PERSONALITY STUDY AND GROUP BEHAVIOUR, 1982, 3, 1.

Factor Extraction : An Examination of Three Methods

Paul T. Barrett and Paul Kline[1]

Abstract : *The Kaiser-Guttman root one criterion, Velicer's minimum average partial correlation and a machine implemented version of Cattell's scree test were used on 26 sets of data, 3 of which were plasmodes. Both principal components and image factors were calculated. The results yielded equivocal estimates of the number of factors in nearly every case. The notion of a test for limiting the number of factors was examined in addition to a detailed analysis of the Kaiser-Guttman logic. From this analysis it was concluded that only the Kaiser argument is of help, given that the implications and constraints in using coefficient α are accepted. Finally, the problem of when to stop factoring was conceived primarily as one that requires test of factor validity, reliability and replicability rather than those with a numerical basis.*

Introduction :

Perhaps the most important stage within a factor analysis (FA) of a set of data is deciding how many factors to extract prior to rotation. Too few restricts the available solution space thus tending to compress the variables onto factors. Extracting too many opens out the solution space to such an extent that variables tend to bunch into 'specific' factors containing, in the case of tests, homogenous item subsets. Unfortunately, there is no exact mathematical technique that permits the precise identification of the factors to be retained. This is not due to any intractable mathematicostatistical problem but rather is a concomitant of the multivariate analysis technique. FA will simply reduce a correlation or covariance matrix to a set of linear combinations (factors) of the variables. Deciding how many factors to extract is dependent upon their conceptual validity and replicability only. Tests for extraction quantity are only lower

1. Department of Psychology University of Exeter Washington Singer Laboratories Exeter Devon EX4 4QG.

bounds or limiting conditions, although using an artificially generated correlation or covariance matrix (plasmode), the test can be made to appear exact. However, 'real' matrices contain an unknown quantity of measurement error which will appear as factors and distort to some extent the structure of the unknown quantity of 'true' factors.

The most appealing solution to this problem is to use plasmodes with or without the addition of measurement error. However, the more 'real' they become, the less exact the tests appear (Linn, 1968). There is simply no way of adding in 'typical error' as Cattell and Vogelmann (1977) assert. Generally plasmodes are structurally too simple with correspondingly simple properties. A special type of plasmode is that based upon observations within a precise measurement domain, e.g. physics or physiology. The principle here is that extensive theory and empirical evidence abounds for expressing phenomena in precise relationships. For example, Cattell and Dickman's (1962) ball problem, Thurstone's (1947) box problem, Coan's (1961) egg factoring and Sokal et al (1980) exercise variable plasmode. Invariably, such plasmodes are structurally simpler than those commonly found in psychological studies. (Breaks in eigenvalue slopes are generally extremely well defined). Thus there is no easy equivalence property here.

The alternative is to use real data matrices in order to evaluate techniques. However, the criterion solutions are those that are considered by factor analysts to be the best in terms of conceptual validity, replicability and usefulness of the factors extracted. There are few solutions that satisfy these criteria. Therefore, examinations of tests of factor extraction have used both types of data to assess their accuracy, with differing results. Tests applied to artificially generated data invariably reach the same conclusions, applying them to real data unfortunately yields equivocal results. (Linn, 1968; Hakstian and Muller, 1973; Cattell and Vogelmann, 1977; Revelle and Rocklin, 1979).

As to the tests designed for factor extraction, they fall into two conceptual categories, the psychometric and the statistical. Psychometric tests are those that accept the initial correlations or covariances at face value and simply decide upon the number of factors that satisfy a particular criterion. Examples of such tests are the Scree (Cattell, 1966), the Kaiser-Guttman (KG) criterion

(Guttman, 1954; Kaiser, 1960).

Horn's (1965) test, Linn's (1968) mean square ratio criterion, Veldman's (1974) Varimax based criterion, Velicer's (1976) MAP test, and Revelle and Rocklin's (1979) Very Simple Structure test. Statistical tests are those which treat the input matrix as a sample from a population matrix, using inferential statistical logic to determine the number of factors to extract. Examples of these tests are Bartlett's (1950) chi square test, Rippe's (1953) large sample chi square criterion, and the Maximum Likelihood factor analysis criterion (Anderson and Rubin, 1956; see also Tucker and Lewis's (1973) derivation of a reliability coefficient). Generally both methods contain flaws, the psychometric for its seemingly arbitrary rules and the statistical because its accuracy depends on the sample size. Clearly there are problems with precise tests of factor extraction. However, the results reported here are more extensive than previous analyses and considerations of conceptual validity and replicability are taken into account.

Three tests of factor extraction are compared below, the KG, an automated Scree test, and Velicer's minimum average partial correlation (MAP) test. Two methods of analysis are used in generating the solutions, principal components analysis (PCA) and Image factoring (Guttman, 1953; Kaiser, 1963). The KG and Scree were chosen primarily because of their popularity, the MAP test because it is relatively new and appears to offer a good conceptual rationale for its use. Additionally, the MAP test is claimed to work on image components as well as PCA.

Method

Data Sets

In total, 29 data sets were used for the analyses. The first 6 are those presented in Velicer (1976) : Harman's (1960) 8 physical variables, Emmett (1949), Maxwell's (1961) 810 normal children, Thurstone (1951), Bechtoldt's (1961) sample I, and Lord's (1956) 33 independent variable matrix. The next 4 are numerically large variable data sets : Kline and Storey (1978), a sample of students (N=406) and a sample of adults (1198) administered the Eysenck Personality Questionnaire (Barrett and Kline, 1980b). Five other data sets with intermediate variable quantities were : Skinner and Rampton (1973), Nesselroade and Baltes (1972), Harman's (1960) 24 psychological variables, and Barrett and Kline (1980c). The next 8 data sets were composed from 3 separate subject samples taken from Barrett and Kline (1980c, 1980d, and 1980e) These data sets are radial parcel (Cattell 1974) factor analytic solutions of the Eysenck Personality Ouestionnaire and 16PF. They are grouped separately because of the nature of the variables used. Radial parcelling results in very clear factor patterns due to the partial elimination of item measurement error with the additional increase in range of correlation coefficients Comrey's (1970) FHID parcel analysis data is also included in this section as it falls into the category of composite variables. The next 3 data sets are plasmodes. The first 2 were generated by the authors. The procedure is presented clearly by Cattell and Vogelmann (1977), the 2 data sets here were generated from orthogonal factor patterns. These contain no measurement error although entering 1.00 in the diagonal of the generated reduced correlation matrix boosts the variance to be extracted by an amount equal to the number of variables minus the sum of the diagonal value of the generated correlation matrix. The communality range for the smaller plasmode was between 0.14 to 0 74, for the larger it was 0.10 to 0 90. The third data set was Kaiser and Horst's (1975) 20 variable box plasmode with added measurement error to make the correlation matrix nonsingular. The final three data sets are included simply as an examination of the automated scree test's consistency of results using the same set of variables with differing subject sample size. The data was taken from Horn and Engstrom's (1979) analysis of the Scree and Bartlett's chi square test.

The Analyses

All numerical analyses were carried out by computer, working to 16 significant digit precision. The PCA routine was used purely on correlation matrices. The image factorings (IFA) were computed from the image covariance matrices :

$$G = R + S^2R^{-1}S^2 - 2S^2$$

rather than from the rescaled correlation matrices :

$$R^* = S^{-1}RS^{-1}$$

where R is the sample correlation matrix and $S^2 = (\text{diag } R^{-1})^{-1}$. Although the orthogonal loading pattern matrix A contains the covariances of variables n with the factors, as :

$$G \rightarrow R - S^2 \text{ as } n \rightarrow \infty$$

so do the loadings g_{ij} of G tend towards the values of the correlation coefficients of variables with the factors. As Harman (1976) indicates, there is very little difference in terms of loading comparison between factoring G or R-S^2.

Velicer (1976) introduced the MAP test for all types of component analyses, i.e. image and PCA. Given A is the n (variables) $\times$ m (component factors) orthogonal pattern matrix resulting from component analysis. For each factor m in A, the partial covariance matrix can be represented as :

$C = R - AA'$ for PCA

or

$C = (S^{-1}RS^{-1}) - AA'$ for rescaled R image factoring.

Thus the matrix of partial correlations is given by :

$R_p = D^{-\frac{1}{2}} C D^{-\frac{1}{2}}$

where

D = Diag C

In order to determine the number of foctors to extract, Velicer proposed the summary statistic :

$$fm = \sum_{i \neq j} \sum (r_{ij})^2 / (n(n-1))$$

Where r_{ij} is the element in row$_i$ and colum j of the matrix R_p in (6). The value of fm is the average of the squared partial correlations after the first on factors are partialled out. Velicer's proposed stopping point is the value of m for which fm is at a minimum; fm ranging between 0 and 1. The logic of the test is that as factors are partialled out, the value of fm decreases to a minimum indicating that the factors are 'common'. When fm begins to increase, the additional factors are viewed as 'specifics' accounting for unique variance only. Thus partialling out a common factor lowers the majority of values of the element in R_p. Partialling out a specific factor, because it has little correlation with the majority of the elements, produces a higher value at f min+1 that at f min : Velicar (1976) demonstrates this with a simple example.

The Scree test is based upon the slope of eigenvalues plotted against their extraction order. When the eigenvalues are successively plotted, a falling curved section followed by a straight line (or several) at a much lesser angle to the horizontal is observed.

The resemblence of these straight line sections to the screes of rock debris running straight at an angle of 'rock stability' at the base of a mountain led Cattell to propose the name 'Scree test'. Cattell (1966) and Cattell and Vogelmann (1977) present some theory for this test in addition to an extensive empirical test of the method versus the K-G on various plasmodes. However, a cursory glance at Cattell and Vogelmann's eigenvalue runs suggests that the plasmodes were too simple in structure. The majority of real data eigenvalue runs do not yield such invitingly clear breakes; Cattell (1978) has suggested four rules for applying the test, stressing that the subjectivity of decision occurs in combining the rules and conditions. Herein lies the element of 'art'. Unfortunately, when faced with complex eigenvalue runs, the element of art expresses itself in different decisions from various investigators. From the literature, it would appear that only a handful of psychometrists are capable of correctly detecting the screes from the retained factors.

In priciple, the concept of the Scree test is that of detecting discontinuities within a two dimensional series of values Simply, given a line function, how does one detect breaks in that function ? Cattell has proposed a set of rules that unfortunately can quite easily be misconstrued by the user. The Scree test in essence is a subjective test. In an attempt to automate the Scree procedure, an algorithm was generated by the authors that attempts to encapsulate, to some extent, some decision processes taking place in choosing scree lines. However, this algorithm is designed purely to find discontinuities in line functions at varying levels of sensitivities, it is not concerned with finding 1, 2, or 3 screes but may find up to 10 or more.

The algorithm, which for convenience will be called AUTOSCREE is certainly one of many possible approaches to the problem and is not necessarily the most efficient. AUTOSCREE works by starting off from the lowest eigenvalues, computing a least squares linear regression on a quantity of eigenvalues n, the tangent of slope about this line, and the value of 1 —coefficient of determination. The next eigenvalue in the series is now added to the previous set of values, the least squares estimation computed, and now the difference between tangents and the difference in error are noted. If either are large than certain test values, the quantity of n eigenvalues are taken to be a scree line.

AUTOSCREE then begins again under the same conditions from where it stopped. The final breaks are detected by an excessive angle deviation between the last scree set and the new scree set. Additionally, should any scree line ascend beyond a 20° slope, the run is terminated for the set of relevant control conditions. Thus summarising the details, there are 3 control parameter fields : group size, angle deviation, and error deviation. The group size is varied from a minimum of 3 values to an approximate maximum of one third of the total number of eigenvalues. The angle deviation (AD) is varied from 1° to 5° in 1° steps. The error deviation (ED) control is varied from 0.0000001 to 0.01 in steps of ED $\times 10^1$. AD controls the final scree line detection, ED controls the within scree cutoffs only. Thus the sequence operates within each AD value, each group size containing the six ED runs. The decision as to the retained factors is made by observing the frequency distribution of stopping points across all control values. Additionally the summary frequency data is split into two groups a high-medium sensitivity angle detection range (1°—3°) and a medium-low sensitivity range (3°—5°). A clear 'true' break yields a clear stopping point across all angle ranges, a slight 'true' break is likely to yield at least two decisions Thus subjectivity now enters the decision process, the investigator having to choose the most likely 'true' break in the AUTOSCREE analysis. Of course, values of the group sizes, angles and errors are ad hoc. The values used here represent the judgement and experience of the authors from a few hundred tests made with AUTOSCREE.

The Scree, MAP, and KG criteria were applied to both factoring methods. While the KG is not normally associated with factor extraction from reduced correlation matrices, as will be made evident in the discussion, there appears to be no sufficient mathematical reason for its lack of use.

Results

Table 1 presents the results of the analyses carried out on the 29 data sets. The multiple values for the Scree tests indicate that these were the breaks detected by AUTOSCREE with either equal or near equal frequency across all parameter groups. Of course, the MAP test could not be computed for the Horn and Engstrom data as only the eigenvalue series were provided by these authors. Notably, their scree solutions yielded two different

Table 1

			Principle Components			Image Components		
Study	No.Subj	No.Var	K-G	Vel.	Scree	K-G	Vel.	Scree
Harman	305	8	2	2	2	2	2	3
Emmett	211	9	2	2	3	1	2	2
Maxwell	810	10	2	2	5:2	1	2	4
Thurstone	213	13	3	2	5	2	3	3
Bechtoldt (SI)	212	17	5	3	2:3	2	3	2:3
Lord	649	33	6	5	3:5	4	5	3:5
Large Variable Data Sets								
Kline and Storey	128	60	15	9	13:10	9	8	6:8
Barrett and Kline (a)	1198	90	23	5	9:10	4	10	5
Barrett and Kline (a)	406	90	29	6	6	6	6	6
Barrett and Kline (b)	491	184	62	11	11	19	11	11
Others								
Skinner and Rampton	2215	20	5	4	5	4	7	3:6
Nesselroade & Baltes	1862	20	5	4	5	3	4	5:6
Harman	145	24	5	4	4	2	4	2
Barrett and Kline (c)	79	24	7	5	7	4	5	6:4
Radial Parcel Solutions								
Barrett and Kline (d)	1198	11	4	3	4	2	3	3
" " "	1198	22	4	4	4:7	4	6	6
" " "	1198	44	8	4	8	4	7	8:7
" " "	406	11	4	2	4	2	2	2
" " "	406	22	4	4	4	3	6	6:4
" " "	406	44	12	4	8	4	4	4
Barrett and Kline (a)	491	46	13	6	6	4	6	7:4
" " "	491	92	26	9	6	6	11	6:11
Comrey	746	44	10	8	5	5	9	3:5
Plasmodes								
10 Vars 3 Facs	—	10	3	2	2:4	1	2	3:4:5
30 Vars 5 Facs	—	30	5	5	6	4	5	3:4:5
Kaiser and Horst	40	20	3	3	5	3	3	3
Scree Analysis								
Horn and Engstrom (L)	103	31	9	—	5	—	—	—
Horn and Engstrom (U)	132	31	7	—	5	—	—	—
Horn and Engstrom (T)	829	31	5	—	3:5:6	—	—	—

values for the samples : L=10, U=10, T=7. If the AUTOSCREE is a 'correct' solution. Horand Engstrom's arguments for the inherent statistical basis of the Scree test are eguivocal. Whether or not 5 or 7 factors is the 'true' value is impossible to answer without knowledge of the factor loading pattern, replicability and validity of the factors themselves. Further empirical evidence of the stability oi AUTOSCREE can be seen in Barret and Kline (1980a) on 4 split sex samples for 406 students and 1198 adults administered the EPQ. The EPQ factors are not overdetermined but simply determined clearly by an adequate variable pattern. One does not overdertermined factors in order to obtain clear solution with various sample sizes. Clearly, the notion of overdetermining factors is fallacious. If the investigator attempts to factor variables that produce weak factor loading pattern with large sample sizes, at low sample sizes the error variance, as Horn and Engstrom point out, would be excessive. Measurement error would be expected to significantly influence the eigenvalue series only when the common factors are defined by weak variable patterns. Such patterns having low factot validity as defined by Cattell and Tsujioka (1964) and therefore theoretically and empirically of little conceptual value.

The use of the MAP test on Equation (1) rather than (2) made no difference to the test results except that a rather uneven function of fm was ovserved. That is instead of fm slowly decreasing to a minimum and then rising to the value of 1.000, fm would beign to decrease, suddenly increase, then decrease again to the minimum before slowly rising to the value of 1.000. Several check tests were made on Equation (2) results to determine the minimum fm value. These checks revealed the same onswer as tests carried out using Equation (1) results. The reason for this uneven function is not at all obvious, especially when one considers the logic of the test. (Of course, computational errors were the most obvious reason but these were ruled out entirely.) Unfortunately, this uneven function did not occur in all data sets but no apparent pattern or systematic reason could be found by the authors. The important results, however, is that the test was valid using Equation (1) as well as (2).

Discussion

From the results of Table 1, three conclusions are evident. Firstly, there is little agreement between the KG, AUTOSCREE, or the MAP test on numerically large data sets. The overestimation of the KG is well known in this context. Secondly, there is little agreement between the KG criterion on PCA and IFA. Thirdly, the MAP test is fairly consistent over PCA and IFA, with AUTO-SCREE in general agreement with the MAP results. If the results were gauged as to the exact agreement between AUTOSCREE and the MAP test, then there is little or no agreement whatsoever. Invariably AUTOSCREE yielded a break in agreement with the MAP criterion but the frequency count was too low to suggest a 'true' break. Overall, the results demonstrate that using any one test of factor extraction is unsatisfactory. In addition, as there is no way of mathematically determining the true number of factors, all tests can only yield estimates or lower bounds. If the estimates from several tests converge on one value, this can be considered evidence for a 'true' quantity. However, as will be argued below, this kind of 'headcounting' is inherently fallacious. As stated above, tests which appear to reach agreement on plasmodes do not produce estimates with the same consistency on real data.

The use of the KG test for both PCA and IFA requires examination in that the simple logic of retaining factors with eigenvalues >1 appears to be in error when using PCA. Its use on IFA would conversely seem to be fully justified given the validity of a crucial assumption and the definition of a gramian matrix. Hakstian and Muller (1973) in their discussion of the KG note 'A widespread, but faulty, interpretation of the algebraic work (on bounds of the rank of a reduced correlation matrix) is that concerning Guttman's (1954) weaker lower bound, in which an elegant proof on the universally weakest lower bound to the rank of a gramian reduce correlation matrix was translated into a rule of thumb concerned with the number of principal components to retain !' Accepting Hakstian and Muller's argument, the only justification for using this test on PCA is Kaiser's (1960, 1965) logic bassed upon the coefficient α gcneralisability of factors. Defining coeficient α as :

$$\alpha = \frac{n}{n-1} \frac{(1-\sum_{i=1}^{n} V_i)}{V_t}$$

when n is the number of variables in a test scaie, V_t is the variance of the composite test scores, and V_i is the variance of the scores on individual variables i. Mulaik (1972) highlights a crucial assumption made by Kaiser in defining the variance of a test variable (p. 210) 'Since in most psychological problems the variance of a variable has no intrinsic meaning, we can assume that the variances of the n variables are arbitrarily set equal to 1.' This assumption reduces (9) to

$$\alpha = \frac{n}{n-1}\left(1 - \frac{1}{\lambda}\right)$$

where λ is the eigenvalue associated with a factor. With this assumption the test is applicable to both common and component factor analysis. One interesting point is that in classical test theory, the square root of coefficient α is the estimated correlation of a test with errorless true scores. Can it thus be said that the root α obtained using eighenvalues indicates the estimated correlation of the observed factor with the 'true' factor ? While Saville and Blinkhorn (1980) have related Cattell and Tsujicka's (1964) coefficient of factor validity (CFV) and the standard form of coefficient α, the interpretation of coefficient α when applied to eigenvalues will be generally unrelated to the values of the CFVs. This is due to the different computational emphases of the coefficients.

However, even assuming the above arguments hold, the Kg test is of little practical value if one continues to regard it as an eigenvalue >1 rule. As Cattell (1966) pointed out, unrotated eigenvalues change size after rotational transformations. Additionally. if a value of 0.7 for α and CFV is taken as an indication of a strong, generalisable, replicable factor, the task of retaining factors becomes simpler. (Note that coefficient α sets an upper limit to the reliability and that the CFV is a multiple correlation, thus a value of 0.7 is conservative). Therefore, using oblique rotation variance coefficients (Cattell, 1978) or orthogonal rotation eigenvalues, the computation of αs and CFVs will provide the basis for decision for either common or component factors. Noticeably, the MAP test achieves much the same results as using αs in terms of the number of factors retained. However, for practical purposes, the number of factors are decided by considering the underlying hypotheses of the study, the number of salient variabless loading a factor, the factor validity and replicability and coefficient. Of course, the 'tests' of factor quantity may be used as guides but must be treated as such. This methodology would finally end the proliferation of reported factors, from the factoring of items from tests, with only 3 or 4 salient loadings that are never replicated and appear to have no validity whatsoever.

While little attention has been paid to statistical extraction criteria, one method perhaps requires discussion within the context of this paper. This is the technique of CMLFA : confirmatory maximum likelihood factor analysis. (Joseskog and Lawley (1969), Joreskog (1967, 1969, 1978)). In contrast to the 'exploratory' factoring techniques used above, this method attempts to fit a linear factor solution to a target factor pattern matrix prescribed by the investigator. The target pattern matrix may contain specified parameters such as the number of factors and their inter correlations in addition to certain loadings. Having specified a target matrix and its associated parameters, the 'best fit' factor solution is then tested for statistical goodness of fit against the X^2 distribution. Thus the problem of factor extraction is now by passed indirectly by expressing it in terms of a global solution hypothesis test. Of course, the statistical criteria given above such as in Rippe (1953) and Anderson and Rubin (1956) remain essentially tests of factor extraction quantity. CMLFA, however, radically departs from this approach by allowing model significance testing directly by a factor analytic procedure. Given the assumptions of multivariate normality and maximum likelihood estimation procedures, a method now exists for apparently totally objective determination of factor analytic solutions. The use of 'apparently' is purposeful in that some subjectivity does enter even this technique. Given a very large sample with a large variable matrix, it has been demonstrated that a class of targeted solutions may be statistically significant, thus requiring the investigator to choose the most 'appropriate' solution. However, given the capability of making such a global hypothesis test, this particular issue can be viewed as being of somewhat secondary importance.

In conclusion, the data and arguments from the study above indicate that tests of factor extraction are an inappropriate answer to the question being asked by an investigator pursuing an exploratory factor analytic strategy. Estimates and lower bounds are appropriate but sometimes of little practical value when faced with a set of factors. An investigator requires factors that represent common variance, are replicable and have some validity outside of the mathematico-statistical framework. In classical test psychometric test theory, factors can be considered as representations of domains of items. Thus, attempting to interpret observed factors defined by a small number of items is inappropriate within this context. The factors are simply not representative of the domains, with all that implies in the use of such factors in any assessment situation. Thus the notion of under-factoring is moderated by general psychometric issues. Minimising the rank of matrix is a psychological as well as a statistical problem. Solutions based upon either method alone are rarely useful.

References

Anderson, T. W., and Rubin, H. Statistical Inference in factor analysis. *Proc. of the Thrid Berkeley Symposium on Mathematical Statistics and Probability.* 5, 1956 111-150.

Barrett, P.T., and Kline, P. Personality factors in the Eysenck Personality Questionnaire. *Personality and Individual Differences.* 1980, In Press (a).

Barrett. P. T., and Kline, P. An Item factor analysis of Cattell's 16PF inventory. 1980, In Preparation (b).

Barrett, P.T., and Kline, P. The location of superfactors P, E, and N within an unexplored personality factor space. *Personality and Individual Differences.* 1980, 1, 239-247 (c).

Barrett, P.T., and Kline, P. Radial Parcel factor analysis. 1980, In Preparation (d).

Barrett, P.T., and Kline, P. Radial Parcel factor analysis of Cattell's 16PF inventory. In Preparation (e).

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

Bachtoldt, H. P. An empirical study of the factor analysis stability hypothesis. *Psychometrika.* 1961, 26, 405-432.

Cattell, R.B. The Scree for the number of factors. *Multivariate Behavioural Research.* 1966. 1, 245-276.

Cartell, R.B. Radial Parcel factoring vs item factoring in defining personality structure in questionnaries : Theory and experimental checks. *Australian Journal of Psychology.* 1974. 26, 103-119.

Cattell, R.B. *The Scientific Use of Factor Analysis in Behavioural and Life Sciences:* New York : Plenum Press. 1978.

Cattell, R.B., and Dickman, K. A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin.* 1962, 59, 389-400.

Cattell, R.B., and Tsujioka, B. The importance of factor-trueness and validity.

versus homogeneity and orthogonality, in test scales. *Educational and Psychological Measurement*, 1964, 24, 3-30.

Cattell, R.B., and Vogelmann, S. A comprehensive trial of the Scree and KG criteria for determining the number of factors. *Multivariate Behavioural Research*, 1977, 12, 289-325.

Coan, R.W. Basic forms of covariation and concomitance designs. *Psychological Bulletin*, 1961, 58, 317-324.

Comrey, A.L. *Manual for the Comrey Personality Scales*. Educational and Industrial Testing Service, 1970.

Emmett, W.C. Factor analysis by Lawley's method of maximum likelihood. *British Journal of Psychology*. Statistical Section, 1949, 2, 90-97.

Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika* 1953, 18, 277-296.

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

Hakstian, A.R., and Muller, V. J. Some notes on the number of factors problem. *Multivariate Behavioural Research*, 1973 8, 461-475.

Harman, H.H. *Modern Factor Analysis (1st Edit.)*. University of Chicago Press. 1960.

Harman, H.H. *Modern Factor Analysis (3rd Edit.)*. University of Chicago Press, 1976.

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

Horn, J.L., and Engstrom, R. Cattell's Scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioural Research*, 1979, 14, 283-300.

Joreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 1967, 32, 443-482.

Joreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis *Psychometrika*, 1969, 34. 183-202.

Joreskog, K. G. Structural analysis of covariance and correlation matrices. *Psychometrika*, 1978, 43, 443-477.

Joreskog, K.G. and Lawley, D.N. New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1968, 21.

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

Kaiser, H.F. Image Analysis. Chapter 9 in C W. Harris (Ed.), *Problems in measuring change*. Madison : University of Wisconsin Press, 1963.

Kaiser, H.F. Psychometric approaches to factor analysis *Proceedings of the 1964 Invitational Conference on Testing Problems*. Princeton : Educational Testing Service, 1965.

Kaiser, H F., and Horst, P. A score matrix for Thurstone's box problem. *Multivariate Behavioural Research*, 1975, 12, 17-25.

Kline, P., and Storey, R. The Dynamic Personality Inventory: What does it measure ? *British Journal of Psychology.* 1978, 69, 375-383.

Linn, R.L. A monte carlo approach to the number of factors problem. *Psychometrika.* 1968, 33, 1, 37-71.

Lord, F.M. A study of speed factors in tests and academic grades. *Psychometrika.* 1956, 21, 31-50

Maxwell, E.A. Recent trends in factor analysis. *Journal of the Royal Statistical Society,* Series A, 1961, 124, 49-59.

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

Nesselroade, J.R., and Baltes P.B. Higher order convergence of two distinct personality systems : Cattell's HSPQ and Jackson's PRF. Paper presented at the meetings of the Society of Multivariate Experimental Psychology, Ft. Worth, Texas, November 1972.

Revelle, W., and Rocklin, T. Very simple structure : an alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioural Research,* 1979, 14, 403-414.

Rippe, D.D. Application of a large sampling criterion to some sampling problems in factor analysis. *Psychometrika,* 1953, 18, 191-205.

Saville, P., and Blinkhorn, S. Item and factor analysis of Cattell's 16PF Inventory, 1980, In Preparation.

Skinner, H.A., and Rampton, G.M. *Evaluation of the Personality Research Form (PRF) in a military environment.* Canadian Forces Personnel Applied Research Unit, April 1973.

Sokal, R.R., Rohlf, F.J., Zang, E., Osness, W. Reification in factor analysis : a plasmode based on human physiology of exercise variables. *Multivariate Behavioural Research,* 1980, 15, 181-202.

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

Thurstone, L.L. The dimensions of temperament. *Psychometrika,* 1951, 16, 11-20.

Tucker, L.R., and Lewis, C. A riliability coefficient for maximum likelihood factor analysis. *Psychometrika,* 1973, 38, 1-10.

Veldman, D.J. Simple structure and the number of factors problem. *Multivariate Behavioural Research,* 1974, 9, 191-200.

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