worked well.

The scree test suffers from subjectivity and ambiguity as noted above, although it may work better when underlying factors are empirically distinct and unambiguously reflected by their measures (Fabrigar et al., 1999). Zwick and Velicer's (1986) comparison concluded that the scree test performed better than the K1 rule, although it was still correct only 57% of the time. Furthermore, in 90% of the cases that the scree test was inaccurate, it was found to overestimate the number of factors. There was also variation in interrater reliability (Zwick & Velicer, 1986).

Less research has evaluated the accuracy of other methods. Humphreys and Montanelli (1975) compared PA with a maximum likelihood procedure and found that the maximum likelihood procedure tended to overfactor. Furthermore, maximum likelihood tests are heavily influenced by sample size, with accuracy declining as sample size increases (Fabrigar et al., 1999). Zwick and Velicer's (1986) comparison found that Bartlett's test was correct about 30% of the time and tended to overfactor, particularly as sample size increased. The MAP procedure was the second most accurate approach in Zwick and Velicer's (1986) study, correct 84% of the time (PA was most accurate). Unlike the other methods evaluated, MAP tended to underfactor when it was inaccurate (90% of the inaccurate cases).

The PA approach tends to perform quite well in research evaluating the accuracy of factor retention methods (Fabrigar et al., 1999). For example, Silverstein's (1987) comparison of PA and the K1 rule concluded that PA was more accurate, although when inaccurate, it tended to overfactor by 1. Humphreys and Montanelli (1975) compared PA and maximum likelihood and concluded that PA was more accurate and almost always correct. Zwick and Velicer's (1986) comparison concluded that PA was the most accurate of the five methods evaluated and that it was correct 92% of the time. They also found a slight tendency of PA to overfactor when it did err (66% of the inaccurate cases). More recently, Eaton, Velicer, and Fava (1999) compared PA, MAP, and the K1 criterion and concluded that PA was the most accurate approach, followed closely by MAP, with K1 being extremely inaccurate.

In reviewing the evidence on factor retention, Glorfeld (1995) concluded that there is little reason to choose any method other than PA. Similarly, *Educational and Psychological Measurement*'s editorial recommendations regarding factor retention indicate that too many researchers rely on K1 when more should be using PA because of its proven merit (Thompson & Daniel, 1996). More recently, Velicer et al. (2000)

reviewed the literature comparing nine alternative methods on rationale, ease of implementation, and accuracy and could recommend only the use of PA and MAP, with the scree used as an adjunct but not by itself.

Evaluating the Use of Factor Retention Criteria

Given the evidence and strong recommendations regarding the accuracy of PA compared to other factor retention methods, especially K1, we might expect that the use of PA would be on the rise relative to other approaches. However, there is substantial evidence that researchers continue to rely on the K1 and scree tests. At least two published reports surveying the use of factor retention rules are available.

Ford et al. (1986) reviewed *Journal of Applied Psychology*, *Organizational Behavior and Human Decision Processes*, and *Personnel Psychology* for the period 1975-1984 and found 152 articles reporting exploratory factor analyses. The most commonly reported factor retention criterion was the K1 rule (21.7%). K1 was also relied on more heavily in the period 1980-1984 than the period 1975-1979, suggesting that its use was increasing. The scree test, a priori theory, and interpretability were each used in 11.2% of the analyses. Another 13.8% reported using some combination of the above methods. There were no reported uses of PA. Furthermore, more than 30% of the analyses failed to indicate what factor retention rules were used, and this omission was higher for the period 1980-1984 than for the period 1975-1979. Ford et al. (1986) concluded both that researchers should report their factor retention criteria and that PA seemed to be an overlooked method.

Fabrigar et al. (1999) reviewed *Journal of Applied Psychology* and *Journal of Personality and Social Psychology* for the period 1991-1995. They found 58 reports of exploratory factor analysis in *Journal of Applied Psychology* and 159 in *Journal of Personality and Social Psychology* and reported very similar results to Ford et al. (1986) in terms of factor retention methods. The K1 rule was the most commonly used (19% for *Journal of Applied Psychology*, 15.7% for *Journal of Personality and Social Psychology*), followed by the scree test (15.5% for *Journal of Applied Psychology*, 15.1% for *Journal of Personality and Social Psychology*). One analysis reported using PA in *Journal of Personality and Social Psychology* and none reported utilizing PA in *Journal of Applied Psychology*. Other methods reported in these journals included prior theory and ease of interpretability. About 20% in both journals reported using some combination of K1, scree, prior theory, and interpretability. Close to 40% of articles reporting the results of factor analysis did not indicate what factor retention criteria were used.

We conducted our own review of two major management journals for the period 1990-1999: *Academy of Management Journal* and *Journal of Management*. Previous reviews have focused largely on psychology journals. Therefore, it may be valuable to determine if the trends noted above hold in the management literature as well. Both *Academy of Management Journal* and *Journal of Management* are prestigious management journals that include a wide range of organizational research.

Our review discovered a similar pattern of results, although more researchers appeared to use multiple methods. One hundred forty-two articles were found that used EFA. Nearly one half of those (67 studies, or 47.2%) reported using some combination of the K1 rule, the scree test, and prior theory to determine the number of factors to retain, with the most popular combination (22 studies, or 15.5%) including K1 and

prior theory. A sizeable percentage relied solely on the K1 rule (36 studies, or 25.4%), despite prior evidence of its limitations. Few studies relied solely on a scree test (8 studies, or 5.6%). None of the researchers reported using other methods such as PA. Notably, more than half of the studies (75 studies, or 52.8%) in our sample did not report which method was used to determine the number of factors to retain.

In summary, our review reveals two important points regarding factor retention. One, PA is one of, if not the, most accurate method for determining the number of factors to retain. Two, despite its accuracy, PA is not widely used in published research, particularly when compared with the less accurate K1 rule. To the extent that researchers are unfamiliar with PA or perceive it as difficult to implement, a clear explication of how to perform PA might be useful. Early criticisms of PA focused on the computational difficulty involved in generating a large number of random data matrices (e.g., Montanelli & Humphreys, 1976; Longman, Cota, Holden, & Fekken, 1989). Silverstein (1977) suggested that the use of PA would increase as computer resources increased in power and affordability.

Concerns about generating random data matrices resulted in efforts to circumvent this requirement. Researchers have offered formulas that approximate the average eigenvalues of random data matrices (e.g., Montanelli & Humphreys, 1976) and provided tables of permutations of eigenvalues for combinations of sample size and number of variables (e.g., Buja & Eyuboglu, 1992). However, Lautenschlager (1989) demonstrated that using equations to predict PA results can be highly inaccurate and recommended against their general use. Eaton et al. (1999) compared several methods of conducting PA and found that the table and regression approaches were more limited in the types of cases that could be analyzed. They concluded that random data generation and the table methods displayed the best overall performance. For most researchers, because of increases in personal computing power, there is no longer cause for concern with generating even a large number of random matrices. Therefore, we provide an example detailing the use and interpretation of PA using one widely used statistical package (SPSS).

Conducting Parallel Analysis

Parallel analysis may be implemented in a number of ways. We have focused on SPSS due to its relative simplicity and widespread use. However, those familiar with SAS, C++, or Fortran could create a similar program to the one illustrated here and follow the steps as outlined in Figure 1. Although SPSS currently has no option for parallel analysis, it is relatively simple to conduct. The analysis can be conducted in four steps, summarized in Figure 1.

The first step involves generating a random data set of the same dimensions as the data being analyzed. That is, the random data set must have the same number of observations (n) and variables (v) as the real data that are being analyzed. This can be quickly achieved using the program shown in Figure 2, which was adapted from Thompson and Daniel (1996). Lines 1 to 22 of this program deal with the creation of a single random data set. Line 3 determines how many observations will be created. Therefore, it is necessary to change n to the actual sample size in the data set of interest. Line 5 determines how many variables the random data set will have. Thus, in line 5, the number of variables should be edited from 50 to the number of items being analyzed in the real data set.

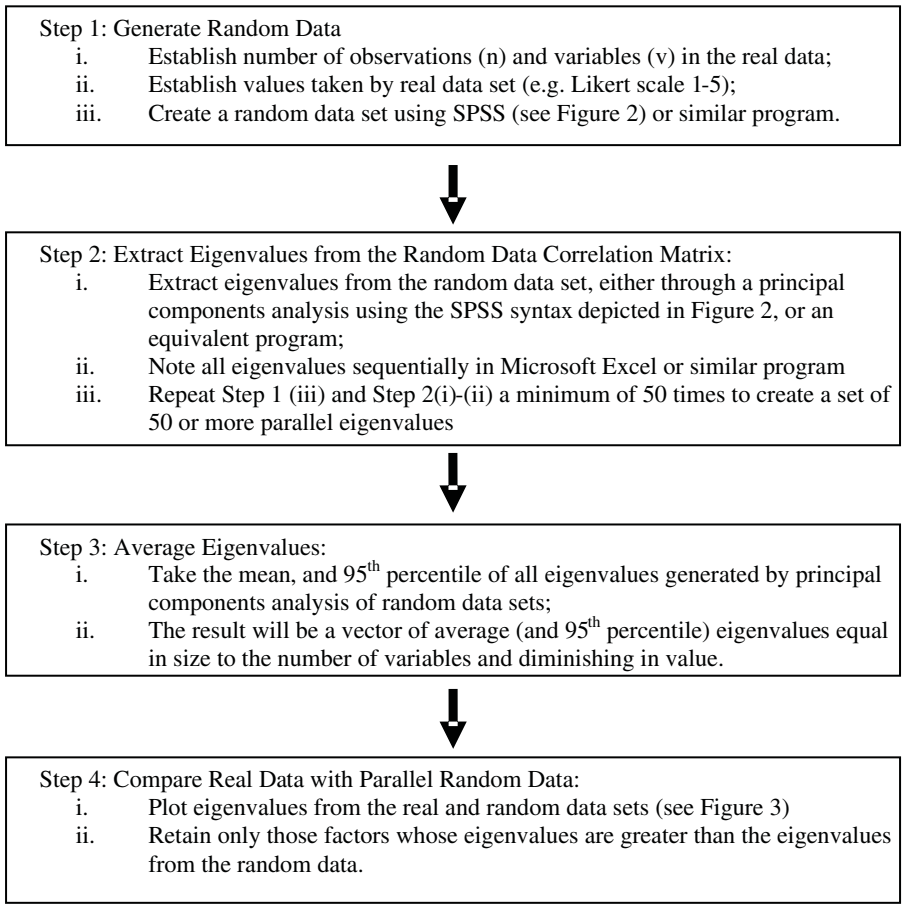

Figure 1: Step-by-Step Guide to Conducting Parallel Analysis

In addition to these changes, it is also important to ensure that the values taken by the random data are consistent with those in the comparison data set. The purpose of line 7 is to ensure that the random variables are normally distributed within the parameters of the real data. Therefore, line 7 must be edited to reflect the maximum and midpoint values of the scales being analyzed. For example, if the measure being analyzed is a 7-point Likert-type scale, then the values 5 and 3 in line 7 must be edited to 7 and 4, respectively. Lines 14 and 15 ensure that the random data assumes only values found in the comparison data and so must be edited to reflect the scale minimum and maximum, respectively.

In the second step, the newly generated random data are subject to principal components analysis to extract the eigenvalues (lines 26-37). A principal components model should be used because the population matrix from which the sampled random data matrices are drawn is itself an identity matrix. In other words, the random data are free from measurement error.[2] The variables identified in lines 27 through 32 should correspond to the number of variables being analyzed. Because these data are random, the

variation of resulting eigenvalues around 1 reflects the effects of sampling error. If we were able to generate an infinitely large random data set and then subject it to a factor analysis, the eigenvalues would all theoretically equal 1. Using just a single random data set risks bias; therefore, Steps 1 and 2 are repeated multiple times. Horn (1965) suggested that the number of random data matrices should be "reasonably large." Turner (1998) employed 100 repetitions. Other researchers recommend as many as 500 to 1,000 repetitions. Crawford and Koopman (1979) compared results from 1 random data set with the averaged eigenvalues across 100 random data sets and found no significant differences. Although there is no standard for the number of times to repeat Steps 1 and 2, 50 appears to be commonly used (G. Lautenschlager, personal communication, April 29, 1997). In general, the greater the number of repetitions, the more accurate will be the final results.

Repeating both steps 50 times thus results in 50 sets of eigenvalues. In the third step, the average of the eigenvalues across the 50 sets is calculated. That is, the average of the first eigenvalue is calculated, then the average of the second eigenvalue, and so on. This results in a single set of average eigenvalues to which we compare the eigenvalues drawn from the actual data set.

In this example, we have used the mean of the randomly generated eigenvalues. However, it has been suggested that using the average eigenvalues is analogous to setting the Type I error rate (α) to .50 (rather than the more common $\alpha = .05$) (Glorfeld, 1995; Harshman & Reddon, 1983). Given that PA has shown a slight tendency to overfactor, Glorfeld (1995) and Harshman and Reddon (1983) suggested that using the 95th percentile of eigenvalues generated from the random data is more conservative. This is also similar to setting α equal to .05, which is the more common standard for Type I error. Therefore, the same basic procedure outlined in Steps 1 through 3 may be used, taking the 95th percentile of each eigenvalue rather than the mean.

In the fourth step, the eigenvalues from the actual data are compared to those from the randomly generated data. Factors from the real data with eigenvalues greater than the corresponding eigenvalue from the random data (either the average or the 95th percentile) would be retained. Thus, the first actual eigenvalue would be compared to the first random eigenvalue, the second actual eigenvalue would be compared to the second random eigenvalue, and so on. Such a comparison can be easily made merely by examining the numbers; however, it may also be useful to plot the average and 95th percentile of the eigenvalues from the random data against the scree plot from the original data.

Table 1 shows the first 10 actual eigenvalues drawn from a sample administration of the 20-item short form of the Minnesota Satisfaction Questionnaire (MSQ) (Weiss, Dawis, England, & Lofquist, 1967) to 579 participants, as well as the average and 95th percentile eigenvalues drawn from random data as described above. The short form of the MSQ contains 20 items anchored with a 5-point response format ranging from *very dissatisfied* to *very satisfied*. Each item represents a single dimension of worker preferences, for example, "On your present job, how satisfied are you with the chance to do things for other people?" In empirical studies, two factors have typically been reported (e.g., Bledsoe & Baber, 1979; Hauber & Bruininks, 1986; Weiss et al., 1967). These reflect the intrinsic and extrinsic sources of job satisfaction.

Examination of the results in Table 1 indicates that only the first 2 actual eigenvalues are greater than those generated by PA (for both the average and 95th percentile criteria) and thus would be retained. Using the K1 or mineigen greater than 1

Table 1
Actual and Random Eigenvalues

Actual Eigenvalue	*Average Eigenvalue*	*95th Percentile Eigenvalue*
6.368	1.348	1.399
1.619	1.292	1.323
1.208	1.248	1.276
1.130	1.206	1.237
0.994	1.173	1.194
0.934	1.142	1.166
0.881	1.106	1.138
0.858	1.076	1.100
0.793	1.047	1.068
0.727	1.020	1.039

criterion would indicate the retention of the first four factors because these actual eigenvalues are greater than 1. Although it is difficult to say what the "correct" number of factors should be, the two-factor solution is more consistent with previous theory and research on the MSQ showing an intrinsic and an extrinsic factor. Furthermore, recall that the K1 criterion has been criticized for its tendency to recommend retaining too many factors. The use of the average versus the 95th percentile criterion made no substantive difference in this example, and we found similar results analyzing data from other job satisfaction measures. Because calculating both adds very little complexity to the process, researchers using PA should consider doing so. Where the two criteria do not agree, the 95th percentile criterion may reduce PA's slight tendency to retain too many factors.

As noted earlier, plotting the actual versus randomly generated eigenvalues may provide a clear visual comparison of the results. Figure 3 shows a plot of the eigenvalues from the MSQ along with the mean and 95th percentiles of the eigenvalues for the random data that were generated in the fashion described above. PA would indicate retaining the two factors whose actual eigenvalues lie above the lines representing the randomly generated eigenvalues.

Discussion

Construct definition, measurement, and validity are critical to the behavioral sciences, and determining the number of meaningful factors represented by measures is an important step. Parallel analysis is one of the most accurate methods of deciding the appropriate number of factors to retain and yet is rarely used in the management and organizational research literature. It is hoped that the step-by-step procedure described in this article will encourage more organizational researchers to use PA.

PA is still not as easy to implement as either the K1 criterion or the scree test. Graduate students at two universities familiar with factor analysis but not necessarily with PA reported taking between 10 and 30 minutes to replicate the example provided here with their own data. Replicating the creation of a random data set 50 times and then calculating the mean and 95th percentile eigenvalues required the most time. However, that time frame does not seem unreasonable to make more accurate factor retention decisions and would likely decrease with experience. Furthermore, the graduate

```
Title “Parallel Analysis”.
1) INPUT PROGRAM.
2) COMMENT This part of the program creates a single random data set.
3) LOOP LOOP#1 = 1 TO n.
4) COMMENT Change n to the actual sample size.
5) DO REPEAT V = V1 TO V50.
6) COMMENT Replace 50 above with the number of variables.
7) COMPUTE V = RND (NORMAL (5/6) + 3).
8) COMMENT This line relates to the response levels
9) COMMENT 5 represents the maximum response value for the scale.
10) COMMENT Change 5 to whatever the appropriate value may be.
11) COMMENT 3 represents the middle response value.
12) COMMENT Change 3 to whatever the actual middle response value may be.
13) COMMENT (e.g., 3 is the midpoint for a 1 to 5 Likert scale).
14) IF (V LT 1)V = 1.
15) IF(V GT 5)V = 5.
16) COMMENT These lines constrain the random numbers to the appropriate response range.
17) COMMENT Change the 5s above to the maximum response value.
18) END REPEAT.
19) END CASE.
20) END LOOP#1.
21) END FILE.
22) END INPUT PROGRAM.
23) COMMENT After creating the random data it must be factored.
24) COMMENT Change the number of variables in the variable and analysis section to the
25) number of variables in the analysis.
26) FACTOR
27)   /VARIABLES v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13 v14 v15 v16 v17 v18
28)   v19 v20 v21 v22 v23 v24 v25 v26 v27 v28 v29 v30 v31 v32 v33 v34 v35 v36 v37
29)   v38 v39 v40 v41 v42 v43 v44 v45 v46 v47 v48 v49 v50 /MISSING LISTWISE
30)  /ANALYSIS v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13 v14 v15 v16 v17 v18 v19
31)   v20 v21 v22 v23 v24 v25 v26 v27 v28 v29 v30 v31 v32 v33 v34 v35 v36 v37 v38
32)   v39 v40 v41 v42 v43 v44 v45 v46 v47 v48 v49 v50
33)   /PRINT INITIAL EXTRACTION
34)   /CRITERIA MINEIGEN(0) ITERATE(25)
35)   /EXTRACTION PC
36)   /ROTATION NOROTATE
37)   /METHOD=CORRELATION.
38)   COMMENT Repeat the entire program 50 times and average the resulting eigenvalues.
```

Figure 2: SPSS Program for Conducting Parallel Analysis[a]

a. This program is adapted from Thompson and Daniel (1996). This program employs principle components analysis; in order to conduct PA using common factor analysis lines 26 through 37 should be replaced with the appropriate SPSS syntax. Remove line references prior to running the program in SPSS.

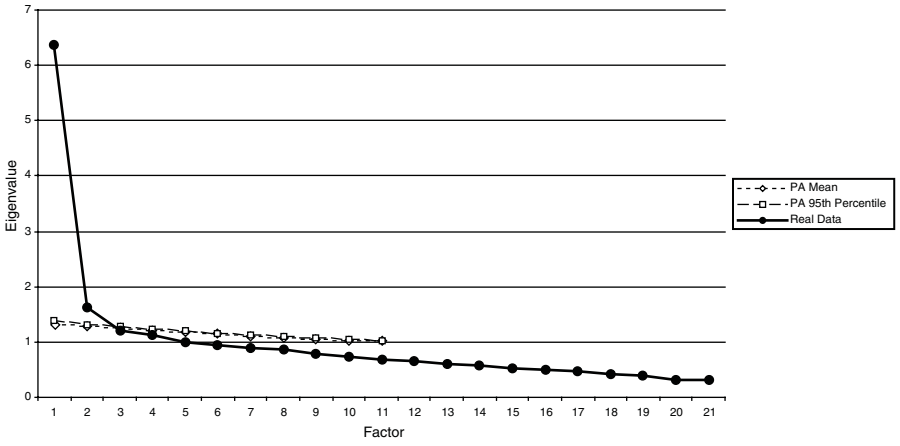

Figure 3: Plot of Actual Versus Randomly Generated Eigenvalues

students performed PA on personal computers, confirming that computational resources are probably much less of a barrier than they used to be.

Given the evidence, organizational researchers should now consider PA as a primary method for factor retention decisions in EFA and should also rely less on the K1 criterion. Using PA with the 95th percentile criterion may be a more conservative approach that mitigates PA's slight tendency to recommend retaining too many factors, although there is usually little substantive difference between the 95th percentile and average random eigenvalues. Some researchers have also suggested that PA should be used in conjunction with the scree test (Fabrigar et al., 1999; Ford et al., 1986) because the scree test is easy to implement and provides a visual component for comparison to the parallel random eigenvalues.

One caution about PA is offered by Turner (1998), who noted that due to the interdependent nature of eigenvalues, the presence of a large first factor in a PA will reduce the size of noise eigenvalues. The consequence is that in certain situations, PA can underfactor, which is potentially more serious than overfactoring. The impact of this limitation is most serious for smaller sample sizes, where there is a high correlation between factors, or where a second factor is based on a relatively small number of items. However, Turner also noted that when used with care, PA remains a potentially "valuable method," although it should be used in conjunction with other methods such as the K1 rule and scree test.

Researchers might also consider using PA in conjunction with the MAP procedure. MAP is the only other factor retention method that approaches the accuracy of PA in comparative studies (e.g., Eaton et al., 1999; Zwick & Velicer, 1986). Furthermore, MAP is also the only method that shows a tendency to recommend retaining too few factors when it is inaccurate, which could prove valuable given PA's slight tendency to recommend retaining too many. Given the scarcity of organizational research reporting the use of the MAP procedure and Velicer et al.'s (2000) rating of its implementation as difficult, researchers might also benefit from a straightforward explanation of how to perform this method as well. The MAP approach is actually quite easily imple-

mented in using the SAS statistical software, which has an option to output the average partial correlation for a given number of factors.[3]

We hope this discussion encourages researchers to carefully consider factor retention decisions in EFA and also to report the factor retention criteria used. We have argued that PA should be used more often and have tried to aid the implementation of PA. Still, the best factor retention approach may be to attempt to gather convergent information from multiple sources, such as PA, scree tests, MAP, and theory, followed by cross-validation and confirmatory approaches.

Notes

1. Note that this differs from Cattell's (1966) original criterion, which also retained the first factor on the scree.

2. Rather than using a principal components model, any method of extracting the eigenvalues from the correlation matrix of random data is appropriate. The principal components approach is one that is readily available to researchers with limited programming skills.

3. Thanks to an anonymous reviewer for pointing this out.

References

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

Bledsoe, J. C., & Baber, W. C. (1979). Factor invariance in the measurement of job satisfaction. *Perceptual and Motor Skills*, *48*, 985-986.

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

Carraher, S. M., & Buckley, M. R. (1991). The effect of retention rule on the number of components retained: The case of the Pay Satisfaction Questionnaire. *Proceedings of the Southern Management Association*.

Carraher, S. M., & Buckley, M. R. (1995). The effect of retention rule on the number of components retained: The case of cognitive complexity and the PSQ. *Proceedings of the Southern Management Association*.

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

Cattell, R. B., & Jaspers, J. (1967). A general plasmode for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, *3*, 1-212.

Cliff, N. (1970). The relation between sample size and population characteristic vectors. *Psychometrika*, *35*, 163-178.

Cliff, N., & Hamburger, C. D. (1967). The study of sampling errors in factor analysis by means of artificial experiments. *Psychological Bulletin*, *68*, 430-445.

Crawford, C. B., & Koopman, P. (1979). Note: Inter-rater reliability of scree test and mean square ratio test of number of factors. *Perceptual and Motor Skills*, *49*, 223-226.

Eaton, C. A., Velicer, W. F., & Fava, J. L. (1999). *Determining the number of components: An evaluation of parallel analysis and the minimum average partial correlation procedures*. Unpublished manuscript.

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.

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.

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

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

Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, *6*, 427-439.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, *19*, 149-162.

Harshman, R. A., & Reddon, J. R. (1983). Determining the number of factors by comparing real with random data: A serious flaw and some possible corrections. *Proceedings of the Classification Society of North America at Philadelphia*, 14-15.

Hauber, F. A., & Bruininks, R. H. (1986). Intrinsic and extrinsic job satisfaction among direct-care staff in residential facilities for mentally retarded people. *Educational and Psychological Measurement*, *46*, 95-105.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *32*, 179-185.

Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*, 193-206.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., et al. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18*, 667-683.

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

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

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

Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research*, *14*, 301-321.

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

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

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

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

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*, 403-414.

Silverstein, A. B. (1977). Comparison of two criteria for determining the number of factors. *Psychological Reports*, *41*, 387-390.

Silverstein, A. B. (1987). Note on the parallel analysis criterion for determining the number of common factors or principal components. *Psychological Reports*, *61*, 351-354.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, *56*, 197-208.

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

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

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

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

Weiss, D. J., Dawis, R. V., England, G. W., & Lofquist, L. H. (1967). Manual for the Minnesota Satisfaction Questionnaire. *Minnesota Studies in Vocational Rehabilitation*, 22.

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*, 254-365.

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). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

James C. Hayton is an assistant professor of management at Utah State University. He received his Ph.D. from the W. T. Beebe Institute of Personnel and Employment Relations at Georgia State University's Robinson College of Business. His research focuses on the links among human resource management, organizational learning and innovation, and corporate entrepreneurship.

David G. Allen is an assistant professor of management in the Fogelman College of Business at the University of Memphis. He earned his Ph.D. from the W. T. Beebe Institute of Personnel and Employment Relations at Georgia State University. His research interests include the flow of human resources into and out of organizations and technology implications for human resource management.

Vida Scarpello received her Ph.D. in Industrial Relations from the University of Minnesota. She has taught at the Universities of Georgia and Florida and was Professor of Management at Georgia State University until retirement in 2002. Her research spans a wide range of topics such as theory and measurement of job attitudes; strategic, justice, and measurement issues in compensation; and international organizational relations. She continues to work with the National Science Foundation's Industry/University Cooperative Research Center Program, is editing a book on human resource management education in the United States, and has reestablished a professional relationship with the University of Florida.