Identifiability of Spurious Factors Using Linear Factor Analysis with Binary Items

Samuel B. Green
Auburn University

The purpose of this study was to evaluate the robustness of some linear factor analytic techniques to violations of the linearity assumption by factoring product-moment correlations computed from data conforming to an extended, three-parameter logistic model of item responding. Three factors were crossed to yield 81 subcases: the number of underlying dimensions (0, 1, or 2), the number of items (10, 15, 20, 25, 30, 35, 40, 45, or 50), and the number of subjects (100, 250, or 500). The mean eigenvalues for the subcases were evaluated using parallel analysis and the scree technique. The mean eigenvectors were visually inspected. For almost all subcases with one or two underlying dimensions, a single spurious factor was able to be identified using parallel analysis. However, in comparison with the nonspurious factors, it was small in magnitude and, in practice, factors of this relative size might be interpreted as trivial. It was concluded that researchers may have some confidence in interpreting linear factor analysis with binary items if they are using a test instrument that has been carefully developed.

Appropriate test interpretation depends upon knowing the number of dimensions underlying the items and the correspondence between the items and these dimensions. The standard method used to assess the dimensionality of a cognitive test is linear factor analysis, which assumes a linear relationship between items and factors. However, most test theorists believe that the assumption of linearity is overly restrictive for items from a cognitive measure and that S-shaped curves such as the logistic function more closely approximate the relationship between items and factors (Hambleton & Cook, 1977; Lord, 1980). The purpose of this paper is to evaluate the robustness of some linear factor analytic techniques to violations of the linearity assumption by factoring product-moment correlations (phi's) computed from data conforming to an extended, three-parameter logistic model of item responding.

Factor analytic writers (e.g., Gorsuch, 1974; Rummel, 1970) have suggested that the consequence of factoring a matrix of phi coefficients is an overestimation of the number of underlying dimensions. They have argued that spurious factors, referred to as difficulty factors, are likely to appear because the magnitude of a phi coefficient is inappropriately sensitive to differences in difficulty levels between items. One possible remedy to this perceived problem is the use of alternative indexes to phi, such as tetrachorics and phi over phi-maximum. However, these coefficients have not escaped criticism. For example, tetrachoric correlations do not necessarily yield correlation-like matrices (Gramian matrices) and may also yield spurious factors (Lord, 1980).

McDonald (1965) and McDonald and Ahlawat (1974) have argued that it is inappropriate to focus

the concern on the choice of correlation indexes. They have shown that if the relationship between items and underlying factors is linear, the factoring of phi correlations does not yield spurious factors. On the other hand, if this relationship is nonlinear, additional noncontent factors may appear. Specifically, given a two-parameter normal ogive model of item responding, McDonald and Ahlawat (1974) showed that, in principle, a series of spurious factors would result. They indicated that, in practice, only one of these spurious factors may be detectable and then only if the items vary in difficulty.

Alternative factor analytic techniques have been developed recently by Christofferson (1975) and Muthén (1978) that allow for nonlinear relationships between items and factors. However, due to the great computational complexities involved with these approaches, they are practically unemployable for medium to large sets of items (Gustafsson, 1980).

Regardless of what should be done to determine test dimensionality, researchers routinely apply linear factor analysis to phi coefficients. They probably tend to make this choice because it is a well-known method that may be accessed on most computer systems and no single alternative approach is perceived as preferable. Given the frequency with which this technique is applied, it becomes critically important that researchers know under what conditions spurious factor(s) are likely to appear and how to differentiate them from content factors. This article attempts to address these issues by inspecting the eigenvalues and factor loadings resulting from the factoring of phi correlations based on computer-generated item data.

To limit the possible cases investigated, the research focused on the determination of unidimensionality. Because unidimensionality is one of the assumptions underlying the application of most item response theories (Lord, 1980), this focus seemed especially appropriate. Three general types of data structures were explored: (1) no common factors, (2) a single common factor, and (3) two common factors. With these as the true structures, the likelihood of correct and incorrect judgments about unidimensionality could be assessed.

Method

Generation Model

The design of the experiment included three crossed factors. The number of common factors (F) had three levels: zero, one, or two common factors. The number of items (K) was set at one of nine levels: 10, 15, 20, 25, 30, 35, 40, 45, or 50. The third factor was sample size (N), which had three levels: 100, 250, or 500. Twenty sets of data were generated for each of the 27 subcases in which sample size equaled 500; however, 30 iterations were necessary to obtain stable results for the 54 subcases in which sample size equaled 100 or 250.

Data were generated for any particular iteration of a subcase using an expanded three-parameter logistic model:

$$P_{ij} = c_j + (1 - c_j) \frac{e^{1.7a_j(\Theta_{ij} - b_j)}}{1 + e^{1.7a_j(\Theta_{ij} - b_j)}} . \quad [1]$$

P_{ij} is the probability of an individual i responding correctly to item j. a_j, b_j, and c_j are the discrimination, difficulty, and guessing parameters, respectively, for item j. Θ_{ij} represents an individual's ability level on the dimension required by item j. An individual i obtains a correct answer on item j if a generated number from a uniform random number generator with a range from zero to one was less than the generated value P_{ij}; otherwise, the response was considered incorrect.

For each iteration, the a_j, b_j, and c_j parameters were generated from univariate normal distributions. The means and standard deviations of these distributions were 1.0 and .25 for the a_j's, 0.0 and 1.0 for the b_j's, and .20 and .05 for the c_j's. The structure of Θ_{ij}'s was necessarily dependent upon the number of common factors. For the no common factor case, the Θ_{ij}'s were multivariately, normally distributed with the correlation between any pair of dimensions being equal to zero. Each item required a separate dimension so that the number of dimensions equaled the number of items for these subcases. For the one common factor case, the Θ_{ij}'s were univariately, normally distributed so that all items were assessing the same dimension. These

data conform to the standard three-parameter logistic model, although the standard model does not necessarily require that the Θ_{ii}'s be normally distributed. For the third case with two common dimensions, the Θ_{ii}'s were bivariately, normally distributed with zero correlation between the dimensions. The first $K/2$ items, if K was even, or $(K - 1)/2$ items, if K was odd, were a function of the first dimension, while the remaining items assessed the second dimension. For all subcases, the means and standard deviations for all dimensions (Θ's) were set equal to 0 and 1, respectively.

The three cases represent very simplistic structures of tests in that no item ever loaded on more than one dimension. Although a more complex extension of the three-parameter logistic model has been suggested by Sympson (1978), such a model was not used for three reasons: (1) no a priori reason could be given for why these more complex structures would alter the likelihood of occurrence of spurious factors; (2) the more complicated models would produce complex factor-loading matrices which would make the effect of the nonlinear generation model on the loadings more difficult to evaluate; (3) an increase in the number of subcases to include these more complex models would have required an excessive amount of computer time.

Factoring of Generated Data

An interitem correlation matrix of phi's was computed for each data set. Item communalities were estimated by calculating squared multiple correlations between each item and all other items. A second matrix was developed by substituting these estimated communalities for the ones along the diagonal of the phi matrix. Both matrices were subsequently factored yielding a principal components solution for the correlation matrix and a principal axes solution for the matrix with communality estimates as diagonal elements.

The mean eigenvalues for each successive root were calculated for both solutions for all 81 subcases. In addition, for the principal axes solutions, the first F and $F + 1$ eigenvectors, both unrotated and rotated, were averaged for subcases which had sample sizes of 100, 250, or 500 and number of

variables of 10, 20, or 30. The F and $F + 1$ factors were evaluated in that they allowed for an inspection of vectors associated with the correct number of factors, F, and with the correct number of factors plus a spurious factor, $F + 1$.

Specifically, with the unrotated factors, a root mean square value was calculated for each loading on the first $F + 1$ factors. (The first $F + 1$ factors incorporated the first F factors.) Prior to averaging these loadings, the items and, thus, their loadings were reordered within each of the dimensions of the generation model according to their p values (proportion of a sample that received credit for an item) so that the relationship between the loadings and the p values could be determined.

Average loadings were also obtained on the first F rotated factors if F was equal to two and first $F + 1$ rotated factors if F was equal to one or two. In other words, factors were rotated and averaged if they yielded a different solution from the unrotated one (i.e., if the number of factors to rotate was greater than one). These factors were rotated using the varimax criterion (Kaiser, 1958). This statistical method, which is the one most frequently used by data analysts, was chosen because the focus of the present study was on problems researchers encounter in applying factor analysis to binary data. Prior to averaging, the rotated eigenvectors were reordered so that the first factor best represented the first dimension of the generation model and so that, if F was equal to two, the second factor best represented the second dimension.

Fit between factors and a generation dimension was evaluated by calculating for each factor a sum of squared loadings across the items associated with a dimension and by choosing the factor with the largest sum of squares as best representing that dimension. For the solutions with $F + 1$ factors, the remaining vector was placed in the $F + 1$ position. Comparable to the unrotated solutions, the loadings of the items associated with each generation dimension were reordered within the vectors. This reordering was based upon the relative magnitude of the items' p values (from highest to lowest) within dimensions. Finally, the root mean square of each loading was obtained.

Criteria for Determining Number of Factors

The mean eigenvalues for each subcase were evaluated to determine the number of factors. The evaluation of eigenvalues is, in practice, made using a variety of methods such as the Kaiser-Guttman criterion (eigenvalues-greater-than-one criterion), significance tests, and graphical techniques (Mulaik, 1972). Consequently, it was initially unclear which approach to apply to the generated data sets. However, given the objectives of the research, it was decided that the chosen methods should be (1) accurate in comparison with the other approaches and (2) insensitive to sampling error. The application of graphical approaches to the mean eigenvalues of a subcase satisfied these requirements.

Based upon Montanelli and Humphreys' (1976) recommendation, two graphical approaches were used: the scree and the parallel analysis techniques. Both methods attempt to differentiate between predominant factors which have large eigenvalues and trivial, specific, or error factors which have small eigenvalues. However, of the two approaches, only parallel analysis uses an objective criterion to distinguish between large and small values.

The parallel analysis technique involves, in standard applications, the following steps:

1. Calculating a correlation matrix based on the K variables and N observations from an investigator's study;
2. Substituting squared multiple correlations between each variable and the other $K - 1$ variables for the diagonal elements of the correlation matrix;
3. Factoring of this matrix using the principal axes method;
4. Plotting the resulting eigenvalues on the ordinate and their ordinal positions among the extracted roots on the abscissa;
5. Plotting a second curve on the same graph based on a set of values obtained from a formula by Montanelli and Humphreys (1976); and
6. Determining the number of factors to rotate as the number of plotted eigenvalues prior to the first crossing of the two curves.

The Montanelli and Humphreys' (1976) formula estimates the mean eigenvalues resulting from the factoring of matrices of correlations between independent, random normal deviates with squared multiple correlations (communality estimates) as the diagonal elements. These random data matrices contain the same number of variables and observations as the data matrix in the study.

In contrast, the scree procedure (Cattell & Vogelmann, 1977) routinely requires the following steps: (1) calculation of a correlation matrix based on the K variables and N observations from a study, (2) factoring of this matrix using the principal components method, (3) plotting the eigenvalues, and (4) determination of the number of factors to rotate by evaluating where the roots first stop forming a falling curve and where they start falling along a straight line.

The mean eigenvalues for the subcases were evaluated using the two graphical methods to determine the number of factors. However, the 162 graphs, two graphs for each of the 81 subcases, could not practically be presented here. Therefore, they have been summarized in two ways: (1) a graph of one of the subcases was selected to illustrate the results and (2) descriptive statistics were used to delineate further the results of the parallel analyses. In particular, correlations between mean and predicted eigenvalues were calculated for each consecutive root across all subcases with the same number of known factors, F. The number of pairs of data points varied across correlations because the number of variables and therefore the number of eigenvalues differed across subcases and because eigenvalues were excluded that were either near zero or negative. Specifically, only the first $(K/2) - 1$ (for an even number of items) or $[(K - 1)/2] - 1$ (for an odd number of items) eigenvalues for any subcase were considered appropriate for these analyses.

Montanelli and Humphreys (1976) computed these correlations to show the accuracy of their formula to estimate mean eigenvalues for independent, random normal deviates. By using similar summary analyses, the generalizability of their formula to non-normal data could be assessed, that is, con-

forming to the zero factor case. Two additional statistics were computed for all subcases. They were mean differences and mean absolute differences between mean and predicted eigenvalues.

If parallel analysis yields accurate results for the zero dimension case, the descriptive statistics should show congruency between the mean and predicted eigenvalues: correlations close to 1 and average absolute differences and average differences close to 0. A lack of congruency in the initial factor(s), with the mean eigenvalues being larger than the predicted one, would suggest that the use of this analysis would lead researchers to interpret spurious factors. On the other hand, if the initial mean eigenvalues were smaller than the predicted eigenvalues, it would be concluded that researchers are unlikely to extract spurious factor(s). If either of these two types of incongruency occur, parallel analysis would probably yield inaccurate results with the one- and two-dimensional cases.

The first factor for the one-dimensional case and the first two factors for the two-dimensional cases should show a lack of congruency for accurate interpretation with parallel analysis. The mean eigenvalues should be larger than the predicted ones. If this lack of congruency maintains itself for successive factor(s), users of parallel analysis are likely to extract spurious factors.

The decision about the number of factors to retain may involve an evaluation not only of eigenvalues but also of eigenvectors. The general principle used to assess eigenvectors is to retain factors only if their loadings form interpretable patterns. Within this context, understanding the effect of nonlinearity on the eigenvectors is important. Consequently, the results are described for the subcases for which averaged eigenvectors were calculated.

Results and Discussion

In Table 1 descriptive statistics are given to indicate the degree of correspondence between the mean and predicted eigenvalues for the generation models with zero, one, and two dimensions. For those subcases with no common factors, excellent congruency was found for each successive factor

between the mean and predicted eigenvalues. The largest amount of incongruency was shown for the initial factors, but even for these the differences were minimal. For example, with the first extracted factor, there was a slight tendency for the mean eigenvalues to be smaller than the predicted ones. The mean difference was $-.006$. The largest difference was for the subcase with 50 variables and 100 subjects, which had a mean eigenvalue of 2.220 and a predicted eigenvalue of 2.168. The mean eigenvalues associated with the first factor were smaller than the predicted ones for 18 of the 27 subcases. However, no clear pattern emerged that differentiated the underpredicted from the overpredicted subcases. Overprediction occurred for subcases with the number of variables (K) equal to 15, 40, 45, or 50 and the number of subjects (N) equal to 100; with $K = 15$, 25, or 30 and $N = 250$; and with $K = 15$ or 20 and $N = 500$.

Although parallel analysis of data from the generation model with no common factors indicated an absence of spurious factors, these results did not generalize to those from the data generated by the models with one or two common factors. As shown in Table 1 the mean eigenvalues were larger than the predicted ones, not only for the first F factors, but also for the $F + 1$ factor. For the subcases with a single underlying dimension, large differences existed between the mean and predicted eigenvalues for the first factor and much smaller differences for the second one. For both factors the mean eigenvalues were almost always larger than the predicted ones. The one exception was for the second factor when K equaled 10 and N equaled 500. After the initial two factors, the mean eigenvalues were consistently below the predicted ones (two exceptions out of 183 for the third through 10th factors). Therefore, for the subcases with a single underlying dimension, parallel analysis generally indicated two factors, but the second factor just barely met the criterion of this analysis.

Similar results were found for the data with two underlying dimensions. The mean eigenvalues for the first two factors were always much larger than the predicted values. For the third factor, the mean eigenvalues maintained their advantage for 24 of

Table 1

Squared Correlations, Average Absolute Differences, and Average
Differences Between Mean and Predicted Eigenvalues for
Generation Models with Zero, One, and Two Dimensions

Number of Each Successive Factor	Number of Dimensions								
	0			1			2		
	r^2	$\lvert \bar{d} \rvert$	$\bar{d}$	r^2	$\lvert \bar{d} \rvert$	$\bar{d}$	r^2	$\lvert \bar{d} \rvert$	$\bar{d}$
1	.999	.029	-.006	.436	4.398	4.398	.572	2.148	2.148
2	.998	.022	.001	.975	.100	.100	.539	1.653	1.653
3	.999	.016	.000	.998	.029	-.027	.997	.033	.029
4	.998	.017	-.001	.999	.047	-.047	.999	.022	-.022
5	.999	.014	.004	.999	.054	-.054	.998	.035	-.034
6	.999	.011	-.001	.999	.062	-.062	.999	.051	-.051
7	.999	.009	-.000	.999	.063	-.063	.999	.055	-.055
8	.999	.006	-.002	.999	.063	-.063	.999	.058	-.058
9	.999	.008	-.006	.999	.062	-.062	.999	.061	-.061
10	.999	.013	-.013	.999	.072	-.072	.999	.069	-.069

Note.   Because the models had at most two underlying dimensions,
statistics are not presented for factors beyond the first ten.
However, very high congruency was found for these later
factors.

the 27 subcases, but the differences between the two sets of values became much smaller. The three exceptions were for the subcase with $K = 10$ and $N = 250$ or 500 and $K = 40$ and $N = 250$. For the remaining factors, the mean eigenvalues were almost always smaller than the predicted ones (three exceptions out of 156 for the fourth through 10th factors).

Figure 1 presents a graph of the mean and predicted eigenvalues for a subcase with one underlying dimension. Also, mean eigenvalues from the principal components solution are plotted on the same graph for the application of the scree technique. This subcase was chosen for presentation because it showed as well as any of the subcases the presence of a spurious factor. The graph indicates that the second factor would be considered meaningful using the objective approach of parallel analysis but might or might not be judged meaningful using the subjective scree technique. Al-

though the second eigenvalue does not fall precisely on the scree line, it is sufficiently close that many data analysts might consider it of trivial size.

Since the graph is of mean eigenvalues, sampling error is minimized. However, the meaningfulness of the $F + 1$ factor may be better understood for the 81 subcases if the variation among the plots of eigenvalues is considered in addition to the mean values. With a single underlying dimension, the second factor's eigenvalue was larger than the predicted one for more than 75% of the iterations with 11 subcases; for 50% or more of the iterations, but less than 75% of the iterations with 14 subcases; and for less than 50% of the iterations for 2 subcases. With two underlying dimensions, the third factor's eigenvalue was larger than the predicted value for more than 75% of the iterations with 8 subcases; for 50% or more of the iterations, but less than 75% of the iterations with 16 subcases; and for less than 50% of the iterations

Figure 1

Mean Eigenvalue Plots for Subcase with
One Underlying Dimension,
30 Variables, and 500 Subjects
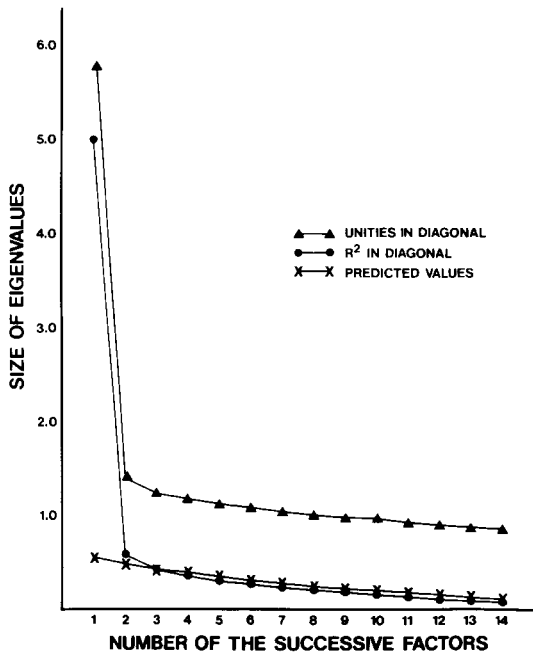
for 3 subcases. In comparison, the initial F factors met the parallel analysis criterion for all iterations for all subcases with one or two underlying dimensions. Consequently, while the initial F factors were properly identified taking into account sampling error, the $F + 1$ factor was not always detected.

Table 2 presents averaged factor loadings for a subcase with a single underlying dimension (the subcase presented in Figure 1). In order to interpret this table, the sign of these loadings for any iteration must be taken into account. The loadings on the first unrotated factor were generally either all positive or all negative for any iteration. The same result held for the loadings of the first and second factors after rotation. However, the items with the 15 highest p values tended to have loadings on the second unrotated factor that were opposite in sign to the others on that factor.

The results for this subcase illustrate those for the subcases with either one or two underlying

common factors For both unrotated and rotated solutions, a quadratic relationship between the loadings on each of the initial F factors and the p values was observed for items that assessed the same underlying dimension. Specifically, items with more extreme p values tended to have smaller loadings, and those items assessing the same underlying dimension generally had the same sign for any factor. For unrotated solutions, a linear relationship between the p values and the loadings on the $F + 1$ factor existed for items that assessed the same dimension. Items with more extreme p values tended to have larger loadings, but items with low p values tended to have signs opposite to those with high p values. Similarly, the p values and the loadings on the $F + 1$ factor tended to be linearly related after rotation for the subcases with two underlying dimensions. However, for the data with a single underlying dimension, the loadings of the $F + 1$ factor after rotation were quadratically related to the p values.

These relationships between the p values and the loading for the initial F and $F + 1$ factors were stronger for the data with only a single underlying dimension. For example, the root mean square correlations between the p values and the $F + 1$ factor loadings for the unrotated solutions were .81 for the subcase with a single underlying dimension presented in Table 2 and .46 for the subcase with the same number of variables and subjects but with two underlying dimensions.

Although the strength of the relationships between the p values and the loadings varied to some extent across subcases, the type of relationships remained relatively constant. Because factor loadings may be influenced by item difficulty, applied researchers may have problems interpreting their factor analytic results. For example, a researcher might conclude that very easy or difficult items do not assess their underlying dimensions when, in fact, they are sensitive measures for individuals whose scores fall at the extremes of their underlying distributions.

Summary and Conclusions

Data were generated using an extended three-parameter logistic model. For the 27 subcases with

Table 2
Average Unsigned Factor Loadings Listed in Order of Item
Difficulty for Subcase with N = 500, K = 30, and F = 1

Unrotated Factors		Rotated Factors		Mean Item
1	2	1	2	Difficulties
.34	.24	.33	.25	.93
.39	.22	.37	.25	.89
.42	.22	.37	.29	.86
.44	.19	.38	.29	.84
.46	.18	.40	.29	.81
.48	.15	.41	.30	.79
.44	.13	.36	.29	.76
.46	.11	.37	.29	.74
.48	.12	.39	.30	.72
.47	.10	.38	.30	.70
.47	.09	.36	.31	.68
.49	.08	.39	.30	.66
.46	.08	.36	.30	.63
.47	.09	.36	.31	.62
.42	.08	.32	.29	.61
.43	.09	.33	.29	.59
.46	.11	.33	.33	.57
.45	.10	.34	.31	.56
.45	.10	.34	.31	.55
.43	.11	.33	.30	.53
.42	.11	.31	.30	.51
.40	.13	.28	.30	.49
.40	.13	.30	.29	.47
.40	.15	.29	.32	.44
.38	.14	.27	.30	.42
.34	.16	.24	.29	.40
.36	.20	.28	.30	.37
.32	.16	.23	.28	.33
.28	.19	.23	.25	.29
.24	.17	.15	.25	.23

no common factors, parallel analysis indicated the presence of no common factors. For the remaining 54 subcases with either one or two common factors, a single spurious factor was almost always detected using parallel analysis. However, an inspection of the eigenvalue plots showed this factor to be very minor in comparison with the factors extracted prior to it. The nonlinear relationship between the items and the underlying dimensions also influenced the factor loadings. The very easy and very difficult

items tended to have smaller loadings on the factor that best represented their underlying dimensions.[1] The factor analyst should not interpret these items with smaller loadings to mean that they do not measure the underlying dimensions. These items are just not sensitive measures of this dimension for the majority of the individuals who have taken the test.

In summary, the findings of this monte carlo study suggested that researchers may be able to apply linear factor analysis to explore the dimensionality of tests and still obtain useful results. They should question the meaningfulness of minor factors and be careful in their interpretation of factor loading matrices. However, these conclusions can be generalized rigorously only to a certain type of test on the basis of this study. The model used to generate the data was defined such that (1) each underlying dimension was represented by at least five items; (2) the discriminations of the items were relatively high, with the *a* parameter values clustered around 1.0; and (3) guessing was minimized, with the values for *c* rarely exceeding .30. In general, actual tests that are likely to exhibit these characteristics are those that have been carefully developed prior to the application of factor analysis. Undoubtedly, tests with poorer psychometric characteristics may have yielded different results. For example, if a measure contains minor factors that are represented by a small proportion of the test items, a factor analyst would probably have difficulty differentiating these minor factors from spurious ones due to a lack of linearity.

It is arguable as to whether the focus of this study should have been on well-developed tests. Nevertheless, the choice can be at least partially justified in that the degree of confidence in factor analytic results is, invariably, directly related to the researcher's care in selecting variables included in an analysis and the quality of them.

References

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

Christoffersson, A. Factor analysis of dichotimized variables. *Psychometrika*, 1975, *40*, 5–32.

Gorusch, R. L. *Factor analysis*. Philadelphia: Saunders, 1974.

Gustafsson, J. E. Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 1980, *33*, 205–233.

Hambleton, R. K., & Cook, L. L. Latent trait models and their use in analysis of educational test data. *Journal of Educational Measurement*, 1977, *14*, 75–96.

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

Lord, F. M. *Application of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum, 1980.

McDonald, R. P. Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1965, *18*, 11–23.

McDonald, R. P., & Ahlawat, K. S. Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 1974, *27*, 82–99.

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

Mulaik, S. A. *Foundations of factor analysis*. New York: McGraw-Hill, 1972.

Muthén, B. Contributions to factor analysis of dichotomous variables. *Psychometrika*, 1978, *43*, 551–560.

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

Sympson, J. B. A model for testing with multidimensional items. In D. J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

---

[1] Besides the 81 subcases discussed, a number of additional subcases with alternative parameter values were nonsystematically explored. For example, data with a single underlying dimension were generated in which half the items had *b* values of 1.0 and the other half had values of − 1.0. The factor loadings for this case had a different pattern of results. Although it is unlikely that *b* values would be bimodally distributed in practice, it is relevant to note that other possible patterns of loadings are possible. For this subcase, parallel analysis of the mean eigenvalues did not indicate the presence of a spurious factor.

Author's Address

Send requests for reprints or further information to Samuel B. Green, Department of Psychology, Auburn University AL 36849, U.S.A.