ISSUES AND RECOMMENDATIONS IN THE USE OF FACTOR ANALYSIS

by

Peter D. MacIntyre

Peter D. MacIntyre is a Ph.D. candidate in the Department of Psychology.

It is not uncommon to encounter the attitude that statistics are "cut and dry" entities - rather boring - without much inherent interest. At some levels this may be true, but as one becomes more involved with solving research problems, it becomes clear that statistics are very useful tools. As with a carpenter, the researcher's choice of tools can influence the quality of his/her work in important ways. Computers have made the most powerful multivariate statistics available to a wide range of users and the skills actually required to generate the statistics are minimal. It is the process of interpreting the results that can trouble even the most experienced of researchers.

The reason for these difficulties originates in the many choices that one makes when analyzing data. Each decision has pitfalls that cannot be avoided. The best that we can do is to be aware of the trips and traps and to make informed judgments rather than acquiescing to the preferences of a computer. Each judgement can be debated and this gives statistics a character all their own.

This paper reviews some of the practical issues involved in performing one powerful statistical procedure: factor analysis. Factor analysis is a tool used to discover the components that underlie, and explain the variation in, several variables that the researcher has measured. It also provides information about the relations of the variables to these underlying components and the relations among the dimensions that they describe. It is this wealth of information that makes factor analysis "powerful" for answering several types of questions. The approach has been employed in a wide range of disciplines, including business, psychology, sociology, agriculture, economics, and others, because of its ability to pull variables apart and to describe them in terms of the "permanent" factors that influence them.

The topic itself would be applicable to anyone with an interest in quantitative data analysis. The discussion of factor analysis that follows assumes that the reader has a basic knowledge of data collection and analysis methods, especially correlation. This paper is intended both for those less familiar with factor analysis and for readers who want a brief review of some recent findings in the literature. A vocabulary has developed along with factor analysis and a glossary has been included at the end of the paper to clarify the jargon that readers will find in most factor analytic reports. Even the term "factor analysis" has developed several meanings based on the several ways that the statistics can be computed and the implications of each method. The differences among methods has been emphasized, however, the methods are highly similar and the term "Factor analysis" will be used here in its most general sense.

Not all of the issues related to factor analyses can be discussed in a single paper. Following a brief description of the purpose and procedures of factor analysis, the focus here will centre on four well debated issues. The first topic under consideration is the sample size required to perform these analyses. Next, five standard methods of extracting factors will be compared. Both of these problems seem to have been adequately solved. The most serious remaining decision is the number of factors to retain. The standard decision rule has been thoroughly discredited and some of the better methods will be examined. Also, the under-emphasized issue of indeterminacy, which has serious implications for the very foundation of factor analysis, will be discussed. Finally, appendix A presents an example analysis that attempts to incorporate the recommendations that will be made throughout this article.

Purpose and Procedure of Factor Analysis

In general, a factor analysis summarizes a correlation (or covariance) matrix of a larger number of variables in terms of a smaller number of combinations of variables called factors. Each factor is presumed to be a dimension underlying the variables associated with it, therefore, the factor aids in explaining the nature of several variables at once. It also points to the properties shared by several variables as well as the combination of forces (factors) influencing any one variable.

The calculations begin by finding a set of coefficients that can be multiplied by the scores on the raw variables. In one approach (called Principal Components) the coefficients or weights are generated such that, when combined with the variables and added together, they account for the maximum amount of variability in the original correlation matrix. When the weights are applied to the data for a given individual and added up, the resulting number is the "factor score" for that person. A simplified version of the equation looks something like this:

Factor score = (weight) * data for Variable 1 +
(weight) * data for Variable 2 +
(weight) * data for Variable 3 +
...

After finding a first set of factor scores, the correlations are reduced by partialling out the scores of the first factor. The amount of variability in the correlation matrix that can be accounted for depends on the variance in the set of factor scores. The variance associated with a set of factor scores is called the eigenvalue. High eigenvalues indicate strong factors that account for a substantial portion of the original correlations.

A second set of weights is then generated such that they are not correlated with the first set. Emphasis is placed on a different set of variables. The resulting scores are partialled out of the remaining correlation matrix. This procedure continues until the elements in the correlation matrix are reduced to near zero, indicating that the original matrix can be adequately represented by the extracted factors. Because of the procedure, the initial set of factors may be difficult to interpret. Usually, these factors are "rotated" in order to aid in interpretation. The rotation of the axes representing two factors is depicted in Figure 1. In this case, the rotation would be similar to the spinning of a propeller. The mathematics of factor analysis would allow the spin to stop anywhere, however, additional criteria have been developed to guide the rotation process. For example, a VARIMAX rotation retains the ability of the original solution to account for the initial correlations while maximizing the correlation between certain variables and one of the factors.

Factor analysis can be used to determine the dimensions that underlie a correlation matrix. For example, factor analysis has been used to describe the dimensions that underlie a battery of psychological tests, such as verbal and non-verbal factors in intelligence tests. Locating and describing these dimensions might be the main objective of a study or the researcher might wish to use the dimensions themselves as variables in a later analysis or study. The dimensions are defined and interpreted in terms of the variables that are highly associated with each factor. The correlation between a variable and a factor is called a factor loading. Unlike simple correlations, however, factor loadings usually have no formal test for significance[1]. The procedure for interpreting loadings is to define an arbitrary cutoff value (such as .3 or .4) and consider variables with loadings above that value to be salient to the definition of the factor.

To summarize, the purpose of Factor Analysis is to reduce a large set of variables to a small set of latent (unobserved) variables called factors. This is accomplished by

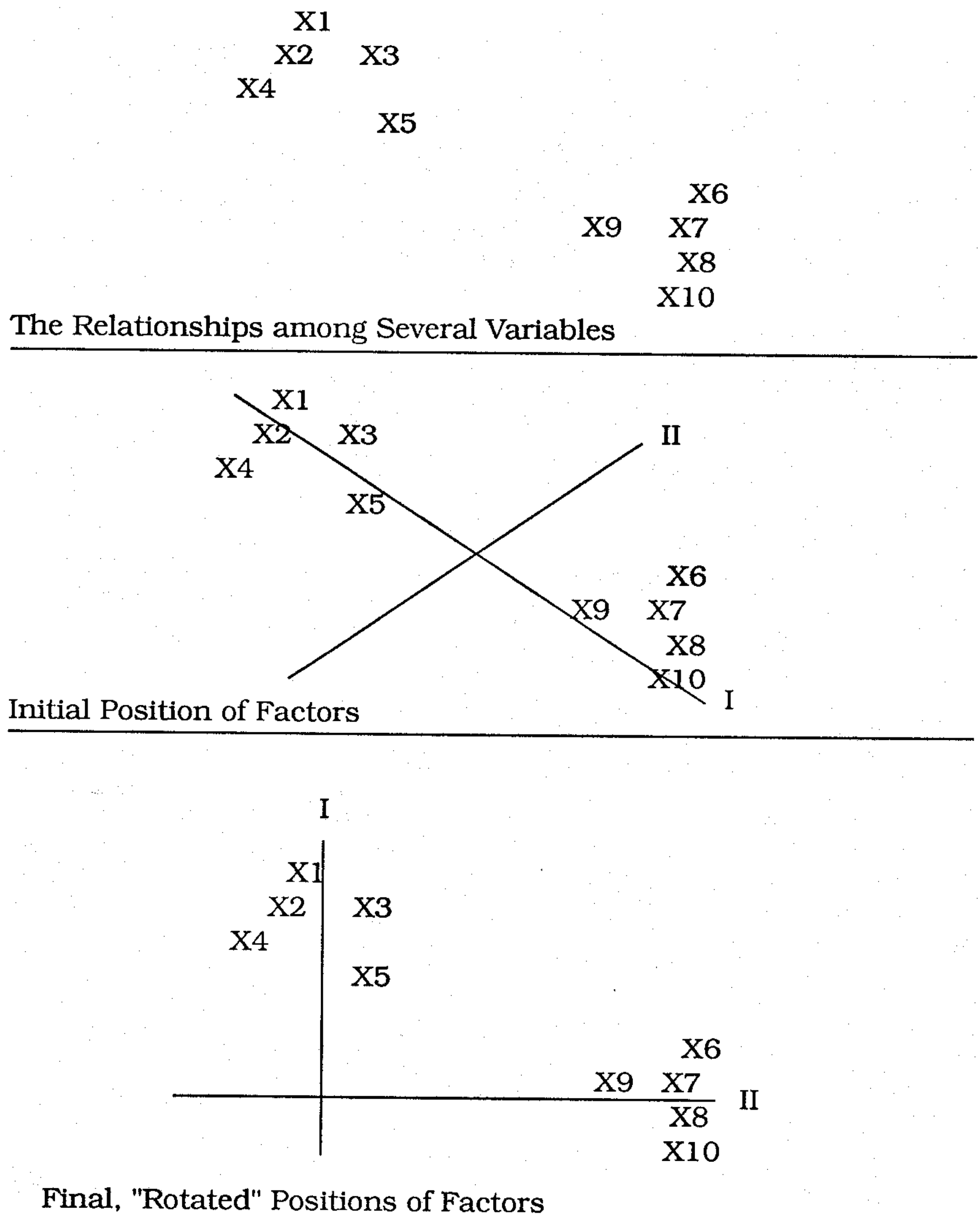

Figure 1: The Rotation Process

finding pockets of related variables that are intercorrelated. The set of variables that define a factor can be uncorrelated with another pocket of variables that are related to a second factor. In this case, independent dimensions that are common to one factor and not the other can be identified. While this introduction is extremely brief, the suggested reading section at the end of this paper lists texts and chapters that cover the topic in more detail.

Attention will now be directed to the issues listed above, namely estimating the required sample size, evaluating various approaches to the calculations required to extract factors, the ways of determining how many factors to extract, and the key problem of indeterminacy in the final solution.

Sample Size

There are several rules-of-thumb with which to determine the sample size (N) required for a factor analysis. N can be set at a definite number, such as 500, regardless of other circumstances (Comrey, 1973), can be related to the number of variables (Baggaley, 1986), the number of factors (Cattell 1966a), the difference between the number of variables and the number of factors (McDonald, 1985), or the ratio of variables to factors (such as the ratios suggested by: Harris (1967) 2:1; Thurstone (1947) 3:1; Gorsuch (1983) 5:1; Marascuilo and Levin (1983) 10:1). The key issue in all of these estimation procedures is to produce reliable, stable factor loadings. Since factor analysis begins with raw correlations, and loadings themselves are correlations, and the standard error of a correlation is defined in terms of N, it would appear that sample size should be a major influence.

It is clear that in any investigation, larger sample sizes will lead to more stable results. There is, however, a limit to the improvements that can be made by simply adding more subjects. Specifically, the limit is related to the reliability of the measures involved and how well each variable defines its associated factor. An analysis with several variables of poor reliability contains errors of measurement that strength of sampling cannot overcome. Alternatively, a hodgepodge of reliable variables with low intercorrelations could produce seemingly meaningful results based on small differences in correlation arising by chance.

In terms of the stability of solutions, Guadagnoli and Velicer (1988) have shown that sample size is almost unimportant under ideal conditions and is the most influential consideration under less favorable conditions. With a few factors that are well defined by several variables, a sample size of 50 was adequate to produce a stable solution. As the number of variables defining each factor diminishes and as the average of the salient loadings diminishes, sample size becomes increasingly important in determining the stability of a solution.

Actio and Anderson (1980) found that N and the number of variables did not interact to produce better results. Therefore, selecting more variables does not de facto require a larger sample size if the "extra" variables are not expected to add more factors. This is analogous to adding highly related items to a test which leads to better measurement, while adding unrelated items would make the test more complex and less consistent. In Figure 1, the addition of more variables close to the ones that already define the factors would not be expected to move the reference axes and therefore not alter the dimension that underlies them.

Therefore, sample size can be seen as less important as the psychometric quality of the investigation improves. To the extent that the following conditions apply, psychometric quality will improve: (1) a large number of variables are expected to correlate with each of the factors, (2) several of the variables are expected to load highly on only one factor, and (3) the overall communality in the matrix is high. When these conditions are less favorable, sample size may be the critical determinant of the stability of the obtained solution. It is recommended that the formula presented by Guadagnoli and Velicer (1988) be used to estimate the required sample size or to test the obtained sample size for adequacy.

Methods of Analysis

There have been several methods of factor analysis developed over the years: Principal Components, Principal Axis, Alpha, Image, and Maximum Likelihood are the most popular methods. With the exception of Maximum Likelihood approaches, much of the difference among the procedures is related to the assumptions about what underlies the diagonal elements of the correlation matrix. The largest difference occurs between the solutions that use a variable's correlation with itself (i.e. 1.0), called components models, versus common factor models that place a measure of communality in the diagonal. Communality refers to that portion of a variable that is shared with the other variables in the matrix. This can be measured by the highest simple correlation between one variable and another, a multiple correlation, squared multiple correlation, or the reliability of each variable. The most popular procedure for extracting factors has been principal components analysis (Actio & Anderson, 1980; Zwick and Velicer, 1982).

The components approach has often been contrasted with the common factors approach, sometimes vigorously. Cattell (1966a) labelled the components approach a "mathematical figment" and claimed that it must be rejected completely in favour of a common factor model. The common factor model sometimes produces mathematically impossible solutions (Steiger, 1979) and rests on conceptually shaky ground, making a components solution far more preferable (Zwick and Velicer, 1986). However, in very rare instances, the communality of a variable can approach 1.0 when it is measured with little error and has high correlations with other variables included in the analysis. Thus, in the limit, the components approach and the common factors approach would begin at the same point. However, such restrictive conditions are not required for the various methods to yield highly similar results.

When the solutions based on several methods are compared, this debate is rendered largely academic. Velicer (1977) analyzed nine well known real data sets comparing Principal Components, Image, and Maximum Likelihood methods. The factor structures produced were extremely similar and rotation did not affect this similarity. When differences were found, they occurred for the final factor which Velicer argues may have been over extracted. Cattell (1966a) notes that error variance can easily creep into the final few factors and the extraction methods handle error somewhat differently. Velicer (1977) concluded that Image analysis was an acceptable alternative to either Principal Components or Maximum Likelihood procedures.

Actio and Anderson (1980) investigated the effects of N, the ratio of variables to factors, and the communality on Principal Components, Image, and Alpha factor methods. In terms of sample size, they found that Image analysis with N = 50 lead to as good a solution as did Principal Components analysis with an N = 300. As the N's increased, however, the difference between the methods became negligible.

Velicer, Peacock, and Jackson (1982) compared Principal Components, Image, and Maximum Likelihood methods under a variety of conditions in simulation matrices created to possess a known set of factors. Overall, the pattern of loadings did not differ much among the methods. The number of errors was also very small. For the few errors that emerged, Maximum Likelihood tended to falsely declare that a loading is not "significant" while Principal Components and Image analysis were biased toward falsely declaring a loading to be "significant". The authors conclude that "... the three methods produce results that, for practical purposes, are indistinguishable" (p. 385).

Actio and Anderson (1980), Dziuban and Harris (1973), and Harris (1967) found differences across methods in certain cases. It should be noted that the latter two authors were working with an unknown number of factors and chose the eigenvalues-greater-than-one rule to determine the number of factors. This rule has been almost completely discredited (see below) and may have extracted too many factors leading to the obtained differences (Velicer, 1977). However, exploratory factor analytic research would be trivial if the number of factors was always known in advance, and therefore the results of these authors should be carefully considered. Harris (1967) recommends the following procedure to guard against potential errors based simply on extraction method:

1) Choose several (4 or 5) competing algorithms that can be applied to the data. Two

of the recommended alternatives are Alpha and Image analysis.
2) Perform the analyses with each of the chosen methods and compare the solutions. There are statistical methods available for estimating factor congruence (McDonald, 1985) or one could "eyeball" the solutions for differences in interpretation. Those analyses in which the investigator stretches the fit between solutions will be unlikely to pass the scholarly review process. Alternatively, rotation to a target matrix could be employed.
3) Retain only those factors that prove to be robust over methods since those specific to one method have not been replicated even within the study. Factor analysis should be evaluated on the basis of the replication of results (Cattell, 1966b; Guilford, 1974; McDonald, 1985).

Thus, practical differences among the methods of extraction are minor in most cases. For a sound study, the differences should be negligible. After all, the ultimate goal of each procedure is the same, the rationale of the procedures is more similar than different, the computations of the results are fairly similar, and those results are interpreted in much the same manner. An investigator should be cautious if different methods of factor extraction yield different interpretations of the factors. Zwick and Velicer (1986) believe that over-extraction may account for most of the discrepancies between components and common factor analyses and that observed differences are more likely to be a function of the number of factors retained than the method of extraction.

The recommended procedure would be to test the solution with several different analyses. Image analysis seems to be gaining acceptance and has received favorable evaluations, however, Principal Components and Principal Axis methods may be chosen for their widespread familiarity. Another analysis might be the Maximum Likelihood procedure. It has the advantage of a chi-square test that, indicates the number of factors to retain but tends to be overly sensitive to sample size. Should discrepancies arise in the rotated solutions, the analysis should be examined for problems with the sample size or the number of factors. If necessary, some variables might be removed and the analysis recalculated.

Number of Factors

The most important decision under the researcher's control in a factor analysis is the number of factors to retain (Zwick & Velicer, 1986). This decision can be the least objective and has the most direct consequences on the interpretation of the results. Too few factors will cause the solution to omit or collapse across relevant dimensions. Too many factors will lead to serious interpretation problems by obscuring the definitions of several factors simultaneously (Cattell, 1966a). The number of factors has a direct influence on the loadings, the estimates of communality, and the quality of the rotation process (Cliff and Hamburger, 1967). Not surprisingly, this issue has received considerable attention producing several rules by which to determine the "correct" number of factors.

The rules to be reviewed below are the "eigenvalues-greater-than-one" rule (ev > 1), the scree test, Bartlett's test, the Maximum Likelihood test, and two less well known procedures, Parallel analysis and the Minimum average partial (MAP) rule. With the exception of ev > 1, the various methods produce adequate results but no firm answer. Since replicating factors is the key issue, a convergence of methods would lead to the most confidence in the results.

EV > 1. Certainly the most common basis on which to make the decision is the size of the eigenvalues, which define the amount of variance accounted for by a factor. The ev > 1 rule simply states that factors with an eigenvalue greater than 1.0 should be retained while those with eigenvalues less than 1.0 should be dropped. Kaiser based this rule on three lines of argument that are criticized by Cliff (1988):
1) Since the variance of standardized variables is 1.0, a factor should account for more variance than a single variable. The reasoning is intuitively appealing, however, it is not a mathematical argument. Also, eigenvalues close to 1 are likely to invite only one large loading and are therefore highly problematic (see next section). Finally, Cattell (1966a)

argues that the variance of a factor can change after rotation, possibly coming above the arbitrary cutoff value.
2) The reliability of a factor with an eigenvalue less than 1 may not be positive. This argument does not hold because Kaiser employed a simplified version of a formula that was based on inappropriate assumptions and, therefore, was invalid. Cliff (1988) has shown that the reliability of a component does not depend on the size of the eigenvalue but rather on the weighted sum of the reliabilities of the observed variables. He concluded that "... the Kaiser rationale for retaining as many components as there are eigenvalues greater than one does not have any logical basis" (p. 276).
3) Ev > 1 was found to be a lower bound for the number of factors in a population (Guttman, 1954). Zwick and Velicer (1986) show that it can also be an upper bound. Finally, Cliff (1988) claims that matrices can be generated to systematically create an underestimation or overestimation of the number of factors using this rule.

Empirical evaluations of this rule have also been far less than kind. Cliff and Hamburger (1967) argue that ev > 1 applies only when dealing with a population matrix and that it tends to overestimate the correct number of factors in a sample matrix. Hakstain, Rogers and Cattell (1982) found that the rule was incorrect in over fifty percent of the cases in sample matrices. Both overestimates and underestimates were found. When compared to two other decision rules, ev > 1 missed the true number of factors to a greater extent in the population but to a lesser extent in sample matrices. Linn (1968) concluded that, in the presence of sampling error as found in any real data set, the ev > 1 rule could not be recommended.

Zwick and Velicer (1982) found that ev > 1 identified the correct number of factors in the population, but consistently and badly overestimated the number of factors with sample matrices from the same population. Also, this rule was found to seriously overestimate the number of factors and was heavily influenced by the number of variables, consistently indicating more factors as the number of variables increased. It was considered the worst of five decision rules tested. "Given the apparent functional relation of the number of components retained by the (ev > 1 rule) to the number of original variables and the repeated reports of the method's inaccuracy, we cannot recommend (this) rule for (Principal Components Analysis)" (p. 439).

Bartlett's Test. This statistic tests the hypothesis that all remaining eigenvalues are equal. This test has been found to be sensitive to sample size, especially as the number of variables approaches N (Zwick & Velicer, 1982). Gorsuch (1983) claims that it may lead to over-factoring when N is large. Zwick and Velicer (1986) found that Bartlett's test was the most variable of the five methods investigated and the least applicable to general usage.

Minimum Average Partial (MAP). The MAP rule takes the average of the squared partial correlations in the residualized matrix after each component is removed. This value reaches a minimum when all of the off-diagonal correlations have been reduced to zero. The value of the MAP tends to increase when the analysis begins to partial out unique variances, that is, when all of the common variance has been extracted (Zwick & Velicer, 1982). This procedure has been reported by Zwick and Velicer (1982, 1986) as one of the most accurate. It is the recommended rule when several variables are present and the investigator is not interested in minor components. Despite its promising results, this rule is not used extensively in the literature.

Comparisons to Random Solutions. Parallel Analysis compares the obtained factor matrix with one based on random variables with the same dimensions and sample size. Eigenvalues from the two matrices are compared retaining only those from the data matrix that exceed those from the random matrix. This is based on the reasoning that, by chance alone, half of the eigenvalues in a random matrix would exceed 1.0 while the other half would be less than 1.0 (Horn, 1965). Eigenvalues from a data matrix that do not exceed ones from a random matrix are assumed to be based on random variance.

Zwick and Velicer (1986) report that this is the most accurate of the methods they tested, estimating the known number of factors within one, 97% of the time even under relatively poor conditions. Although promising, this procedure's drawback lies in the generation of the random matrix which cannot be described as a routine procedure. Another problem lies in the fact that, beyond the first factor, the hypothesis being tested may not be appropriate[2]. It also has not been used frequently in the literature.

Another approach, suggested by Horn, is to add random variables to a data matrix. Factors defined by mostly random variables could be assumed to be random, error factors. This approach has been criticized for deliberately introducing error into the analysis and the number of random variables that are added could influence the obtained solution. The use of random variables has not been employed extensively in factor analytic investigations.

Scree. Cattell (1966b) developed the scree test based on the assumption that when the eigenvalues are nearly equal, their plot will show a flat slope when compared with the plot of the eigenvalues of the meaningful factors. When evaluating the scree plot, the researcher should eliminate the largest eigenvalue and look for definite breaks in the remaining values. A straight line should be drawn from the bottom up through the smaller eigenvalues. More than one "break" may make it necessary to draw more than one line. The first eigenvalue above the scree should be taken as the final factor. Although this may indicate error variance, the rotation should indicate the factors that are based on error.

The scree test is based on a "speculative proof" offered by Cattell (1966b) who notes that its value has been demonstrated empirically. Conceptually, the scree test and the Bartlett test share the same reasoning, that is, stop factoring when the eigenvalues reach zero, or close to it, at which time they will show a flatter slope (see Appendix A). Linn (1968) found that the scree test performed quite well when there were clear breaks in the plots. Matrices of high psychometric quality tend to produce clearly discernible breaks (Guttman, 1954). Hakstain, Rogers and Cattell (1982) found the scree test to be quite good when the factor pattern was simple. As the variables become more complex or as their communalities drop, the scree test tends to overestimate the number of factors as the breaks become less discernable. Cliff and Hamburger (1967) found a plot of the eigenvalues helpful for N's as low as 100. Zwick and Velicer (1982) found that the scree test had high inter-rater reliability and was consistently the most accurate test. Zwick and Velicer (1986) found that when in error, 90% of the time the scree test overestimated by one or two factors. In addition, they found that the scree test was less sensitive to N, number of variables, and the ratio of variables to factors.

The scree test is not mathematical and requires a researcher's judgement which may actually be an advantage because the mathematical solutions (such as ev > 1) offer a false sense of precision. By only approximating, and perhaps slightly overestimating, the number of factors, it forces the investigator to contemplate several potential solutions. A researcher who observes a clear break in the plot of the eigenvalues can be fairly confident in the estimation of the number of factors and, more times than not, will have converging evidence from other tests. If no clear break exists, the researcher may find that there is no clearly superior solution, that is, the number of factors to retain is not clearly indicated (Linn, 1968). The other methods reviewed here will provide an answer to the number of factors question, but the scree plot should be used routinely in conjunction with the mathematical methods.

A recommendation is difficult to make, save to say that the ev > 1 rule should never be the basis for a decision on the number of factors. Objective procedures such as the MAP and Parallel Analysis have yet to stand the rigors of repeated use because the ev > 1 rule has been so dominant. Convergence is the most desirable condition, however, when discrepancies exist, it is reasonable to test several solutions for the most theoretically meaningful solution, as opposed to the most mathematically preferable one. Replication in a future study would be required to confirm the decision. It is unfortunate that the most important decision must be made with this amount of uncertainty and that the dominant rule is woefully inadequate.

The Indeterminacy Problem

While the number of factors is the most important decision facing a factor analyst, the indeterminacy problem is the most limiting feature of the procedure. And there is little that a researcher can do other than to be cognizant of the problem. It would not be an overstatement to suggest that the issue of indeterminacy is one that is critical to the understanding of the limitations of factor analysis.

To say that a solution is determined means that it has one value. An indeterminate solution has an infinite number of possible values. Division by zero is one instance of an indeterminate solution. For example, the fraction 12/2 is determined since 2 can be subtracted from 12 exactly 6 times. The fraction 12/0 is not determined because one can subtract 0 from 12 an infinite number of times and no number of times is exactly correct.

The usual explanation of the purpose behind factor analysis is to reduce the number of variables in a given matrix by locating the essential, underlying (latent) constructs that influence the scores of a larger group of observed variables. Attention is then directed to the latent constructs as variables themselves. Another way to conceptualize this is to view factor analysis as essentially a regression equation with the criterion having the unfortunate property of being un-observable. With regression, a group of variables are combined in such a way as to optimally predict some criterion variable. Each subject has scores for the predictor variables, a score on the criterion, and an estimated score on the criterion. The estimated criterion score can be compared to (i.e. subtracted from) the actual score as a measure of fit between the regression model and the actual data.

Indeterminacy in factor analysis arises from not having scores on the underlying variables. Therefore, we must make some assumptions about the nature of those variables. Most factor analytic algorithms extract factors such that as much variance as possible is included in the first factor. As much remaining variance as possible is then allocated to factor 2, and so on until either all of the variance has been accounted for or the researcher decides to stop extracting factors.

Usually, the factors that emerge from this "initial solution" are not pleasing theoretically and would be difficult to interpret. But as noted above, we are not forced to retain this as the final solution. Figure 1 shows two pockets of interrelated variables and the lines represent the underlying factors that will be used to summarize the variables. The proximity of an individual variable to a given line corresponds to the correlation between that variable and the underlying factor. It is possible to adjust, through rotation, the location of the dimensions without altering either the relations among the variables or the relations among the underlying factors. By "spinning" the cross in Figure 1, however, we do alter the distances between the variables and the factors: some variables move away from the axes and others move closer. This retains the ability of the dimensions to describe the factors; it just changes the numbers used to do it. These distances, as measured by correlations, are the factor loadings that most researchers use to assess the nature of the dimension.

The key problem then is that there are an infinite number of possible rotated solutions and therefore an infinite number of possible values that the loadings could take. Therefore, how can we trust our interpretation of the factor based on its loadings if those loadings may be high or low based on a rather arbitrary decision? (Guttman, 1954; McDonald, 1985). Thirty years of research into intelligence have been questioned on this basis alone (see Guilford, 1974; Horn & Knapp, 1974)! The implications of this argument are unsettling and represent the most serious weakness of factor analysis (Guilford, 1974). Steiger (1979) notes that this issue was discovered in the 1920's and its implications fully developed during the 1930's. The procedure in the period from 1940's - 1970's was to ignore indeterminacy, assuming to have resolved the debate by indicating simple structure (Thurstone, 1947). Simple structure has a number of rules associated with it but essentially attempts to find variables that are not associated with given factors, thus reducing the number of factors that would be required to describe that variable. The simplest case would be when only one of the extracted factors is needed to describe a variable and the rest are not related to that variable (i.e. have zero loadings). This serves as a criterion to which a rotation can be directed and, because it is an objective criterion, only one solution emerges. Algorithms such as VARIMAX are based, more or less, on the notion of simple structure and define their rotational criteria unambiguously. Therefore, the VARIMAX rotated solution looks quite determinate, for a particular matrix it gives one set of loadings and gives the same set every time.

There is a preferable solution to the problem that requires matrices of higher psychometric quality. The logic is as follows: As a test becomes infinitely large, it becomes almost perfectly reliable as problems with individual items on the test balance each other. The same reasoning applies to factor analysis. As a factor becomes better defined by an infinitely large battery of tests, it becomes almost perfectly determined, errors of measurement even themselves out, and the position of its reference vector becomes stable, even if more subjects or variables are added. The implication is that with a very well defined factor, the factors can be located exactly. As the factor becomes more indeterminate, the arbitrary nature of any factor loading increases rapidly.

At this point, a major difference between the common factors and the principal components approach can be seen. Indeterminacy is a significant problem for the common factor model and, at times, "impossible" solutions will arise. It is a problem inherent in the calculations and may be traced back to the research question. It therefore must be considered in advance of any experimental design. Component models do not suffer from this problem, at least within the computations. If no factors are truncated, Principal Components will reproduce the correlation matrix exactly, every time. However, a components analysis that did not truncate the "trivial" factors would be little more than a mathematical exercise. To the extent that a number of trivial factors are eliminated, we conceptually, if not mathematically, approach the same type of indeterminacy problem.

As noted above, many factor analyses are performed to find pockets of variables that can be combined to form a well defined latent variable. To this end, scores can be generated that correspond to a particular subject's score for each of the latent variables. These scores, because they are based on factor loadings, also can be indeterminant. A potential solution to the indeterminacy problem is not to deal with factor scores at all. Rather, scale scores are computed by simply aggregating the variables that load significantly on each factor. In this case, the factor analysis serves to suggest variables that may be inter-related. It is then a researcher's decision, based on the theory related to those variables, to select the appropriate ones for aggregation. The aggregation stands on its own merits and must be validated by conventional methods, with factor analysis providing supporting evidence for a given combination of variables.

Before leaving the topic, there is another level at which indeterminacy can pose a problem for any type of factor analysis. When a factor is defined by only one or two salient loadings, the rest being zero loadings, they can take almost any value (McDonald, 1985). Because of the loose mathematical restrictions associated with rotation, it takes, at the bare minimum, three variables with high loadings to define a factor. Any analysis that presents a factor with less than three substantial loadings on it would do well to reduce the number of factors and recalculate the analysis. In no case should such a factor be trusted.

It is recommended that when designing a study, the investigator postulate a target factor matrix. Included here would be an indication of the number of factors expected, the relative size and signs of the salient loadings, and the location of zero loadings. Each factor should have four or more variables defining it with one or two "marker" variables expected to load very highly and exclusively on a factor. If a factor is expected to emerge with only one or two associated variables, eliminate it from the study, do not include it in the Factor Analysis, or supplement that factor with other related variables. Although this does not solve the indeterminacy problem, it does help the researcher ensure that the factors are as well defined as relevant theory allows.

Recommendations for Reporting Results

There are several pieces of information required before a factor analysis can be properly evaluated. Articles that do not report sufficient information are not in the interest of science. "If we are given only, say, the rotated factor matrix, it is perhaps better to regard the study as never having been carried out at all" (McDonald, 1985, p. 77). Factor analysis is an extremely powerful multivariate statistic (there being statistics, damned statistics, and multivariate statistics).

The onus must be on the investigator to provide information sufficient to evaluate the analysis. The first criterion is to establish the adequacy of the data matrix for factor analysis. Dziuban and Shirky (1974) list three methods for determining the adequacy of a matrix. The recommended statistic is the Kaiser-Meyer-Olkin measure of sampling adequacy. This statistic ranges from zero to one and assesses the degree to which the variables belong to a family. Values less than 0.5 mean that the matrix is not suited for factor analysis, values between 0.5 and 0.7 are marginal and values greater than 0.7 are adequate. If possible, the original correlation matrix should be presented.

After establishing that a factor analysis is appropriate, it must be described in unambiguous terms. Throughout this paper the term factor analysis has referred to a collection of procedures and therefore is not a recommended phrase. "Principal Components Analysis", "Alpha", "Image", and "Maximum Likelihood" methods of extraction have quite standard uses and are not ambiguous terms. When a Principal Axis solution is generated, the author should be careful to indicate the manner in which the communalities were estimated.

Since the original solution is almost always rotated, usually to a VARIMAX criterion, only the interpreted solution needs to be presented, since equal amounts of information are contained in any rotation. It is imperative that this solution be presented in full, especially in cases when the correlation matrix has not been shown. The annoying and potentially misleading practice of blotting out the "non-significant" loadings must be avoided especially since the definition of "significant loading" is quite arbitrary. Those loadings of significance to the investigation, which is the only appropriate definition of significance in this case, can be underlined, starred, or bolded for emphasis.

It has been argued above that several solutions, based on different algorithms, should be generated for any analysis. This serves to replicate the findings and lends credibility to the interpretations. When possible, all of the rotated matrices should be presented or, in their place, some measure of the degree of similarity between the various solutions can be reported. The solution presented should be augmented by the eigenvalues, especially if space is available for a scree plot.

Conclusion

Factor analysis is a powerful multivariate statistic and the options available may make it more difficult to interpret the results. This article has addressed a few of the major issues but several others remain. Several excellent references are available to those interested in a more detailed discussion.

Glossary

<u>Alpha</u> - a technique for factor analysis that uses diagonal values equal to the reliability of the variable in question.
<u>Communality</u> - the portion of the variance in a variable that is accounted for by the other variables included in the correlation matrix. The remaining portion is unique to the variable and has nothing in common with the rest of the variables.
<u>Eigenvalues</u> - The percentage of variance accounted for by a factor. High eigenvalues indicate that the factor accounts for a substantial portion of the variance.
<u>Image</u> - a technique for factor analysis that uses diagonal values equal to the multiple correlation between a particular variable and all of the remaining variables in a matrix.
<u>Loading</u> - the correlation between a variable

and a factor.

Principal Components Analysis - a technique for factor analysis that uses diagonal values equal to one (1.0).

Rotation - a procedure used to purify the factors that capitalizes on the fact that an infinite number of patterns can equally reproduce a correlation matrix.

Simple Structure - A criterion for rotation that, in simple terms, increases the chances that each variable will correlate highly with only one factor and near zero with the rest.

Varimax - A rotation method that seeks to maximize that variance of a factor by generating as many high loadings on the factors as possible.

Weights - coefficients, ranging from 0 to 1, that are multiplied by the variables when aggregating them. Low coefficients remove a variable from the equation while high coefficients emphasize a variable.

APPENDIX A

Example Analysis

The following represents an attempt to display the above recommendations with respect to an actual Factor Analysis. The data come from 95 students and represents scores on several anxiety scales. A previous investigation found two factors in a smaller set of scales.

The intercorrelations among the eleven anxiety scales were factor analyzed (see Figure 2). Five different approaches (Principal Components, Principal Axis, Alpha, Image, and Maximum Likelihood methods) to the analysis were taken in order to ensure that the method of extraction did not influence the interpretation of the results. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.71, indicating that the matrix was appropriate for Factor Analysis. According to a scree test (see Figure 3), the chi-square test from the Maximum Likelihood solution (chi-square = 30.6, df = 25, $p < .20$), and the theory associated with the variables, three factors with eigenvalues greater than one were retained. Analyses with two and four factors were attempted but were judged to be inadequate since they violated the tests for the number of factors and did not provide theoretically meaningful results. The initial solutions accounted for 65% of the variance in the correlation matrix.

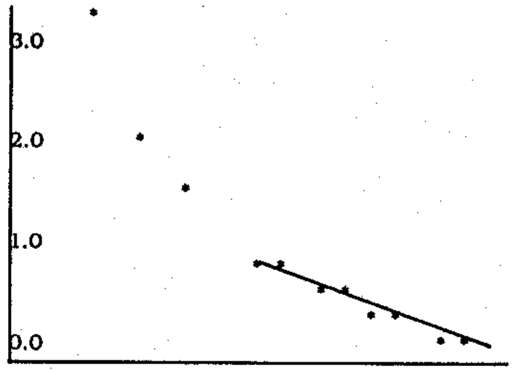

Figure 3: The Scree Test

Since all eleven measures are scales of anxiety, loadings above 0.4 will be interpreted as representative of the factor. Inspection of the rotated factor matrices (see Figure 4) show that the interpretations based on each of the methods are similar. For the Principal Components solution, Factor 1 is defined by variables related to communication apprehension (PRCA), audience anxiety (AUD), social evaluation anxiety (EVAL), and interpersonal anxiety (INTER). Each of these scales assess anxiety in social or communication situations and will be labelled *Social Anxiety*. The second factor was defined by the three administrations of a state anxiety scale and will be called *Experiment Anxiety*. Finally, the four scales on Factor 3 are French class anxiety (FRCA), debilitating French test anxiety (DEBIL), facilitating French test anxiety (FACIL), and French use anxiety (FRUSE). All

	FRCA	FRUSE	DEBIL	FACIL	STATE1	STATE2	STATE3	PRCA	AUD	INTER	EVAL
FRCA	1.00										
FRUSE	.47	1.00									
DEBIL	.49	.37	1.00								
FACIL	-.37	-.12	-.40	1.00							
STATE1	.16	.26	.15	-.09	1.00						
STATE2	.13	.23	.34	-.15	.42	1.00					
STATE3	.05	.20	.31	-.03	.40	.79	1.00				
PRCA	.29	.23	.17	-.13	.23	.17	.08	1.00			
AUD	.18	.16	.08	-.01	.15	.19	.12	.75	1.00		
INTER	.09	.10	.05	.02	.25	.28	.16	.52	.53	1.00	
EVAL	.09	.15	.26	-.14	.27	.29	.16	.65	.51	.37	1.00

Figure 2: The Correlation Matrix

Principal Components

PRCA	**.89**	.01	.21
AUD	**.86**	.04	.05
EVAL	**.72**	.18	.14
INTER	**.72**	.21	-.07
STATE3	.01	**.91**	.06
STATE2	.14	**.89**	.14
STATE1	.22	**.61**	.12
FRCA	.14	-.02	**.82**
DEBIL	.03	.28	**.75**
FACIL	.01	.01	**-.67**
FRUSE	.14	.21	**.60**

Image

PRCA	**.78**	.03	.21
AUD	**.73**	.07	.09
EVAL	**.62**	.17	.15
INTER	**.55**	.18	.01
STATE3	.05	**.77**	.10
STATE2	.15	**.76**	.17
STATE1	.20	**.40**	.15
FRCA	.14	.01	**.60**
DEBIL	.06	.26	**.55**
FRUSE	.13	.16	**.44**
FACIL	-.02	-.04	**-.42**

Principal Axis

PRCA	**.92**	.00	.21
AUD	**.80**	.06	.06
EVAL	**.63**	.18	.14
INTER	**.59**	.19	-.01
STATE2	.15	**.88**	.16
STATE3	.04	**.87**	.08
STATE1	.22	**.43**	.15
FRCA	.13	-.02	**.82**
DEBIL	.04	.28	**.75**
FRUSE	.14	.01	**-.67**
FACIL	.01	-.03	**-.47**

Alpha

PRCA	**.92**	.00	.22
AUD	**.81**	.05	.04
EVAL	**.61**	.18	.16
INTER	**.60**	.21	-.03
STATE3	.02	**.87**	.06
STATE2	.15	**.84**	.17
STATE1	.23	**.46**	.16
FRCA	.12	.00	**.81**
DEBIL	.04	.23	**.73**
FACIL	.01	-.02	**-.44**
FRUSE	.14	.21	**.42**

Maximum Likelihood

PRCA	**.93**	.00	.21
AUD	**.78**	.06	.07
EVAL	**.66**	.18	.11
INTER	**.57**	.19	.00
STATE2	.15	**.91**	.06
STATE3	.06	**.84**	.19
STATE1	.21	**.41**	.16
FRCA	.12	-.01	**.80**
DEBIL	.04	.26	**.63**
FRUSE	.13	.15	**.53**
FACIL	-.02	.05	**-.47**

Figure 4: Five Extraction Methods

four of these variables are associated with the anxiety of being evaluated in French class or in speaking in French and will be referred to as *French Anxiety*.

These results differ from previous analyses using similar scales. The difference lies in the second factor which would be assumed to be unstable since state anxiety is a transient emotion. Thus the two factors of Social Anxiety and French Anxiety can be considered as replicating the stable factors found in the previous studies. The presence of Experiment Anxiety seem to be specific to the particular experimental conditions associated with the current study.

Scales representing each of these factors were created by combining the variables that showed substantial loadings, as listed above. These aggregated variables can now be correlated with other variables.

This is only one example of reporting a factor analysis. The figures that accompany the text present some of the detail that is required but would make for tedious reading.

ACKNOWLEDGEMENTS

Many thanks to Dr. R.C. Gardner, Dr. R. Harshman, and Dr. P. A. Vernon for their contribution to the development of my interest in statistics, factor analysis in particular. Preparation of this manuscript was financially assisted, in part, by SSHRCC Doctoral Fellowship # 452-89-1394.

NOTES

1. One of the approaches to calculating a factor analysis, the Maximum Likelihood model, does have a test for significance because it is based on a set of assumptions that are fairly different from the other, more common approaches.

2. This argument itself is beyond the scope of this paper. It should also be noted that, in spite of its theoretical shortcomings, empirical tests have shown Parallel Analysis to be useful in determining the number of factors to retain (see Cota, Longman, Holden, & Fekken, 1990).

LITERATURE CITED

Acito, F. & Anderson, R. D. 1980. A Monte Carlo comparison of factor analytic methods. *Journal of Marketing Research*, *17*, pp. 228-236.

Browne, M. W. 1968. A comparison of factor analytic techniques. *Psychometricka*, *33*, pp. 267-344.

Baggaley, A. R. 1986. Heterogeneous correlations and estimates of required sample size in Factor Analysis. *Multivariate Experimental Clinical Research*, *8*, pp. 161-163.

Cattell, R. B. 1966a. The meaning and strategic use of factor analysis. *Handbook of Multivariate experimental psychology*. R. B. Cattell (ed.). Rand-McNally, Chicago, pp. 174-243.

Cattell, R. B. 1966b. The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, pp. 145-276.

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

Cliff, N. 1988. The eigenvalues greater than one rule and the reliability of components. *Psychological Bulletin*, *103*, pp. 276-279.

Comrey, A. L. 1973. *A first course in factor analysis*. Academic Press, New York.

Cota, A. A., Longman, R. S., Holden, R. R., Fekken, G. S. 1990. Choosing the number of factors to retain: A guide to parallel analysis. *Canadian Psychology*, *31*, p. 367.

Dziuban, C. D. & Harris, C. W. 1973. On the extraction of components and the applicability of the factor model. *American Educational Research Journal*, *10*, pp. 93-99.

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

Gorsuch, R. L. 1983. *Factor Analysis*. Erlbaum, Hillsdale, New Jersey.

Guadnoli, E., & Velicer, W. F. 1988. Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*, pp. 265-275.

Guilford, J. P. 1974. Rotation problems in factor analysis. *Psychological Bulletin*, *81*, pp. 498-501.

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

Hakstian, A. R., Rogers, W. D., & Cattell, R. B. 1982. The behavior of numbers factors rules with simulated data. *Multivariate Behavioral Research*, *17*, pp. 193-219.

Harmann, H. H. 1976. *Modern Factor Analysis* (2nd ed.). University of Chicago Press, Chicago.

Harris, C. W. 1967. On factors and factor scores. *Psychometrika*, *32*, pp. 363-379.

Horn, J. L. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, p. 179.

Horn, J. L. & Knapp, J. R. 1974. Thirty wrongs do not make a right: reply to Guilford. *Psychological Bulletin*, *81*, pp. 502-504.

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

Kaiser, H. F. 1961. A note of Guttman's lower bound for the number of common factors. *British Journal of Mathematical and Statistical Psychology, 14*, p. 1.

Linn, R. L. 1968. A monte carlo approach to the number of factors problem. *Psychometricka, 33*, pp. 37-71.

Marascuilo, L. A. & Levin, J. R. 1983. *Multivariate Statistics in the Social Sciences: A Researcher's Guide.* Brooks/Cole. Monterey, Calif.

McDonald, R. P. 1985. *Factor Analysis and Related Methods.* Erlbaum, Hillsdale, New Jersey.

McDonald, R. P. & Mulaik, S. A. 1979. Determinacy of common factors: A non-technical review. *Psychological Bulletin, 86*, pp. 297-306.

Mulaik, S. A. 1966. Inferring the communality of a variable in a universe of variables. *Psychological Bulletin, 66*, pp. 119-124.

Steiger, J. H. 1979. Factor indeterminacy in the 1930's and the 1970's: Some interesting parallels. *Psychometricka, 44*, pp. 355-360.

Thurstone, L. L. 1947. *Multiple factor analysis.* University of Chicago Press, Chicago.

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

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

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

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

RECOMMENDED READING

The following sources provide a concise, readable introduction to Factor Analysis:

Fruchter, B. 1954. *Introduction to Factor Analysis.* D. Van Nostrand, Toronto.

Kim, J-O, & Meueller, C. W. 1978. *Factor analysis: What it is and how to do it.* Sage, Beverly Hills.

Kim, J-O, & Meueller, C. W. 1978. *Factor analysis: statistical and practical issues.* Sage, Beverly Hills.

Stevens, J. 1986. Chapter 11, Principal Components. *Applied Multivariate Statistics for the Social Sciences.* Erlbaum, Hillsdale, New Jersey.