UNDERSTANDING STATISTICS, *2*(1), 13–43

TEACHING ARTICLES

Repairing Tom Swift's Electric Factor Analysis Machine

Kristopher J. Preacher and Robert C. MacCallum
Department of Psychology
The Ohio State University

Proper use of exploratory factor analysis (EFA) requires the researcher to make a series of careful decisions. Despite attempts by Floyd and Widaman (1995), Fabrigar, Wegener, MacCallum, and Strahan (1999), and others to elucidate critical issues involved in these decisions, examples of questionable use of EFA are still common in the applied factor analysis literature. Poor decisions regarding the model to be used, the criteria used to decide how many factors to retain, and the rotation method can have drastic consequences for the quality and meaningfulness of factor analytic results. One commonly used approach—principal components analysis, retention of components with eigenvalues greater than 1.0, and varimax rotation of these components—is shown to have potentially serious negative consequences. In addition, choosing arbitrary thresholds for factor loadings to be considered large, using single indicators for factors, and violating the linearity assumptions underlying EFA can have negative consequences for interpretation of results. It is demonstrated that, when decisions are carefully made, EFA can yield unambiguous and meaningful results.

exploratory factor analysis, principal components, EFA, PCA

Exploratory factor analysis (EFA) is a method of discovering the number and nature of latent variables that explain the variation and covariation in a set of measured variables. Because of their many useful applications, EFA methods have enjoyed widespread use in psychological research literature over the last several decades. In the process of conducting EFA, several important decisions need to be

Requests for reprints should be sent to Kristopher J. Preacher, 142 Townshend Hall, 1885 Neil Avenue Mall, The Ohio State University, Columbus, OH 43210–1222. E-mail: preacher.2@osu.edu

made. Three of the most important decisions concern which model to use (common factor analysis vs. principal components analysis[1]), the number of factors to retain, and the rotation method to be employed. The options available for each decision are not interchangeable or equally defensible or effective. Benefits of good decisions, based on sound statistical technique, solid theory, and good judgment, include substantively meaningful and easily interpretable results that have valid implications for theory or application. Consequences of poor choices, on the other hand, include obtaining invalid or distorted results that may confuse the researcher or mislead readers.

Although the applied factor analysis literature contains many superb examples of careful analysis, there are also many studies that are undoubtedly subject to the negative consequences just described due to questionable decisions made in the process of conducting analyses. Of particular concern is the fairly routine use of a variation of EFA wherein the researcher uses principal components analysis (PCA), retains components with eigenvalues greater than 1.0, and uses varimax rotation, a bundle of procedures affectionately termed "Little Jiffy" by some of its proponents and practitioners (Kaiser, 1970).

Cautions about potential negative consequences of this approach have been raised frequently in the literature (notably, Fabrigar, Wegener, MacCallum, & Strahan, 1999; Floyd & Widaman, 1995; Ford, MacCallum, & Tait, 1986; Lee & Comrey, 1979; Widaman, 1993). However, these cautions seem to have had rather little impact on methodological choices made in many applications of EFA. Articles published in recent years in respected journals (e.g., Beidel, Turner, & Morris, 1995; Bell-Dolan & Allan, 1998; Brown, Schulberg, & Madonia, 1995; Collinsworth, Strom, & Strom, 1996; Copeland, Brandon, & Quinn, 1995; Dunn, Ryan, & Paolo, 1994; Dyce, 1996; Enns & Reddon, 1998; Flowers & Algozzine, 2000; Gass, Demsky, & Martin, 1998; Kier & Buras, 1999; Kwan, 2000; Lawrence et al., 1998; Osman, Barrios, Aukes, & Osman, 1995; Shiarella, McCarthy, & Tucker, 2000; Yanico & Lu, 2000) continue to follow the Little Jiffy approach in whole or in part, undoubtedly yielding some potentially misleading factor analytic results. Repeated use of less than optimal methods reinforces such use in the future.

In an effort to curtail this trend, we will illustrate by example how poor choices regarding factor analysis techniques can lead to erroneous and uninterpretable results. In addition, we will present an analysis that will demonstrate the benefits of making appropriate decisions. Our objective is to convince researchers to avoid the Little Jiffy approach to factor analysis in favor of more appropriate methods.

[1]Use of the word *model* usually implies a falsifiable group of hypotheses describing relationships among variables. We use the term here in a broader sense, namely an explanatory framework leading to understanding, without necessarily the implication of falsifiability. By using the word *model* we simply mean to put principal components analysis and exploratory factor analysis on the same footing so that they may be meaningfully compared.

THE ELECTRIC FACTOR ANALYSIS MACHINE

In 1967 an article entitled "Derivation of Theory by Means of Factor Analysis or Tom Swift and His Electric Factor Analysis Machine" (Armstrong, 1967)[2] was published. The intended point of this article was to warn social scientists against placing any faith in EFA when the intent is to develop theory from data. Armstrong presented an example using artificial data with known underlying factors and then performed an analysis to recover those factors. He reasoned that because he knew what the recovered factor structure *should* have been, he could assess the utility of EFA by evaluating the degree to which the method recovered that known factor structure. We make use of the Armstrong example here because his factor analysis methods correspond closely to choices still commonly made in applied factor analysis in psychological literature. Thus, we use the Armstrong article as a surrogate for a great many published applications of EFA, and the issues we address in this context are relevant to many existing articles as well as to the ongoing use of EFA in psychological research. Although the substantive nature of Armstrong's example may be of little interest to most readers, we urge readers to view the example as a proxy characterized by many of the same elements and issues inherent in empirical studies in which EFA is used. Generally, data are obtained from a sample of observations on a number of correlated variables, and the objective is to identify and interpret a small number of underlying constructs. Such is the case in Armstrong's example, and the conclusions drawn here apply more generally to a wide range of empirical studies.

Armstrong presented the reader with a hypothetical scenario. In his example, a metals company received a mysterious shipment of 63 box-shaped, metallic objects of varying sizes. Tom Swift was the company's operations researcher. It was Swift's responsibility to develop a short, but complete, classification scheme for these mysterious objects, so he measured each of the boxes on 11 dimensions:

(a) thickness
(b) width
(c) length
(d) volume
(e) density
(f) weight
(g) total surface area
(h) cross-sectional area
(i) total edge length
(j) length of internal diagonal
(k) cost per pound

[2]The Armstrong (1967) article was available, at the time of this writing, at http://www-marketing.wharton.upenn.edu/forecast/papers.html.

The measurements were all made independently. In other words, Swift obtained volume by some means other than by multiplying thickness, width, and length together, and weight was obtained by some means other than calculating the product of density and volume.

Swift suspected (correctly) that there was overlap between some of these dimensions. He decided to investigate the structure of the relationships among these dimensions using factor analysis, so he conducted a PCA, retained as many components as there were eigenvalues greater than 1.0, and rotated his solution using varimax rotation. As noted earlier, this set of techniques is still widely used in applied factor analysis research. Armstrong noted at this point that all of the available information relied on 5 of the original 11 variables, because all of the 11 measured variables were functions of thickness, width, length, density, and cost per pound (functional definitions of Swift's 11 variables are shown in Table 1). This implies that an EFA, properly conducted, should yield five factors corresponding to the 5 basic variables. Swift's analysis, however, produced only three underlying components that he called *compactness*,[3] *intensity*, and *shortness*, the first of which he had some difficulty identifying because the variables that loaded highly on it did not seem to have much in common conceptually. These components accounted for 90.7% of the observed variance in the original 11 variables. Armstrong's reported rotated loadings for these three components are presented in Table 2 (note that only loadings greater than or equal to 0.7 were reported). Armstrong pointed out that we should not get very excited about a model that explains 90.7% of the variability using only three factors, given that we know that the 11 variables are functions of only five dimensions.

Because Swift had trouble interpreting his three components, Armstrong suggested that if Swift had relaxed the restriction that only components with eigenvalues greater than 1.0 be retained, he could retain and rotate four or even five components. This step might have seemed reasonable to Armstrong's readers because they happened to know that five factors underlay Swift's data even though Swift did not know that. However, because Swift had no prior theory concerning underlying factors, he may not have considered this alternative.

Nevertheless, Swift repeated his analysis, this time retaining four components. On examination of the loadings (presented in Table 3), he termed these components *thickness*, *intensity*, *length*, and *width*. The problem with this solution, according to Armstrong, was that the model still did not distinguish between density and cost per pound, both of which had high loadings on a single component even though they seemed conceptually independent. Dissatisfied with his results, Swift sought to more fully identify the underlying dimensions by introducing nine additional measured variables:

---

[3]Because some variables and components share labels, component and factor labels will be italicized.

TABLE 1
Functional Definitions of Tom Swift's Original 11 Variables

Dimension	*Derivation*
Thickness	x
Width	y
Length	z
Volume	xyz
Density	d
Weight	$xyzd$
Total surface area	$2(xy + xz + yz)$
Cross-sectional area	yz
Total edge length	$4(x + y + z)$
Internal diagonal length	$(x^2 + y^2 + z^2)^2$
Cost per pound	c

TABLE 2
Armstrong's (1967) Three-Factor Solution for Original 11 Variables

	Dimension		
	Compactness	*Intensity*	*Shortness*
Thickness	0.94	—	—
Width	0.74	—	—
Length	—	—	0.95
Volume	0.93	—	—
Density	—	0.96	—
Weight	0.72	—	—
Total surface area	0.86	—	—
Cross-sectional area	—	—	0.74
Total edge length	0.70	—	—
Internal diagonal length	—	—	0.88
Cost per pound	—	0.92	—

Note. An em dash (—) = a loading not reported by Armstrong.

(l) average tensile strength
(m) hardness
(n) melting point
(o) resistivity
(p) reflectivity
(q) boiling point
(r) specific heat at 20°C
(s) Young's modulus
(t) molecular weight

TABLE 3
Armstrong's (1967) Loadings on Four-Rotated Components for Original 11 Variables

	Dimension			
	Thickness	*Intensity*	*Length*	*Width*
Thickness	0.96	—	—	—
Width	—	—	—	0.90
Length	—	—	0.99	—
Volume	0.85	—	—	—
Density	—	0.96	—	—
Weight	0.71	—	—	—
Total surface area	0.73	—	—	—
Cross-sectional area	—	—	—	0.72
Total edge length[a]	—	—	—	—
Internal diagonal length	—	—	0.84	—
Cost per pound	—	0.93	—	—

Note. An em dash (—) = a loading not reported by Armstrong.
[a]Total edge length was excluded from Swift's findings presumably because it had no loadings greater than 0.7.

Swift conducted a components analysis on the full set of 20 variables using the same methods he used earlier. This time, however, he retained five components that explained 90% of the variance. (Even though it was not explicitly stated, it can be assumed that Swift retained as many components as there were eigenvalues greater than 1.0.) He interpreted the varimax-rotated components as representing *impressiveness*, *cohesiveness*, *intensity*, *transference*, and *length*. The reported loadings on these components for 20 variables are presented in Table 4. Swift then noticed that, in some cases, items loading highly on the same components seemed to have little to do with each other conceptually. For example, density and cost per pound still loaded onto the same component when they were clearly (to him) independent concepts.

An overall evaluation of Tom Swift's results would lead most readers to the conclusion that his analysis failed to uncover the known factors underlying the observed variables. Armstrong concluded that because EFA was employed with no a priori theory, Swift had no criteria by which to judge his results. According to Armstrong, factor analysis would be better suited to evaluate a prior theory rather than to generate a new one (i.e., it would have been better to use factor analysis in a confirmatory way rather than in a purely exploratory way). In other words, Swift may have explained 90% of the variability in his data at the cost of retaining a bogus and nonsensical collection of factors. Armstrong ended the article by saying that conclusions based on factor analytic techniques may be unsupported or misleading.

ASSESSING THE DAMAGE

Whereas Armstrong wanted to argue that even a properly conducted factor analysis can lead to the wrong conclusions, he instead inadvertently demonstrated how poor choices regarding factor analytic techniques can lead to wrong conclusions.

There are at least six major methodological shortcomings in Armstrong's example that call into question his conclusions. Unfortunately, many modern applications of EFA exhibit several, and sometimes all, of these same shortcomings (Fabrigar et al., 1999; Floyd & Widaman, 1995). These shortcomings are the facts that Swift (a) confused EFA with PCA, (b) retained components with eigenvalues greater than 1.0, (c) used orthogonal varimax rotation, (d) used an arbitrary cutoff for high factor loadings, (e) used only one indicator for one of his components, and (f) used several variables that violated the assumption that measured variables (MVs) are linearly related to latent variables (LVs). Each of these deficiencies will be addressed in turn.

TABLE 4
Armstrong's (1967) Loadings on Five-Rotated Components for 20 Variables

	Dimension				
	Impressiveness	*Cohesiveness*	*Intensity*	*Transference*	*Length*
Thickness	0.92	—	—	—	—
Width	0.80	—	—	—	—
Length	—	—	—	—	0.92
Volume	0.98	—	—	—	—
Density	—	—	0.96	—	—
Weight	0.76	—	—	—	—
Total surface area	0.95	—	—	—	—
Cross-sectional area	0.74	—	—	—	—
Total edge length	0.82	—	—	—	—
Internal diagonal length	—	—	—	—	0.76
Cost per pond	—	—	0.71	—	—
Tensile strength	—	0.97	—	—	—
Hardness	—	0.93	—	—	—
Melting point	—	0.91	—	—	—
Resistivity	—	—	—	–0.93	—
Reflectivity	—	—	—	0.91	—
Boiling point	—	0.70	—	—	—
Specific heat	—	—	–0.88	—	—
Young's modulus	—	0.93	—	—	—
Molecular weight	—	—	0.87	—	—

Note. An em dash (—) = a loading not reported by Armstrong.

EFA Versus PCA

EFA is a method of identifying unobservable LVs that account for the (co)variances among MVs. In the common factor model, variance can be partitioned into *common variance* (variance accounted for by common factors) and *unique variance* (variance not accounted for by common factors). Unique variance can be further subdivided into *specific* and *error* components, representing sources of systematic variance specific to individual MVs and random error of measurement, respectively. Common and unique sources of variance are estimated separately in factor analysis, explicitly recognizing the presence of error. Common factors are LVs that account for common variance only as well as for covariances among MVs.

The utility of PCA, on the other hand, lies in data reduction. PCA yields observable composite variables (components), which account for a mixture of common and unique sources of variance (including random error). The distinction between common and unique variance is not recognized in PCA, and no attempt is made to separate unique variance from the factors being extracted. Thus, components in PCA are conceptually and mathematically quite different from factors in EFA.

A problem in the Armstrong (1967) article, as well as in much modern applied factor analysis literature, is the interchangeable use of the terms *factor analysis* and *principal components analysis*. Because of the difference between factors and components just explained, these techniques are not the same. PCA and EFA may seem superficially similar, but they are very different. Problems can arise when one attempts to use components analysis as a substitute or approximation for factor analysis. The fact that Swift wanted to describe his boxes on as few underlying dimensions as possible sounds at first like simple data reduction, but he wanted to account for correlations among MVs and to lend the resultant dimensions a substantive interpretation. PCA does not explicitly model error variance, which renders substantive interpretation of components problematic. This is a problem that was recognized over 60 years ago (Cureton, 1939; Thurstone, 1935; Wilson & Worcester, 1939; Wolfle, 1940), but misunderstandings of the significance of this basic difference between PCA and EFA still persist in the literature.[4]

Like many modern researchers using PCA, Armstrong confused the terminology of EFA and PCA. In one respect, the reader is led to believe that Swift used PCA because he refers to "principal components" and to "percent variance accounted for" as a measure of fit. On the other hand, Armstrong also refers to "fac-

[4]Other important differences between the two methods do not derive directly from the differences in their intended purposes, but may nevertheless have a bearing on a given analysis. In factor analysis, for example, models are testable, whereas in principal components analysis (PCA) they are not. PCA typically overestimates loadings and underestimates correlations between factors (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Floyd & Widaman, 1995; Widaman, 1993).

tors," which exist in the domain of factor analysis, and attempts to lend substantive meaning to them. Researchers should clarify the goals of their studies, which in turn will dictate which approach, PCA or EFA, will be more appropriate. An investigator wishing to determine linear composites of MVs that retain as much of the variance in the MVs as possible, or to find components that explain as much variance as possible, should use PCA. An investigator wishing to identify interpretable constructs that explain correlations among MVs as well as possible should use factor analysis. In EFA, a factor's success is not determined by how much variance it explains because the model is not intended to explain optimal amounts of variance. A factor's success is gauged by how well it helps the researcher understand the sources of common variation underlying observed data.

The Number of Factors to Retain

One of the most important decisions in factor analysis is that of how many factors to retain. The criteria used to make this decision depend on the EFA technique employed. If the researcher uses the noniterative principal factors technique, communalities[5] are estimated in a single step, most often using the squared multiple correlation coefficients (SMCs).[6] The iterative principal factors technique requires the number of factors to be specified a priori and uses some approach to estimate communalities initially (e.g., by first computing SMCs), but then enters an iterative procedure. In each iteration, new estimates of the communalities are obtained from the factor loading matrix derived from the sample correlation matrix. Those new communality estimates are inserted into the diagonal of the correlation matrix and a new factor loading matrix is computed. The process continues until convergence, defined as the point when the difference between two consecutive sets of communalities is below some specified criterion. A third common technique is maximum likelihood factor analysis (MLFA), wherein optimal estimates of factor loadings and unique variances are obtained so as to maximize the multivariate normal likelihood function, to maximize a function summarizing the similarity between observed and model-implied covariances. It should be emphasized that these three methods represent different ways of fitting the same model—the common factor model—to data.

Criteria used to determine the number of factors to retain fall into two broad categories depending on the EFA technique employed. If iterative or noniterative principal factors techniques are employed, the researcher will typically make the

[5]Communality represents the proportion of a variable's total variance that is "common," that is, explained by common factors.

[6]Squared multiple correlation coefficients are obtained by using the formula $h_{jj} = 1 - (P_{jj}^{-1})^{-1}$, where P_{jj}^{-1} represents the diagonal elements of the inverse of the matrix.

determination based on the eigenvalues of the *reduced* sample correlation matrix (the correlation matrix with communalities rather than 1.0s in the diagonal). Unfortunately, there is no single, fail-safe criterion to use for this decision, so researchers usually rely on various psychometric criteria and rules of thumb. Two such criteria still hold sway in social sciences literature.

The first criterion—that used by Swift—is to retain as many factors (or components) as there are eigenvalues of the *unreduced* sample correlation matrix greater than 1.0. This criterion is known variously as the Kaiser criterion, the Kaiser–Guttman rule, the eigenvalue-one criterion, truncated principal components, or the K1 rule. The theory behind the criterion is founded on Guttman's (1954) development of the "weakest lower bound" for the number of factors. If the common factor model holds exactly in the population, then the number of eigenvalues of the unreduced population correlation matrix that are greater than 1.0 will be a lower bound for the number of factors. Another way to understand this criterion is to recognize that MVs are typically standardized to have unit variance. Thus, components with eigenvalues greater than 1.0 are said to account for at least as much variability as can be explained by a single MV. Those components with eigenvalues below 1.0 account for less variability than does a single MV (Floyd & Widaman, 1995) and thus usually will be of little interest to the researcher.

Our objection is not to the validity of the weakest lower bound, but to its application in empirical studies. Although the justifications for the Kaiser–Guttman rule are theoretically interesting, use of the rule in practice is problematic for several reasons. First, Guttman's proof regarding the weakest lower bound applies to the population correlation matrix and assumes that the model holds exactly in the population with *m* factors. In practice, of course, the population correlation matrix is not available and the model will not hold exactly. Application of the rule to a sample correlation matrix under conditions of imperfect model fit represents circumstances under which the theoretical foundation of the rule is no longer applicable. Second, the Kaiser criterion is appropriately applied to eigenvalues of the unreduced correlation matrix rather than to those of the reduced correlation matrix. In practice, the criterion is often misapplied to eigenvalues of a reduced correlation matrix. Third, Gorsuch (1983) noted that many researchers interpret the Kaiser criterion as the actual number of factors to retain rather than as a lower bound for the number of factors. In addition, other researchers have found that the criterion underestimates (Cattell & Vogelmann, 1977; Cliff, 1988; Humphreys, 1964) or overestimates (Browne, 1968; Cattell & Vogelmann, 1977; Horn, 1965; Lee & Comrey, 1979; Linn, 1968; Revelle & Rocklin, 1979; Yeomans & Golder, 1982; Zwick & Velicer, 1982) the number of factors that should be retained. It has also been demonstrated that the number of factors suggested by the Kaiser criterion is dependent on the number of variables (Gorsuch, 1983; Yeomans & Golder, 1982; Zwick & Velicer, 1982), the reliability of the factors (Cliff, 1988, 1992), or on the MV-to-factor ratio and the range of communalities (Tucker, Koopman, & Linn, 1969). Thus, the general conclusion is that there is little justification for using the Kaiser criterion to decide how many factors to retain. Swift chose to use the Kaiser criterion even though it is probably a poor approach to take. There is little theoretical evidence to support it, ample evidence to the contrary, and better alternatives that were ignored.

Another popular criterion (actually a rule of thumb) is to retain as many factors as there are eigenvalues that fall before the last large drop on a scree plot, which is a scatter plot of eigenvalues plotted against their ranks in terms of magnitude. This procedure is known as the *subjective scree test* (Gorsuch, 1983); a more objective version comparing simple regression slopes for clusters of eigenvalues is known as the Cattell–Nelson–Gorsuch (CNG) scree test (Cattell, 1966; Gorsuch, 1983). The scree test may be applied to either a reduced or unreduced correlation matrix (Gorsuch, 1983). An illustration of a scree plot is provided in Figure 1, in which the eigenvalues that fall before the last large drop are separated by a line from those that fall after the drop. Several studies have found the scree test to result in an accurate determination of the number of factors most of the time (Cattell & Vogelmann, 1977; Tzeng, 1992).

A method that was developed at approximately the same time as the Tom Swift article was written is called parallel analysis (Horn, 1965; Humphreys & Ilgen, 1969). This procedure involves comparing a scree plot based on the reduced correlation matrix (with SMCs on the diagonal rather than communalities) to one derived from random data, marking the point at which the two plots cross, and counting the number of eigenvalues on the original scree plot that lie above the intersection point. The logic is that useful components or factors should account not for more variance than one MV (as with the Kaiser criterion), but for more variance than could be expected by chance. The use of this method is facilitated by an equation provided by Montanelli and Humphreys (1976) that accurately estimates the expected values of the leading eigenvalues of a reduced correlation matrix

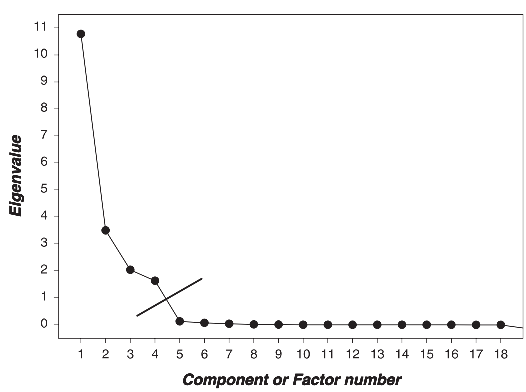

FIGURE 1 A scree plot for a correlation matrix of 19 measured variables. The data were obtained from a version of the Tucker matrix (discussed later) with additional variables. There are four eigenvalues before the last big drop, indicating that four factors should be retained.

(with SMCs on the diagonal) of random data given sample size and number of MVs. An illustration of parallel analysis is provided in Figure 2, in which the number of factors retained would be equal to the number of eigenvalues that lie above the point marked with an *X*. This method has been found to exhibit fairly good accuracy as well (Humphreys & Montanelli, 1975; Richman, as cited in Zwick & Velicer, 1986), but it should nevertheless be used with caution (Turner, 1998).

The second broad class of methods used to determine the number of factors to retain requires the use of maximum likelihood (ML) parameter estimation. Several measures of fit are associated with MLFA, including the likelihood-ratio statistic associated with the test of exact fit (a test of the null hypothesis that the common factor model with a specified number of factors holds exactly in the population) and the Tucker–Lewis index (Tucker & Lewis, 1973). Because factor analysis is a special case of structural equation modeling (SEM), the wide array of fit measures that have been developed under maximum likelihood estimation in SEM can be adapted for use in addressing the number-of-factors problem in MLFA. Users can obtain a sequence of MLFA solutions for a range of numbers of factors and then assess the fit of these models using SEM-type fit measures, choosing the number of factors that provides optimal fit to the data without overfitting. Browne and Cudeck (1993) illustrated this approach using fit measures such as the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), expected cross-validation index (ECVI; Cudeck & Browne, 1983), and test of exact fit. If Swift had known about maximum likelihood estimation and the accompanying benefits, he could have performed a series of analyses in which he retained a range of number of factors. Examination of the associated information about model fit could help him determine the best solution, and thus an appropriate number of factors.

Swift did not use ML estimation, which means that he had to choose from among the Kaiser criterion, the scree test, and parallel analysis. An appropriate approach would have been to examine the scree plot for the number of eigenvalues

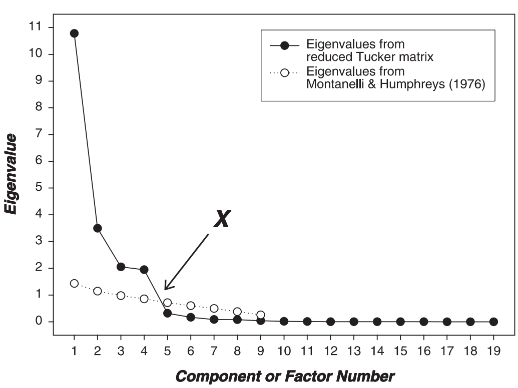

FIGURE 2 Superimposed scree plots for correlation matrices of random and real (expanded Tucker matrix) data sets including 19 measured variables, $N = 63$. Four eigenvalues fall before the point at which the two plots cross, which indicates that four factors should be retained. Note that the eigenvalues produced by Montanelli and Humphreys's (1976) procedure are real numbers only up to the ninth eigenvalue.

that fall before the last large drop and to conduct a parallel analysis. Both methods have shown fairly good performance in the past. The determination of the appropriate number of factors to retain always has a subjective element, but when the scree test and parallel analysis are combined with judgment based on familiarity with the relevant literature and the variables being measured, an informed decision can be made. When results provided by different methods are inconsistent or unclear as to the number of factors to retain, it is recommended that the researcher proceed with rotation involving different numbers of factors. A judgment about the number of factors to retain can be based on the interpretability of the resulting solutions. Swift's approach of using a single rule of thumb has a strong likelihood of producing a poor decision.

Oblique Versus Orthogonal Rotation

A third important decision in factor analysis involves rotation. One of Thurstone's (1935, 1947) major contributions to factor analysis methodology was the recognition that factor solutions should be rotated to reflect what he called *simple structure* to be interpreted meaningfully.[7] Simply put, given one factor loading matrix, there are an infinite number of factor loading matrices that could account for the variances and covariances among the MVs equally as well. Rotation methods are designed to find an easily interpretable solution from among this infinitely large set of alternatives by finding a solution that exhibits the best simple structure. Simple structure, according to Thurstone, is a way to describe a factor solution characterized by high loadings for distinct (non-overlapping) subsets of MVs and low loadings otherwise.

Many rotation methods have been developed over the years, some proving more successful than others. *Orthogonal* rotation methods restrict the factors to be uncorrelated. *Oblique* methods make no such restriction, allowing correlated factors. One of the most often used orthogonal rotation methods is *varimax* (Kaiser, 1958), which Swift presumably used in his analysis (Armstrong stated that Swift used orthogonal rotation; it is assumed that varimax was the specific method employed because it was by far the most popular method at the time).

For reasons not made clear by Armstrong (1967), Swift suspected that the factors underlying the 11 MVs would be statistically independent of each other. Consequently, Swift used orthogonal rotation. Whereas this may sound like a reasonable approach, it is a strategy that is very difficult to justify in most cases. We will show that it is a clearly inappropriate strategy in the present case.

The thickness, width, and length of the boxes in Armstrong's artificial study were determined in the following manner (see footnote 2 in Armstrong, 1967):

[7]Factor loadings can be thought of as vector coordinates in multidimensional space. Hence, the terminology used to describe rotation methods relies heavily on vector geometry.

(a) "Random integers from 1 to 4 were selected to represent width and thickness with the additional provision that the width ≥ thickness."
(b) "A random integer from 1 to 6 was selected to represent length with the provision that length ≥ width."
(c) "A number of the additional variables are merely obvious combinations of length, width, and thickness."

There is no fundamental problem with generating dimensions in this fashion. However, doing so introduces substantial correlations among the dimensions, and in turn among the underlying factors. Using Armstrong's method with a sample size of 10,000, similar length, width, and thickness data were generated and Pearson correlation coefficients were computed. The three dimensions were moderately to highly intercorrelated ($r_{xy} = .60$, $r_{xz} = .22$, $r_{yz} = .37$). Of course, Tom had no way of knowing that these dimensions were not independent of each other, but his ignorance is simply an additional argument in favor of using oblique rotation rather than orthogonal rotation. Using orthogonal rotation introduced a fundamental contradiction between his method and the underlying structure of the data.

In general, if the researcher does not know how the factors are related to each other, there is no reason to assume that they are completely independent. It is almost always safer to assume that there is not perfect independence, and to use oblique rotation instead of orthogonal rotation.[8] Moreover, if optimal simple structure is exhibited by orthogonal factors, an obliquely rotated factor solution will resemble an orthogonal one anyway (Floyd & Widaman, 1995), so nothing is lost by using oblique rotation. Oblique rotation offers the further advantage of allowing estimation of factor correlations, which is surely a more informative approach than assuming that the factors are completely independent. It is true that there was no completely satisfactory analytic method of oblique rotation at the time the Tom Swift article was written, but Armstrong could have (and probably should have) performed oblique graphical rotation manually. If Swift were plying his trade today, he would have ready access to several analytic methods of oblique rotation, including direct quartimin (Jennrich & Sampson, 1966) and other members of the Crawford–Ferguson family of oblique rotation procedures (Crawford & Ferguson, 1970). For a thorough review of available factor rotation methods, see Browne (2001).

[8]It is interesting to note that Armstrong's (1967) example strongly resembles Thurstone's (1935, 1947) box problem, which also involved dimensions of boxes and an attempt to recover an interpretable factor solution. Interestingly, Thurstone performed his factor analysis correctly, over 30 years before Armstrong wrote the Tom Swift article. As justification for using oblique rotation, Thurstone pointed out that a tall box is more likely than a short box to be thick and wide; thus, the basic dimensions may be correlated to some extent. Use of orthogonal rotation in the box problem and similar designs such as that constructed by Armstrong is simply unjustified.

Retaining Factor Loadings Greater Than an Arbitrary Threshold

It is generally advisable to report factor loadings for all variables on all factors (Floyd & Widaman, 1995). For Swift's first analysis, he chose to interpret factor loadings greater than 0.7 as "large." He reported only loadings that exceeded this arbitrary threshold. Following such rules of thumb is not often the best idea, especially when better alternatives exist and when there is no logical basis for the rule of thumb. As a researcher trying to establish the relationships of the latent factors to the observed variables, Swift should have been interested in the complete pattern of loadings, including low loadings and mid-range loadings, not simply the ones arbitrarily defined as large because they were above 0.7. Moreover, depending on the field of study, large may mean around 0.3 or 0.4. It would have been more informative for the reader to have seen the other loadings, but Swift did not report them. It is not clear why he chose the threshold he did, but it *is* clear that no absolute cutoff point should have been defined.

In addition, because factor loadings will vary due to sampling error, it is unreasonable to assume that loadings that are high in a single sample are correspondingly high in other samples or in the population. Therefore, there is no reasonable basis for reporting only those sample loadings that lie beyond a certain threshold. Recent developments in factor analysis have led to methods for estimating standard errors of factor loadings (Browne, Cudeck, Tateneni, & Mels, 1998; Tateneni, 1998), which allow researchers to establish confidence intervals and conduct significance tests for factor loadings.

In general, it is advisable to report all obtained factor loadings so that readers can make judgments for themselves regarding which loadings are high and which are low. Additionally, because the necessary technology is now widely available (Browne et al., 1998), it is also advisable to report the standard errors of rotated loadings so that readers may gain a sense of the precision of the estimated loadings.

Using a Single Indicator for a Latent Variable

In factor analysis, a latent variable that influences only one indicator is not a common factor; it is a specific factor. Because common factors are defined as influencing at least two manifest variables, there must be at least two (and preferably more) indicators per factor. Otherwise, the latent variable merely accounts for a portion of unique variance, that variability which is not accounted for by common factors. One of Swift's components—*cost per pound*—is found to have only one indicator (the MV cost per pound), even after increasing the set of MVs to 20.

Not only should Swift have included more indicators for *cost per pound*, but also for *density*, as it had only two indicators. In fact, as we shall see in the results of a later analysis, the cost per pound MV correlates highly with the density and weight MVs and loads highly on the same factor as density and weight. Swift would have been safer had he considered all three MVs to be indicators of the same common factor (call it a *density/cost per pound* factor). Because Armstrong did not supply his method for determining cost per pound, we do not know for certain how it is related to density, but the evidence suggests that the two MVs are related to the same common factor. Fabrigar et al. (1999) recommended that at least four indicators be included per factor. In empirical studies using *exploratory* factor analysis, the researcher may not always have enough information to ensure that each factor has an adequate number of indicators. However, there should be some rough idea of what the MVs are intended to represent, which should allow the researcher to make an educated decision about which (and how many) MVs should be included in the analysis.

Violation of the Linearity Assumption

One of the assumptions underlying the common factor model is that the MVs are linearly dependent on the LVs (not that MVs ought to be linearly related to other MVs). Of course, in the real world few variables are exactly linearly related to other variables, but it is hoped that a linear model captures the most important aspects of the relationship between LVs and MVs.

A potentially major problem with Armstrong's analysis involves the issue of linearity. The Tom Swift data include three MVs that reflect the three basic dimensional factors of *thickness*, *width*, and *length*. However, the data also include many variables that are nonlinear functions of these three basic dimensional variables. It can be assumed that these MVs are related to the basic dimensional factors in a nonlinear way. For example, the thickness and width MVs are linearly related to the *thickness* and *width* LVs. Given that assumption, it is also safe to conclude that the volume MV (*xyz*; see Table 1) is not linearly related to any of the three factors for which it is assumed to serve as an indicator. Similar violations of the linearity assumption exist for most of the other MVs included in Swift's analysis. Thus, it was not reasonable of Armstrong to criticize the methodology for Swift's failure to exactly recover five dimensions when there were inherent incompatibilities between the model (linearity assumed) and the data (nonlinear). Violation of the linearity assumption does not completely compromise the chances of recovering an interpretable factor solution (e.g., Thurstone, 1940), but it may have contributed to the recovery of poor or uninterpretable factor solutions in the past, including that reported by Tom Swift.

Summary of Tom Swift's Assessment

Armstrong's article suffers from several misconceptions. First, even though Armstrong wanted us to believe that the underlying factor structure was clear, it clearly was not. Of the five latent factors proposed by Armstrong, three were highly intercorrelated. Of the remaining two, one had only one indicator and the other had only two. In addition, there were strong nonlinear relationships between factors and most of the MVs, constituting a violation of a basic assumption of the common factor model. Second, Swift's analysis followed the common pattern of choosing PCA, retaining factors with eigenvalues greater than 1.0, and using orthogonal varimax rotation. In addition, he chose an arbitrary cutoff for high loadings. These misconceptions and uses of dubious techniques describe a recipe for ambiguous results.

It should be pointed out that the Little Jiffy method does not necessarily always produce distorted or ambiguous results, nor does it always fail to recover strong underlying factors. In fact, there are certain circumstances under which Little Jiffy might work quite well. For example, when the communalities of the MVs are all high (and therefore unique variances are all low), PCA and EFA yield similar results. The Kaiser criterion will at least occasionally yield a correct estimate of the number of factors to retain. In addition, when the underlying LVs are nearly or completely uncorrelated, orthogonal rotation can yield undistorted, interpretable results. However, this combination of circumstances is probably rather rare in practice and, in any case, the researcher cannot know a priori if they hold.

In the end, Armstrong defeated the purpose of his own article. Rather than demonstrate the general failure of factor analysis, he illustrated the necessity of using the correct methods. A series of poor choices in technique led to confusing results and erroneous conclusions that were unfairly generalized to all studies employing factor analysis.

REPAIRING THE MACHINE

We repeated Tom Swift's "factor analysis" using an analogous data set. Swift's data set was unavailable. Following methods analogous to those used by Armstrong (1967), Ledyard Tucker (personal communication, 1970) generated a population correlation matrix for Swift's original 11 variables. That correlation matrix, hereafter referred to as the Tucker matrix, was used to investigate the relevance of the choice of factor analysis techniques to the results obtained by Swift.

Swift's findings were first verified by conducting a PCA on the Tucker matrix, retaining components for eigenvalues greater than 1.0, and using varimax rotation. All factor loadings from the present analysis, as well as the reported loadings from Swift's analysis, are presented in Table 5.

Using the Tucker matrix, three principal components were retained, just as in Swift's analysis. Factor loadings from the Tucker data were only slightly different from those reported by Swift. If 0.65 were used as the criterion for "high loading" instead of 0.7 (Swift's criterion), almost exactly the same pattern of high loadings is obtained. Inspection of the full set of loadings obtained from the Little Jiffy analysis of the Tucker matrix reveals that Swift's choice of 0.7 as the cutoff point was a questionable decision because it resulted in ignoring some substantial loadings. As seen in Table 5, some of the other loadings were very close to, or above, Swift's cutoff (e.g., those for cross-sectional area on the first component and surface area and edge length on the second).

Swift's four-component solution was also replicated using the Tucker data. Loadings are presented in Table 6. Using Swift's criterion of 0.7 as a high loading, exactly the same loading pattern was obtained. Again, several loadings approached the arbitrary threshold, such as those for edge length on the first, second, and third components and weight on the fourth, calling into question the use of an arbitrary cutoff for loadings.

Performing the Analysis Correctly

The Tucker matrix was analyzed again using more justifiable methods. These techniques included the use of factor analysis rather than PCA, using multiple methods to determine the proper number of factors to retain, and using oblique rather than orthogonal rotation. Ordinary least squares (OLS; equivalent to iterative principal factors) parameter estimation was used rather than maximum likelihood (ML) estimation, which may be preferable. ML estimation will not work in this situation because the population matrix supplied by Tucker is singular. The minimization function associated with ML estimation involves computing the logarithm of the determinant of the correlation matrix, a value that is undefined in the case of a singular matrix.

TABLE 5
Loadings on Three Principal Components Versus Tom Swift's Loadings

	Rotated Factor Pattern					
	Component 1		*Component 2*		*Component 3*	
	Tucker	*Swift*	*Tucker*	*Swift*	*Tucker*	*Swift*
Thickness	0.95	0.94	0.09	—	–0.02	—
Width	0.65	0.74	0.54	—	–0.02	—
Length	0.06	—	0.95	0.95	0.03	—
Volume	0.88	0.93	0.42	—	0.00	—
Density	0.03	—	–0.02	—	0.94	0.96
Weight	0.71	0.72	0.29	—	0.51	—
Total surface area	0.80	0.86	0.59	—	0.00	—
Cross-sectional area	0.51	—	0.81	0.74	0.00	—
Total edge length	0.65	0.70	0.76	—	0.00	—
Internal diagonal length	0.46	—	0.88	0.88	0.01	—
Cost per pound	–0.01	—	–0.01	—	0.91	0.92

Note. An em dash (—) = a loading not reported by Armstrong.

TABLE 6
Loadings on Four-Principal Components Versus Tom Swift's Loadings

	Rotated Factor Pattern							
	Component 1		*Component 2*		*Component 3*		*Component 4*	
	Tucker	*Swift*	*Tucker*	*Swift*	*Tucker*	*Swift*	*Tucker*	*Swift*
Thickness	0.96	0.96	0.09	—	0.17	—	–0.03	—
Width	0.38	—	0.24	—	0.89	0.90	0.01	—
Length	0.11	—	0.99	0.99	0.13	—	0.01	—
Volume	0.84	0.85	0.35	—	0.38	—	–0.01	—
Density	0.04	—	–0.01	—	–0.03	—	0.94	0.96
Weight	0.71	0.71	0.25	—	0.25	—	0.50	—
Total surface area	0.72	0.73	0.48	—	0.50	—	–0.01	—
Cross-sectional area	0.34	—	0.61	—	0.70	0.72	0.01	—
Total edge length	0.57	—	0.65	—	0.49	—	0.00	—
Internal diagonal length	0.41	—	0.81	0.84	0.41	—	0.00	—
Cost per pound	–0.01	—	–0.01	—	0.01	—	0.91	0.93

Note. An em dash (—) = a loading not reported by Armstrong.

Results using eigenvalues from both unreduced and reduced[9] versions of the Tucker matrix are presented in Table 7. Eigenvalues from the unreduced matrix are the appropriate values to examine when the Kaiser criterion is used. Inspection of Table 7 reveals that use of Kaiser's criterion would erroneously suggest that three factors be retained. The results of a scree test and parallel analysis using the eigenvalues from the reduced correlation matrix may be inferred from the plot in Figure 3. Random eigenvalues for use in parallel analysis were generated using an equation provided by Montanelli and Humphreys (1976). Both the scree test and parallel analysis indicate that four factors should be retained.

In summary, the various criteria for number of factors are not completely consistent in this case, but they do agree closely. The Kaiser criterion suggests that at least three factors should be retained, whereas the scree test and parallel analysis both suggest four. Although there is some convergence of these criteria on four factors as an appropriate number, a final resolution would depend on the interpretability of rotated solutions.

[9]Because a singular matrix has no inverse, a small constant (.001) was added to each diagonal element before inversion (Tucker, personal communication). In addition to using the method described earlier, squared multiple correlation coefficients (SMCs) were also estimated by using a procedure developed by Tucker, Cooper, and Meredith (1972). The SMCs found by both methods were identical to two decimal places. Given that the SMCs are so similar, it makes little difference which are used as starting values in iterative least squares factor analysis (Widaman & Herringer, 1985), so those from the first method were used.

TABLE 7
Eigenvalues of Unreduced and Reduced Versions of the Tucker Matrix

	Unreduced	*Reduced*
1	6.8664	6.8457
2	1.9765	1.6541
3	1.1035	1.0778
4	0.5283	0.5168
5	0.3054	0.0860
6	0.1436	0.0082
7	0.0685	–0.0013
8	0.0144	–0.0023
9	0.0024	–0.0031
10	0.0012	–0.0037
11	0.0008	–0.1333

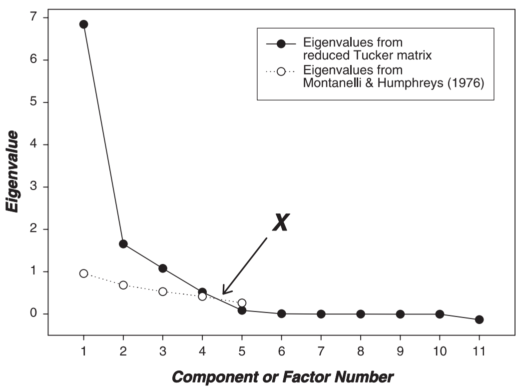

FIGURE 3 Superimposed scree plots for eigenvalues of random and real data sets including 11 measured variables, $N = 63$. There are four eigenvalues before the last big drop, indicating that four factors should be retained. It is evident by using parallel analysis that four factors should be retained.

As explained earlier, we consider Swift's choice of orthogonal varimax rotation to be inadvisable because it assumes the factors to be uncorrelated. The alternative is oblique rotation to simple structure, for which there are many widely accepted methods available. We used direct quartimin rotation. Factor loadings for the rotated four-factor solution are presented in Table 8. Judging by the highest factor loadings for the basic five variables (thickness, width, length, density, and cost per pound), the factors can be interpreted as *thickness*, *width*, *length*, and *density/cost per pound*, respectively. Notice that the thickness, width, and length MVs have high loadings only on factors of the same name. Volume, surface area, edge length, and diagonal length have non-zero loadings on the three dimensional factors, and cross-sectional area has non-zero loadings on factors corresponding to the two dimensions that contributed to it (*width* and *length*). The cost per pound, density, and weight MVs all have non-zero loadings on the last factor (*density/cost per pound*).

TABLE 8
Direct Quartimin Rotated Loadings for the Four-Factor Solution

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*
Thickness (*x*)	1.06	–0.10	–0.10	–0.04
Width (*y*)	0.03	1.02	–0.12	0.00
Length (*z*)	–0.05	–0.08	1.06	0.00
Volume	0.81	0.17	0.13	–0.03
Density (*d*)	–0.04	–0.02	–0.01	1.00
Weight	0.60	0.11	0.09	0.47
Total surface area	0.60	0.33	0.24	–0.02
Cross-sectional area	0.00	0.74	0.36	0.00
Total edge length	0.39	0.35	0.46	–0.02
Internal diagonal length	0.21	0.26	0.69	–0.01
Cost per pound (*c*)	–0.03	–0.01	–0.01	0.75

Note. In obliquely rotated solutions, factor loadings greater than 1.0 are not only admissible, but also routinely encountered.

The only unexpected finding is that weight does not load particularly highly on width and length.[10] Nevertheless, it must be acknowledged that the direct quartimin rotated solution is much more interpretable than the varimax rotated solution reported by Armstrong. Furthermore, this solution corresponds well to the known structure underlying these data.

A five-factor solution was also subjected to rotation. The five-factor solution introduced a weak factor with modest (.2 to .4) loadings for the volume, weight, surface area, and cross-sectional area measured variables, each of which represents some multiplicative function of length, width, and thickness. Retention of this factor did not enhance interpretability. Furthermore, the presence of a factor with only weak loadings can be considered a sign of overfactoring. Based on the interpretability of these rotated solutions, it was decided that it would be most appropriate to retain four factors.

One of the advantages associated with using oblique rotation is that factor correlations are estimated. They are presented in Table 9. As might be guessed from the dependency among the three primary dimensional variables demonstrated earlier, the *thickness*, *width*, and *length* factors are moderately to highly intercorrelated. The *density/cost per pound* factor, however, is not highly correlated with the other three. These correlations among factors reinforce the validity of the choice of an oblique rotation method, and also reveal why solutions obtained using orthogonal rotation were somewhat distorted.

[10]However, because of the way in which the measured variables were generated, weight is essentially a third-order function of thickness multiplied by density. Thus, it is the variable most in violation of the linearity assumption.

AN EMPIRICAL EXAMPLE

We consider it important to demonstrate that the consequences of poor choices outlined in the Tom Swift example also apply in practice, when the object is to understand the number and nature of factors underlying real psychological data. For this purpose, we selected the "24 abilities" data set supplied by Holzinger and Swineford (1939). This data set is widely available, and has been used frequently as an example (e.g., Browne, 2001; Gorsuch, 1983; Harman, 1967). The data set consists of 24 mental ability tests administered to $N = 145$ students in an Illinois elementary school circa 1937. For convenience we used the corrected Holzinger and Swineford (1939) correlation matrix supplied by Gorsuch (1983, p. 100). We compared the success of the Little Jiffy approach and the recommended approach in determining the factor structure underlying these 24 tests.

Determining the Number of Factors

First, eigenvalues of the unreduced correlation matrix were obtained (see Table 10). Using the Kaiser criterion, five factors should be retained because there are five eigenvalues greater than 1.0. The recommended approach entails using several criteria to converge on an appropriate number of factors. The scree plot of eigenvalues of the reduced correlation matrix (see Figure 4) shows a marked drop from the fourth eigenvalue to the fifth eigenvalue, with no appreciable difference between the fifth and sixth eigenvalues, indicating that four factors ought to be retained. Parallel analysis, the results of which may also be inferred from Figure 4, indicates that four factors should be retained.

Choosing a Model

The use of Little Jiffy entails using PCA rather than common factor analysis and orthogonal rather than oblique rotation. It has been argued earlier that use of PCA is

TABLE 9
Factor Correlations for the Direct Quartimin Rotated Four-Factor Solution

	Thickness	*Width*	*Length*	*Density*
Thickness	1.00			
Width	0.65	1.00		
Length	0.42	0.57	1.00	
Density	0.10	0.05	0.04	1.00

TABLE 10
Eigenvalues of Unreduced and Reduced Versions of Gorsuch's (1983) Version of the Holzinger and Swineford (1939) Matrix

	Unreduced	*Reduced*
1	8.163	7.691
2	2.065	1.641
3	1.701	1.215
4	1.522	0.942
5	1.014	0.421
6	0.918	0.400
7	0.891	0.339
8	0.837	0.296
9	0.771	0.232
10	0.727	0.190
11	0.643	0.107
12	0.538	0.039
13	0.535	0.021
14	0.498	–0.022
15	0.464	–0.044
16	0.405	–0.069
17	0.397	–0.100
18	0.343	–0.122
19	0.328	–0.133
20	0.317	–0.150
21	0.290	–0.218
22	0.263	–0.233
23	0.200	–0.252
24	0.171	–0.261

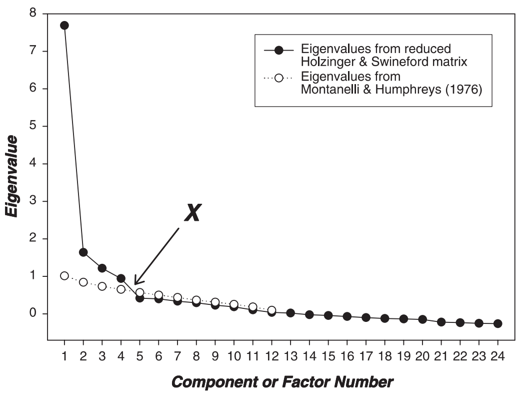

FIGURE 4 Superimposed scree plots for eigenvalues of random and real data sets for the 24 ability variables (N = 145) taken from Holzinger and Swineford (1939), corrected by Gorsuch (1983). There are four eigenvalues before the last big drop, indicating that four factors should be retained. It is evident by using parallel analysis that four factors should be retained.

inappropriate in situations such as the current one, so it would usually not be worthwhile to lend meaning to components. Nevertheless, for the sake of comparison the Holzinger and Swineford matrix was submitted to both PCA and EFA. A five-component PCA solution and a four-factor EFA solution were obtained.

Choosing a Rotation Method

The choice of rotation method is crucial to a clear understanding of the factor structure underlying the Holzinger and Swineford data set. The five-component PCA solution was submitted to orthogonal varimax rotation to remain consistent with the Little Jiffy method, whereas the four-factor EFA solution was submitted to oblique direct quartimin rotation. Rotated loadings for the two solutions are presented in Tables 11 and 12. Factor correlations for the four-factor oblique solution are presented in Table 13. The factor correlations in the oblique solution are consistently greater than zero, demonstrating not only why orthogonal rotation methods are unnecessarily restrictive, but also calling into question the legitimacy of interpreting loadings from an orthogonal rotation.

Interpreting the Results

It has already been demonstrated that the attempt to lend substantive interpretations to components obtained from PCA is inappropriate. Components do not represent latent variables that account for covariances among observed variables. Rather, they represent composite variables that account for observed variances, making no distinction between common and unique variance. However, again for the sake of comparison, let us consider the interpretation of the rotated component loadings in Table 11 as if this solution reflected influences of underlying latent variables.

A comparison of the factor solution in Table 12 and the components solution in Table 11 reveals resemblance between the four factors and the first four components. The fifth component appears to borrow items from components resembling the *memory/recognition* and *spatial/visual* factors reported by Holzinger and Swineford (1939) and Gorsuch (1983). However, there are important differences between the EFA and PCA solutions. Most significantly, the simple structure in the factor solution is clearly superior to that in the components solution, as indicated by the magnitude of the small loadings in each solution. The small loadings in the factor solution are quite consistently, and often substantially, smaller than the corresponding small loadings in the components solution. This distinction suggests better simple structure and thus more precise definition of the constructs in the factor solution than in the components solution. Furthermore, the orthogonal rotation prescribed by the Little Jiffy approach forces these components to be uncorrelated. Together, the overfactoring and orthogonality characterizing the PCA solution render the results difficult to interpret meaningfully. Interpretation of an obliquely rotated EFA solution is, however, appropriate. Table 12 shows a loading pattern quite similar to that reported by Holzinger and Swineford (1939) and Gorsuch (1983), with factors clearly corresponding to *verbal ability*, *speed/numerical ability*, *spatial/visual*, and *memory/recognition* factors, respectively.

Overall, the factor solution is clearly superior. The components solution is distorted by the retention of the fifth component, causing difficulty in interpretation. This solution also displays poorer simple structure than the factor solution and gives the user the misleading impression that the underlying constructs are independent. These failings are attributable to the use of poor technique. As illustrated both with Tom Swift's data and with the empirical data from Holzinger and Swineford (1939), such failings might well be avoided through better decisions about methods.

TABLE 11
Loadings on Five-Orthogonal Components for the Holzinger and Swineford (1939) Data

	Component 1	*Component 2*	*Component 3*	*Component 4*	*Component 5*
1	0.17	0.20	0.70	0.08	0.18
2	0.08	0.10	0.65	0.09	–0.15
3	0.79	0.22	0.16	0.10	–0.01
4	0.81	0.08	0.17	0.18	0.08
5	0.85	0.16	0.14	0.04	0.08
6	0.65	0.24	0.27	0.03	0.19
7	0.85	0.06	0.14	0.16	0.09
8	0.18	0.84	–0.11	0.09	0.00
9	0.20	0.63	0.06	0.29	0.14
10	0.03	0.80	0.21	0.03	–0.02
11	0.20	0.62	0.41	–0.05	0.14
12	0.22	0.08	0.00	0.70	0.12
13	0.09	0.10	0.13	0.74	–0.06
14	0.06	0.09	0.49	0.55	0.14
15	0.17	0.26	–0.05	0.56	0.45
16	–0.00	0.40	0.29	0.37	0.37
17	0.16	0.16	0.09	0.15	0.80
18	0.44	0.09	0.47	0.35	–0.04
19	0.19	0.50	0.43	0.16	0.06
20	0.43	0.12	0.38	0.25	0.24
21	0.43	0.23	0.51	0.18	0.14
22	0.40	0.54	0.11	0.16	0.27
23	0.16	–0.09	0.56	–0.06	0.50
24	0.26	0.07	0.61	0.00	0.11

TABLE 12
Loadings on Four-Oblique Factors for the Holzinger and Swineford (1939) Data

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*
1	0.07	0.01	0.69	0.06
2	0.05	–0.01	0.44	0.03
3	0.78	0.11	0.01	–0.04
4	0.81	–0.06	0.00	0.07
5	0.85	0.05	0.02	–0.09
6	0.57	0.14	0.20	–0.02
7	0.87	–0.09	–0.03	0.07
8	0.07	0.87	–0.17	0.06
9	0.08	0.44	0.05	0.27
10	–0.10	0.69	0.25	–0.01
11	0.09	0.46	0.46	–0.08
12	0.11	–0.02	–0.09	0.57
13	0.01	–0.02	0.00	0.54
14	–0.06	–0.07	0.33	0.51
15	0.04	0.13	–0.10	0.63
16	–0.11	0.23	0.23	0.45
17	0.07	0.06	0.13	0.36
18	0.33	–0.04	0.29	0.24
19	0.07	0.34	0.34	0.14
20	0.32	–0.03	0.28	0.25
21	0.30	0.09	0.40	0.15
22	0.29	0.42	0.03	0.22
23	0.08	–0.16	0.55	0.07
24	0.18	–0.04	0.50	0.01

TABLE 13
Factor Correlations for the Oblique Four-Factor Solution for the Holzinger and Swineford (1939) Data

	Factor 1	*Factor 2*	*Factor 3*	*Factor 4*
Factor 1	1.00			
Factor 2	0.30	1.00		
Factor 3	0.40	0.25	1.00	
Factor 4	0.43	0.32	0.37	1.00

Note. All correlations in this matrix are significantly greater than zero.

CONCLUSIONS

These demonstrations have shown that the choices involved in factor analysis make a difference. A major benefit of making appropriate decisions in factor analysis is a much improved chance to obtain a clear, interpretable set of results. On the other hand, the consequences of making poor decisions often include erroneous, uninterpretable, or ambiguous results. Unfortunately, the methods used by Swift are still commonly used in applied EFA studies today.

As stated earlier, we chose to use the Armstrong article as a surrogate for many EFA studies in the social sciences using similar methods of analysis. Many of these studies undoubtedly suffer the same consequences as those that occurred in Swift's results. A reviewer pointed out that Armstrong's goals were different than those of many applications of EFA in modern literature. The context of Armstrong's article is that of theory development. Many modern applications of EFA, such as those focusing on scale development, assume theory already exists and use it as a basis for interpreting factor solutions. We used Armstrong's article as a case in point, but our criticisms apply just as much to modern applications of EFA as to Armstrong's analysis. Regardless of the researcher's reasons for using EFA, the methodological issues are same, and therefore the consequences of the researcher's choices will be the same.

Three recommendations are made regarding the use of exploratory techniques like EFA and PCA. First, it is strongly recommended that PCA be avoided unless the researcher is specifically interested in data reduction.[11] If the researcher wishes to identify factors that account for correlations among MVs, it is generally more appropriate to use EFA than PCA. Detractors of common factor analysis often raise the specter of factor indeterminacy—the fact that infinitely many sets of unobservable factor scores can be specified to satisfy the common factor model (for an overview of the indeterminacy issue, see Mulaik, 1996; Steiger, 1979). However, factor indeterminacy does not pose a problem for the interpretation of factor analytic results in most circumstances because factor scores need not be computed in the first place, as in the Tom Swift example. If the purpose is to compute factor scores to represent individual differences in a latent variable, and then to use those factor scores in subsequent analyses, SEM can usually be employed instead, thus eliminating the need to obtain factor scores.

Second, it is recommended that a combination of criteria be used to determine the appropriate number of factors to retain, depending on the EFA method used (e.g., OLS vs. ML). Use of the Kaiser criterion as the sole decision rule should be

[11]See Cliff and Caruso (1998) for discussion of a special form of components analysis called reliable component analysis (RCA), which they suggested may offer a useful approach to exploratory factor analysis.

avoided altogether, although this criterion may be used as one piece of information in conjunction with other means of determining the number of factors to retain.

Third, it is recommended that the mechanical use of orthogonal varimax rotation be avoided.[12] The use of orthogonal rotation methods, in general, is rarely defensible because factors are rarely if ever uncorrelated in empirical studies. Rather, researchers should use oblique rotation methods. When used appropriately, EFA is a perfectly acceptable method for identifying the number and nature of the underlying latent variables that influence relationships among measured variables.

ACKNOWLEDGMENTS

We thank Ledyard Tucker for providing the authors with a reconstructed version of Tom Swift's data and Paul Barrett, Michael Browne and several anonymous reviewers for helpful comments.

REFERENCES

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

Beidel, D. C., Turner, S. M., & Morris, T. L. (1995). A new inventory to assess childhood social anxiety and phobia: The Social Phobia and Anxiety Inventory for Children. *Psychological Assessment, 7,* 73–79.

Bell-Dolan, D. J., & Allan, W. D. (1998). Assessing elementary school children's social skills: Evaluation of the parent version of the Matson Evaluation of Social Skills With Youngsters. *Psychological Assessment, 10,* 140–148.

Brown, C., Schulberg, H. C., & Madonia, M. J. (1995). Assessing depression in primary care practice with the Beck Depression Inventory and the Hamilton Rating Scale for Depression. *Psychological Assessment, 7,* 59–65.

Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika, 33,* 267–334.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36,* 111–150.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (1998). CEFA: Comprehensive Exploratory Factor Analysis [Computer software and manual]. Retrieved from quantrm2.psy.ohio-state.edu/browne/

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

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Journal of Educational Measurement, 14,* 289–325.

[12]Browne (2001) suggested that the continued popularity of varimax is due mainly to ease of implementation rather than to any intrinsic superiority.

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

Cliff, N. (1992). Derivations of the reliability of components. *Psychological Reports, 71,* 667–670.

Cliff, N., & Caruso, J. C. (1998). Reliable component analysis through maximizing composite reliability. *Psychological Methods, 3,* 291–308.

Collinsworth, P., Strom, R., & Strom, S. (1996). Parent success indicator: Development and factorial validation. *Educational and Psychological Measurement, 56,* 504–513.

Copeland, A. L., Brandon, T. H., & Quinn, E. P. (1995). The Smoking Consequences Questionnaire–Adult: Measurement of smoking outcome expectancies of experienced smokers. *Psychological Assessment, 7,* 484–494.

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

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18,* 147–167.

Cureton, E. E. (1939). The principal compulsions of factor analysts. *Harvard Educational Review, 9,* 287–295.

Dunn, G. E., Ryan, J. J., & Paolo, A. M. (1994). A principal components analysis of the dissociative experiences scale in a substance abuse population. *Journal of Clinical Psychology, 50,* 936–940.

Dyce, J. A. (1996). Factor structure of the Beck Hopelessness Scale. *Journal of Clinical Psychology, 52,* 555–558.

Enns, R. A., & Reddon, J. R. (1998). The factor structure of the Wechsler Adult Intelligence Scale–Revised: One or two but not three factors. *Journal of Clinical Psychology, 54,* 447–459.

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.

Flowers, C. P., & Algozzine, R. F. (2000). Development and validation of scores on the Basic Technology Competencies for Educators Inventory. *Educational and Psychological Measurement, 60,* 411–418.

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

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

Gass, C. S., Demsky, Y. I., & Martin, P. C. (1998). Factor analysis of the WISC–R (Spanish version) at 11 age levels between 62 and 162 years. *Journal of Clinical Psychology, 54,* 109–113.

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

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

Harman, H. H. (1967). *Modern factor analysis*. Chicago: University of Chicago Press.

Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs.* Chicago: University of Chicago,

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

Humphreys, L. G. (1964). Number of cases and number of factors: An example where *N* is very large. *Educational and Psychological Measurement, 24,* 1964.

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

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

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

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Educational and Psychological Measurement, 23,* 770–773.

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

Kier, F. J., & Buras, A. R. (1999). Perceived affiliation with family member roles: Validity and reliability of scores on the Children's Role Inventory. *Educational and Psychological Measurement, 59,* 640–650.

Kwan, K.-L. K. (2000). The internal–external ethnic identity measure: Factor-analytic structures based on a sample of Chinese Americans. *Educational and Psychological Measurement, 60,* 142–152.

Lawrence, J. W., Heinberg, L. J., Roca, R., Munster, A., Spence, R., & Fauerbach, J. A. (1998). Development and validation of the Satisfaction With Appearance Scale: Assessing body image among burn-injured patients. *Psychological Assessment, 10,* 64–70.

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.

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

Mulaik, S. A. (Ed.). (1996). [Special issue on factor score indeterminacy]. *Multivariate Behavioral Research, 31.*

Osman, A., Barrios, F. X., Aukes, D., & Osman, J. R. (1995). Psychometric evaluation of the social phobia and anxiety inventory in college students. *Journal of Clinical Psychology, 51,* 235–243.

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.

Shiarella, A. H., McCarthy, A. M., & Tucker, M. L. (2000). Development and construct validity of scores on the Community Service Attitudes Scale. *Educational and Psychological Measurement, 60,* 286–300.

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

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

Tateneni, K. (1998). *Use of automatic and numerical differentiation in the estimation of asymptotic standard errors in exploratory factor analysis*. Unpublished doctoral dissertation, Ohio State University.

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1940). Current issues in factor analysis. *Psychological Bulletin, 37,* 189–236.

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

Tucker, L. R, Cooper, L. G., & Meredith, W. (1972). Obtaining squared multiple correlations from a correlation matrix which may be singular. *Psychometrika, 37,* 143–148.

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

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

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.

Tzeng, O. C. S. (1992). On reliability and number of principal components: Joinder with Cliff and Kaiser. *Perceptual and Motor Skills, 75,* 929–930.

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

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

Wilson, E. B., & Worcester, J. (1939). Note on factor analysis. *Psychometrika, 4,* 133–148.

Wolfle, D. (1940). Factor analysis to 1940. *Psychometric Monographs* (No. 3). Chicago: University of Chicago Press.

Yanico, B. J., & Lu, T. G. C. (2000). A psychometric evaluation of the Six-Factor Self-Concept Scale in a sample of racial/ethnic minority women. *Educational and Psychological Measurement, 60,* 86–99.

Yeomans, K. A., & Golder, P. A. (1982). The Guttman–Kaiser criterion as a predictor of the number of common factors. *The Statistician, 31,* 221–229.

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

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