Person. individ. Diff. Vol. 3, pp. 259 to 270, 1982
Printed in Great Britain.

0191-8869 82 030259-12$03.00 0

AN ITEM AND RADIAL PARCEL FACTOR ANALYSIS OF THE 16PF QUESTIONNAIRE

PAUL BARRETT and PAUL KLINE

Department of Psychology, University of Exeter, Exeter EX4 4QG, Devon, U.K.

(*Received* 1 *October* 1981)

Summary—An extensive series of analyses were carried out on a sample of data from 491 undergraduate university students who completed Form A of Cattell's 16PF questionnaire. The data was item analysed, factored using both principal component and image analyses, and radial parcelled. However, even though five different factor solutions were rotated to a maximum simple structure, the 16 factors did not emerge as expected. Radial parcelling also yielded equivocal results. Using only psychometric criteria to guide the analysis, three new factor scales were generated that satisfied the test of high factor validity and high coefficient alpha simultaneously for each scale. The overall solution yielded seven factored scales. Additionally, results were reported of a scale factoring of the 16 scales yielding a replicable 4-factor solution. An alternative 7-factor solution was not replicable among subsamples taken from the total data set.

INTRODUCTION

The 16PF Questionnaire (Cattell *et al.*, 1970) is perhaps one of the most widely used psychometric instruments for the measurement of personality. The questionnaire was the outcome of Cattell's researches in the late 1940s and early 1950s, attempting to encompass the 'sphere' of personality initially defined by ratings. Howarth (1976) provides an excellent account of the detailed procedure adopted by Cattell in reducing 18000 dictionary terms relating to temperament to the 400 or so questionnaire items used in Forms A, B, C and D of the standardized questionnaire. Recently Cattell and Delhees (1973) have extended the number of factors to be found in the 16PF to 23, the supplemental scale information and augmented items provided by DeVoogd and Cattell (1973).

Recently, there have been many divergent findings reported by investigators who have attempted to replicate Cattell's factor structures. The more important of these are Levonian (1961a, b), Eysenck and Eysenck (1969), Howarth and Browne (1971), Comrey (1973) and Howarth (1976). Cattell has recently answered some of these criticisms with powerful arguments as in his reply to Eysenck (Cattell, 1972) and in his book on psychometric methods (Cattell, 1973). Invariably his points have turned on the methodology of these investigators, Cattell claiming that poor methodology has resulted in the divergent results. However, Cattell himself is not always consistent in his methods as Howarth (1976) shows quite clearly. In addition, Gorsuch and Cattell (1967) demonstrating the second-order scale factor-structure of the 16PF used parcels of items to assist in the factor procedures. These parcels, contrary to Cattell's own statements (Cattell, 1974; Cattell and Burdsall, 1975) on radial parcelling techniques, were *ad hoc* collections of items utilizing marker variables as parcel cluster centroids. This is most surprising given his vehement objections to such theoretically poor methods of clustering items.

The investigation below is an attempt to replicate the factor structure of the 16PF using methodology very similar to Cattell's. Both item and scale principal component and image analyses, radial parcel analyses, factor validity analyses and classical item analyses are undertaken. Although the data are from Form A only, this nevertheless is the most popular form being used by other investigators and occupational psychologists. It is of interest to discover whether the scales are still factorially valid and reliable within a student population. Results from Barrett and Kline (1980a) and Kline *et al.* (1980) using the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975) have demonstrated the efficacy of item factor analysis in delineating good replicable item factor scales. This is contrary to some investigators who believe that it is impossible to obtain reliable and replicable factor structures using single items as variables (e.g. Nunnally, 1978).

METHOD

Subjects

The sample was composed of 241 female and 250 male undergraduate university students. The 16PF Form A questionnaires were administered under group testing conditions with no more than 20 students completing the forms at any one time. Participation of students in this study was entirely voluntary.

Procedure

For scoring purposes, the S's responses were extracted from their answer sheets and coded on to punched cards. These data were then machine-scored yielding individual item and scale information. For both the male and female groups, means and SDs for each scale were computed. Following this, the male and female data were combined to form a sample of 491 Ss, this sample being used for all subsequent analyses.

Item analysis

Classical item analysis was then carried out on this data using uncorrected point biserial coefficients as estimates of item-total correlations. Additionally, coefficient alphas (Cronbach, 1951) were computed for each scale.

Principal component and image factor analysis

A principal component factor analysis (PCA) was implemented upon a 184 × 184 variable Pearson correlation matrix. As both the work of Velicer (1976a, 1977) and Nunnally (1978) demonstrates, with n (the number of variables) >20, very little difference exists between factor patterns computed from correlation matrices with either unities or communality estimates in the main diagonal. However, due to the sensitive nature of this particular study, image analysis (IFA: Guttman, 1953; Kaiser, 1963) was also undertaken in order to provide information as to the level of overestimation of the PCA loadings.

Factor extraction tests

For both the PCA and IFA three tests of factor extraction were undertaken:

(a) The Kaiser factor alpha criterion (Kaiser, 1960; Kaiser & Caffrey, 1965; Barrett and Kline, 1982). This criterion is based upon Kaiser's derivation of coefficient alpha for a factor. For each eigenvalue $\lambda_1, \lambda_2, \ldots, \lambda_n$, a coefficient alpha estimate of reliability can be computed using:

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

where n = No. of variables in the correlation matrix.

Thus, for any $\lambda_i < 1.00$, $\alpha\lambda_i$ is negative. However, unlike the now obsolete PCA Kaiser–Guttman criterion (Hakstian and Muller, 1973) factor extraction proceeds upon the basis of accepting and rotating factors with satisfactory alphas. What constitutes 'satisfactory' is, of course, dependent upon an investigator's particular interpretation of the term 'reliability'. However, very few investigators would accept a coefficient <0.3 as being of any particular value. For most sizes of λ_i of approx. 1.9–2.0 yield alphas just greater than 0.5. This is the bound value adopted within this particular study.

(b) The Velicer Minimum Average Partial Correlation test (MAP: Velicer, 1976b). Velicer introduced the MAP test for all types of component analyses. 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:

$$C = R - AA^1 \text{ (for PCA)} \tag{2}$$

where R = the sample correlation matrix or

$$C = (S^{-1} RS^{-1}) - AA^1 \text{ (for rescaled } R \text{ image factoring)} \tag{3}$$

where

$$S^{-1} = \sqrt{(\text{diag } R^{-1})}.$$

Thus the matrix of partial correlations is given by:

$$R_p = D^{-1/2}\, CD^{-1/2} \tag{4}$$

where

$$D = (\text{diag } C).$$

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

$$f_m = \sum_{i \neq j}\sum (r_{ij})^2(n(n-1)) \tag{5}$$

where r_{ij} is the element in row i and column j of the matrix R_p in (4).

The value of f_m is the average of the squared partial correlations after the first m factors are partialled out. The stopping point is the value of m for which f_m is at a minimum; f_m ranging between 0 and 1. The logic of the test is that as factors are partialled out, the value of f_m decreases to a minimum indicating that the factors are 'common'. When f_m 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 elements 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 than at f min: Velicer (1976b) demonstrates this with a simple example.

(c) The machine implemented Scree test (Cattell, 1966) known as AUTOSCREE (Barrett and Kline, 1982).

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 resemblance 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 vs 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 breaks! 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 principle, 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 Barrett and

Kline 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 has been 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 larger 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 summarizing the details, there are three control parameter fields: group size, angle deviation, and error deviation. The group size is varied from a minimum of three 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.

As a simple test of the efficacy of AUTOSCREE compared with that of both an experienced judge and a novice, 21 sets of eigenvalues were computed from data matrices containing from 8 to 90 variables. Three of the matrices were plasmodes, the other 18 were real data matrices. (See Barrett and Kline, 1982, for information concerning these matrices). The experienced judge was the second author of this paper, who has used the Scree test for many years and worked with Cattell himself. Thus the 21 sets of computer plotted eigenvalue series, each set comprising a $1{:}1 \times \lambda_i$, $1{:}2 \times \lambda_i$, and $1{:}3 \times \lambda_i$ ratio plots were presented to each judge individually. No information as to the origin of the eigenvalues was given to either judge. The results of this procedure are given in Table 1. The error data was computed by subtracting the judge's cutoff for a particular eigenvalue set from the AUTOSCREE cutoff. As can be seen from Table 1 the AUTOSCREE algorithm provides results similar to that of an experienced judge.

Factor rotation

From the consideration of the results from these three tests, the retained factors were rotated using a modified direct oblimin (Jennrich and Sampson, 1966) procedure, with the δ parameter swept from −30.0 to 0.5 in steps of 0.5. The convergence criterion was set at 0.00001 with a maximum of 400 iterations per value of δ. The associated overall hyperplane count (HC: ± 0.1) for each of these rotations was noted, the appropriate solution being given by the maximum HC and its associated δ. Then the rotation was again carried out around this δ value in steps of 0.1 to 'fix' the solution. (Direct oblimin has the attractive characteristic that obliquity can be varied from near orthogonality to strong obliquity by varying the parameter δ, thus the solution is virtually unconstrained by the rotation method).

Table 1. Human judgement vs AUTOSCREE

	Mean error	Mean squared error	SD of errors
Experienced judge	0.6	3.9	1.9
Novice	−3.0	22.6	3.7

Factor validity

Finally, factor validity coefficients (Cattell and Tsujioka, 1964) were computed for each of the factor scales A through Q4. The factor validity coefficient of a scale or set of items may be defined as the ratio of mean validity (mean item-factor correlation) to mean homogeneity (mean inter-item correlation). The factor validity coefficient is, in fact, a multiple correlation coefficient, thus it can be viewed as indirectly assessing the similarity of a set of items that load upon a factor. If the items are no more than reworded counterparts of one or two basic items, then their homogeneity is likely to be high with the result that the mean item-factor correlation will reach a ceiling value. The factor validity will then be low. Of course, in the converse limit, the factor validity will also be low. In this way, the coefficient can be used to assess the validity of any set of items that the investigator intends to call a factor scale or at least an item group that measures a common component.

Radial parcel factor analysis

In addition to the item factor analyses, radial parcel factor analysis (Cattell, 1974; Cattell and Burdsall, 1975; Barrett and Kline, 1981) was implemented using the full 184 × 184 variable intercorrelation matrix. The procedure simply forms parcels of items and factors the resulting intercorrelation matrix generated from the parcel scores. Parcels in this study were of size 2- and 4-items. The 2-item parcelling procedure was as follows: from an unrotated V_0 retained factor pattern matrix of item loadings congruences between each row (the item vector) and all others were computed. The highest absolute value congruence coefficient indicated the formation of the first 2-item radial parcel. These two items were then eliminated from any further searching. That is, all congruences computed using either of these two items were removed from the congruence matrix. This simple search procedure continued until all items were parcelled. Although the absolute congruence values were used in the search procedure, the sign of this value indicated a possible reflection of scoring for one of the variables in order to make the congruence positive. Thus, when rescoring the data prior to factor analysis, the addition of the values of two items was moderated by a reflection constant of −1 or +1 operating on one item's response. Having obtained the rescored 2-item parcel data, these 92 parcel variables were submitted to PCA, followed by AUTOSCREE and the MAP test, and subsequently the direct oblimin rotation procedure as indicated above.

In order to obtain 4-item parcels, one can either operate the above procedure on the unrotated 2-item parcel V_0 matrix, or proceed directly from the 2-item parcel identification stage operated on the unrotated V_0 item matrix. For the purpose of this analysis, the latter procedure was adopted. Having identified the 2-item parcel components and having noted the sign of the congruences, the task is now of one generating a new composite V_0 matrix. By adding or subtracting the loadings across factors for each of the two items in a parcel, a new row in the composite V_0 is generated for that parcel variable. If the congruence between rows was positive, add the row values, if negative subtract all values from one another. Thus a 45 parcel variable composite V_0 matrix is found. Generally, parcelling of items reduces individual item measurement error and thus provides a more clear representation of the factor structure. Full details of the technique are given in Barrett and Kline (1981).

Before any tests of factor extraction were made, it was decided to extract and rotate 16 factors, and as in Cattell's (1972) reply to Eysenck, 19 factors. Every effort was made in this study to attain the factorial structure represented by Cattell's 16 factor scales.

Table 2. Scale means and SDs

Scale	Male (N = 250) Mean	SD	Female (N = 241) Mean	SD
A	9.312	3.192	10.129	3.023
B	9.428	1.827	9.390	1.955
C	14.032	4.071	13.701	3.934
E	14.336	4.617	12.531	4.146
F	15.636	5.187	16.436	4.634
G	9.952	4.283	9.929	3.733
H	13.116	6.401	13.427	5.538
I	9.824	4.034	13.672	3.410
L	9.632	3.366	8.826	3.647
M	14.576	3.520	14.639	3.600
N	8.404	2.965	8.846	2.776
O	10.884	4.458	11.809	3.848
Q1	10.848	3.323	9.183	3.110
Q2	11.800	3.317	11.461	3.477
Q3	10.160	3.577	9.697	3.416
Q4	13.132	5.185	15.004	4.782
Age	23.224	8.286	22.203	6.762

RESULTS AND DISCUSSION

Scale means and SDs

Table 2 presents the means and standard deviations for each of the 16 Form A scales taken over the 491 *Ss*. This table indicates that by no standards could the scores of our student group be regarded as abnormal.

Item analysis and coefficient alphas

Following the scoring of the data, an item analysis was initiated. The results from this were indicative of generally non-homogenous item-factor scales. This is demonstrated in Table 3 (below) which presents the alpha coefficients for each of the 16 scales. For those coefficients less than about 0.4, some of the scale items correlated less than $1/\sqrt{n}$ (the average correlation of all the n items in a scale with the total scale score). Noticeably, the O and Q4 items were predicting both their own and the O or Q4 scale score. Nunnally (a severe psychometrist) has asserted that

> "... in many applied settings, a reliability of 0.80 is not nearly high enough. In basic research, the concern is with the size of correlations and with the differences in means for different experiment treatments, for which purposes a reliability of 0.80 for the different means involved is adequate. In many applied problems, a great deal hinges on the exact score made by a person on a test..." (1978, p. 245)

If this view is accepted, all the 16PF Form A scales are of little or no practical value, they remain useful for research purposes only. Saville and Blinkhorn's (1976) alternate forms reliabilities are no better, even with their very large sample sizes. The obvious way to improve the reliability estimates is to include Form B scales with the Form A scales thus doubling the scale lengths. However, this of course assumes that Form B is in fact an alternate form of A. Nevertheless, as Cattell has frequently noted, low homogeneity or internal consistency is not necessarily a bad property of a scale, given that its factor validity is high. These validities will be examined below.

Principal components and image factor analysis

The PCA and IFA of the data was subsequently implemented. Both the MAP and AUTOSCREE tests indicated 11 factors from both methods of analysis. The PCA factor alpha for 11 factors = 0.58, for 12 = 0.53 and for 16 = 0.49. The Kaiser–Guttman lower bound—strictly applicable only to image eigenvalues (because of the IFA properties of a

Table 3. Coefficient alpha and validities for the 11, 16 and 19 factor solutions

Scale	Coefficient alpha	16 factor $N = 491$ Validity	Position	19 factor $N = 780$ Cattell validity	Position	19 factor $N = 491$ Validity	Position	11 factor $N = 491$ Validity	Position
A	0.39	0.42	4	0.63	1	0.42	1	0.42	4
B	0.44	0.70	10	0.61	2	0.70	15	0.71	10
C	0.52	0.66	1	0.55	3	0.48	3	0.69	1
E	0.59	0.56	2	0.54	4	0.54	1	0.62	2
F	0.70	0.77	3	0.77	5	0.76	2	0.73	4
G	0.59	0.65	7	0.69	6	0.61	9	0.70	3
H	0.81	0.76	2	0.79	7	0.75	1	0.77	2
I	0.64	0.84	5	0.74	8	0.83	6	0.84	5
L	0.45	0.60	4	0.45	9	0.65	4	0.56	6
M	0.28	0.39	7	0.52	10	0.40	9	0.40	3
N	0.19	0.43	8	0.41	11	0.45	8	0.46	11
O	0.56	0.73	1	0.66	12	0.59	7	0.77	1
Q1	0.40	0.59	6	0.48	13	0.60	5	0.55	8
Q2	0.41	0.49	3	0.57	14	0.59	13	0.52	4
Q3	0.47	0.46	7	0.64	15	0.47	11	0.48	3
Q4	0.69	0.81	1	0.73	16	0.67	3	0.80	1

gramian reduced rank correlation matrix)—indicated 19 factors. Thus five solutions were rotated: the 11-factor PCA and IFA V_0 matrices, the 16-factor PCA V_0 matrix and the 19-factor PCA and IFA V_0 matrices. Generally, the only differences between the PCA and IFA solutions were in the second decimal place of the loadings. No gross loading pattern changes were observed. The δ values and hyperplane count percentages for each of the solutions were as follows: PCA 11 factors = δ of 0.65, 59%, IFA 11 factors = δ of 0.45, 66%, PCA 16 factors = δ of 0.45, 64%, PCA 19 factors = δ of 0.2, 67%, IFA 19 factors = δ of 0.3, 73%. The value of 73% of loadings in the hyperplane for the IFA 19-factor solution compares favourably with Cattell's (1972) value of approx. 75% for his data. Both values being significant $P < 0.01$ by the Bargmann test (1953).

However, the 16- and 19-factor pattern matrices do not represent a clear 16 scale structure. Only the I and G scales appear to be loaded clearly by their scale items. The remaining 14 scales are impossible to ascertain when attempting to interpret the factors by accepting loadings $> |0.30|$ as being indicative of significance. Thus the alternative method of interpreting these matrices is by the use of factor validities. All solutions were checked for difficulty factors based upon concordant item response splits; there were no recognizable difficulty factors.

Factor validities

Table 3 above presents the factor validity coefficients for the sample data calculated from the 16 and 19 PCA factor pattern rotated solutions. In addition, Cattell's (1972) factor validities for Form A from his sample of 780 adults are presented for comparison. (His validities are based upon a principal factor analysis loading pattern.) The multiple correlation coefficients were left uncorrected (Claudy, 1978) as the effects of such corrections are minimal given the sample size. The calculation of the factor validities proceeded by taking each questionnaire scale in turn, and calculating its validity across all factors in the relevant pattern matrix. The highest validity coefficient for that scale on a particular factor was taken as the scale/factor validity. Thus the column of figures (Table 3) entitled POSITION provide the factor number on which the scale has most validity. Notably, all other validity coefficients were very low relative to each scale's maximum.

Thus, in comparison with Cattell's coefficients, a superficial similarity is noted—however, regarding the factor positions at which these maximum values occur, it appears that many of the scales are not uniquely valid. For example, scales C, O and Q4 for the 16 factor pattern all have their highest validity on factor 1. Cattell's (1972) factor positions were taken from the order in which he presents his factor pattern matrix. Noticeably, all his coefficients are uniquely determined.

Ignoring these positional effects, it is pertinent to ask what size of coefficient should be regarded as conceptually significant. Taking values greater than 0.6 as being indicative of some validity, factors C, E, L, M, N, Q1 and Q2 in Cattell's pattern fail. In the 16 factor $N = 491$ sample solution, A, E, M, N, Q1, Q2, and Q3 fail, and in the 19 factor $N = 491$ sample solution, A, C, E, M, N, O, Q2 and Q3 fail. Of course, the value of 0.6 is arbitrary but nevertheless this problem of conceptual significance of these coefficients requires careful consideration.

Given the above results, it was concluded that the 16 and 19 factor solutions had failed to yield Cattell's 16 item factors. Thus, 11 PCA factors (as suggested by both the MAP and AUTOSCREE factor extraction test results) were rotated to a maximum simple structure. Once again, simple visual interpretation of the pattern matrix was impossible, especially when attempting to locate the 16 factor scales. Thus recourse was sought in the factor validities for the 16 scales. (See Table 3). Given that searching for 16 scales among 11 factors will yield some validity coefficient overlap, the amount encountered once again invalidated the solution.

Radial parcel analyses

At this stage, it was decided that radial parcelling of the 11 and 16 factor unrotated V_0 matrices might yield better results than those from any item factoring. Thus size 2 and 4 parcels of items were generated from these matrices. However, having obtained the item parcels, it was obvious that the procedure paired off the items that loaded 'highly' on each item factor such that any parcel solution rotations would yield images of the item factors. This was expected given the results of Barrett and Kline (1980b), however, with the unknown quantity of individual item error involved, it was possible that some 'sorting' would have taken place.

From a consideration of all the results presented so far, it was concluded that Cattell's 16 factors were not represented in this sample data using Form A of the 16 PF questionnaire.

Statistically determined factoring

Thus attempts were made to extract factors using statistical and psychometric criteria, ignoring Cattell's hypothesized structure. These criteria were that factors must yield good coefficient alphas, and high factor validity coefficients, and their items must show clear item total correlations in a classical item analysis.

The initial attempt accepted the 11-item factor pattern above as indicative of 11 new scales. All loadings in this pattern $> |0.3|$ were taken as scale items. The new scales thus formed contained a few items that were not unique, however, this number was minimal. Subsequent item analysis of these 11 new scales yielded extremely poor coefficient alphas. PCA factoring of the correlation matrix yielded AUTOSCREE and MAP test results of nine and six factors, respectively. Although these solutions were rotated, the results of both were, because of very low (even negative) alphas and equivocal factor extraction results, discarded.

The next step was to return to the factor validity coefficients in Table 3. Using these as the prime indicators of possible scales, two possible sets of scales were composed. Of course, these scales would be composed from whole 16PF questionnaire scales, either appended to one another or single scales as in the original questionnaire form. The choice of the new scales was determined by the coefficient alpha, factor validity, and the sharing of maximum validity coefficients among scales across the 11, 16 and 19 PCA factor solutions. From these considerations, seven new scales were chosen for subsequent analysis: (1) C + O + Q4; (2) E + H; (3) G + Q3; (4) I; (5) L; (6) B; (7) Q1. Thus scales A, F, M, N and Q2 were discarded from this analysis. Scales A, M and N had alphas and factor validities of such low values that no serious consideration could be given to their continued use. Scales F and Q2 present a different problem. Q2 has a very low alpha of only 0.41 while its overall validity is not much higher. However, it shares its validity with the F scale which has both a high alpha (0.70) and high average validity (0.75). This

Table 4. Coefficient alphas and factor validities for the seven new scales

C scale missing				C scale inclusive			
Scale	Coefficient alpha	Factor validity	Position	Scale	Coefficient alpha	Factor validity	Position
O + Q4	0.73	0.83	2	C+O+Q4	0.34	0.58	1
E + H	0.81	0.83	1	E + H	0.81	0.81	2
G + Q3	0.71	0.87	3	G + Q3	0.71	0.84	3
I	0.64	0.86	5	I	0.64	0.81	5
L	0.45	0.54	6	L	0.45	0.52	6
Q1	0.40	0.54	4	Q1	0.40	0.43	4
B	0.44	0.76	7	B	0.44	0.71	7

appears to be the same situation as for the scales G and Q3, however, the Q3 alpha (0.47) is higher—suggesting a psychometrically more sound set of items. Thus while G and Q3 were retained (a borderline decision), F and Q2 were eliminated from this analysis.

The raw data was thus rescored and subsequently item analysed. The coefficient alphas for these seven new scales are given in Table 4. The worrying feature here was the behaviour of the C scale items embedded within the C + O + Q4 composite scale. These items were not correlating in any significant fashion with the total scale score, thus apparently reducing the scale alpha quite considerably. However, it was decided to proceed with a PCA and direct oblimin analysis on the 128 variables, obtaining factor validities accordingly. Factor extraction test results were equivocal with the MAP test yielding a decision of 6 components, AUTOSCREE a value of 10, with the factor alphas ranging from $6\alpha = 0.59$, $9\alpha = 0.51$, $10\alpha = 0.48$, $11\alpha = 0.47$. Given that an attempt was being made to find seven factor scales, the extraction test results were overridden. Seven factors were rotated to a maximum simple structure and factor validities computed accordingly. The associated direct oblimin δ and overall hyperplane count percentage were: $\delta = 0.5$, 53%. The factor validities are reported in Table 4. The coefficients for scales E + H, G + Q3, I and B are satisfactory. However, the values for scales C + O + Q4, L, and Q1 are disappointing. Given the equivocal factor extraction test results and the correspondingly poor alpha and validity coefficients for the C + O + Q4 composite, it was decided to discard the C scale altogether and begin again with the remaining 115 items.

An item analysis of the composite O + Q4 scale yielded a much improved coefficient alpha of 0.73. The 115 variables were then PCA factored. The MAP test indicated retention of seven components, AUTOSCREE indicated 6 with the factor alphas ranging from 6 = 0.58, 7 = 0.54, 8 = 0.50, 9 = 0.47. These results provided sound evidence for the retention and subsequent rotation of seven components. The direct oblimin δ and overall hyperplane count percentage were: $\delta = -0.7$, 55%. The resulting factor validities are reported in Table 4. The coefficients for scales O + Q4, E + H, G + Q3, I and B are most satisfactory, those for L and Q1 are improved from the $n = 128$ variable factoring but still a little disappointing. However, the results as a whole for this set of scales are generally satisfactory.

At this stage of the investigation, it was decided to postpone any further factor investigations until a new sample of 16PF data could be factored, the results compared to those above, and some evidence of replicated factorial structure found.

One final result relevant to these findings is some scale factoring of these data reported in Barrett and Kline (1980b) where the effects of subject to variable ratios in factor analysis were examined. The results again failed to support the hypothesized second-order structure of the 16PF. The study required the compilation of six randomly drawn samples from the 491 *Ss* having *S* to variable ratios of $1\frac{1}{4}$:1, 2:1, 3:1, 6:1, 12:1, 18:1. The total sample data from the 491 *Ss* was initially PCA scale factored. It was not clear, from the factor extraction tests, just how many factors should be rotated. The MAP test indicated only 2, AUTOSCREE indicated 7 and the factor alpha coefficients suggested 4. It was decided to rotate both the 7 and a 4 factor solution. These two solutions would

Table 5. Subsample MAP and AUTOSCREE decisions

Sample ratio	MAP	AUTOSCREE
$1\frac{1}{4}$:1	2	7
2:1	4	8
3:1	4	9
6:1	2	4
12:1	4	6
18:1	4	7 or 4

then be compared with the subsample factor patterns. From the final factor patterns, Pearson correlations and Tucker congruences were computed between the total sample and all subsample factors (reflecting and correcting for factor position where necessary). Because the underlying issue in this study was stability of factors—only 4 and 7 factor solutions were rotated for each subsample pattern matrix. Table 5 presents the MAP and AUTOSCREE results for each subsample data set. Having thus rotated all 4 and 7 factor V_0 matrices, the associative analysis was undertaken. Coefficients >0.75 were taken as indicative of solution similarity. From these results it was obvious that the 7-factor solution was unsatisfactory. This was viewed as a direct result of overfactoring. (The values in Table 5 for the MAP especially and the AUTOSCREE appear to indicate a consistent 4 factor structure.) This overfactoring suggests that a large quantity of error variance is being distributed across the solutions leading to factor instability. The factor alpha coefficients for the rotated 6th and 7th factors from the total sample analysis were very low indeed: 6 = 0.27, 7 = 0.09. Therefore, this particular solution was discarded in place of the 4-factor representation.

This 4-factor solution was found to be extremely stable in that the subsample factor patterns matched the total sample pattern very closely with associative coefficients >0.80 for all subsamples except the $1\frac{1}{4}$:1 ratio data. The total sample ($N = 491$) factor pattern is given in Table 6. This result obviously presents an intriguing problem. The very scales that were discarded from analyses above are now loading highly on these second order factors. The scales do not appear as item factors yet behave as good scales when all the items are used as a composite variable. Noting above that radial parcelling failed to yield clarity of the item factor scales, it appears that for meaningful measurement, only the whole scale score is valid. Yet, the factor validity coefficients are consistently low in all rotations. Until the new sample data is analysed—further speculation is pointless. In

Table 6. 4-factor scale pattern for 16PF ($N = 491$)

Scale	Factor 1	Factor 2	Factor 3	Factor 4
A	0.094	0.280	0.732	0.067
B	0.216	0.011	−0.150	−0.219
C	0.807	0.046	0.099	−0.017
E	0.023	−0.718	0.266	−0.100
F	−0.021	−0.281	0.679	−0.231
G	0.176	0.188	0.043	0.720
H	0.291	−0.313	0.682	−0.054
I	−0.122	0.562	0.281	−0.576
L	−0.481	−0.528	0.150	0.073
M	0.323	−0.093	−0.058	−0.617
N	0.011	0.609	−0.033	0.088
O	−0.775	0.142	−0.095	−0.040
Q1	0.153	−0.652	−0.052	−0.120
Q2	0.081	0.043	−0.677	−0.106
Q3	0.474	0.087	−0.035	0.658
Q4	−0.820	0.092	−0.008	−0.139

Loadings > |0.4| are underlined.

contrasting this factor pattern with that reported by Gorsuch and Cattell (1967), factors 1 and 3 in Table 6 are similar to their factors I and II. However, the last three factors of Gorsuch and Cattell are not really justified as second orders. With only one or two loadings marking each 'factor', it suggests that overfactoring has been carried out. Extracting eight instead of five factors will necessarily lead to this result, artificially splitting broad factors into smaller composites.

CONCLUSIONS

An extensive factorial study of a sample ($N = 491$) of data on Form A of the 16PF has not yielded 16 factor scales. Rather, the results have suggested that between seven and nine factors will have both satisfactory coefficient alphas and factor validities. The psychological meaning of the factors generated within this study has not been discussed—such a task is premature given the relatively small sample size. A new sample of data is being obtained by the authors to cross check the above results. From the results of the new data and the old, a better picture of the factor structure of the items contained within the 16PF questionnaire may emerge.

REFERENCES

BARGMANN R. (1953) The statistical significance of simple structure in factor analysis. *Frankfurt am Main. Hochschule Pedagogische Forschung.*

BARRETT P. T. and KLINE P. (1980a) Personality factors in the Eysenck Personality Questionnaire. *Person. individ. Diff.* **1**, 317–333.

BARRETT P. T. and KLINE P. (1980b) The observation to variable ratio in factor analyses. *Person. Group Behav.* **1**, 23–33.

BARRETT P. T. and KLINE P. (1981) Radial parcel factor analysis. *Person. individ. Diff.* **2**, 311–318.

BARRETT P. T. and KLINE P. (1982) Factor extraction: an examination of three methods. *Person. Group Behav.* In press.

CATTELL R. B. (1966) The Scree test for the number of factors. *Multivar. behav. Res.* **1**, 140–161.

CATTELL R. B. (1972) The 16PF and basic personality structure. A reply to Eysenck. *J. behav. Sci.* **1**, 169–187.

CATTELL R. B. (1973) *Personality and Mood by Questionnaire.* Jossey-Bass, London.

CATTELL R. B. (1974) Radial parcel factoring vs item factoring in defining personality structure in questionnaires: theory and experimental checks. *Aust. J. Psychol.* **2**, 103–119.

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

CATTELL R. B. and BURDSALL J. R. (1975) The radial parcel double factoring design: a solution to the item vs parcel controversy. *Multivar. behav. Res.* **10**, 165–179.

CATTELL R. B. and DELHEES K. H. (1973) Seven missing and normal personality factors in the questionnaire primaries. *Multivar. behav. Res.* **8**, 173–194.

CATTELL R. B. and TSUJIOKA B. (1964) The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales. *Educ. psychol. Measur.* **24**, 3–30.

CATTELL R. B. and VOGELMANN S. (1977) A comprehensive trial of the Scree and KG criteria for determining the number of factors. *Multivar. behav. Res.* **12**, 289–325.

CATTELL R. B., EBER H. W. and TATSUOKA M. M. (1970) *Handbook for the Sixteen Personality Factor Questionnaire.* IPAT, Champaign, Illinois.

CLAUDY J. G. (1978) Multiple regression and validity estimation in one sample. *Appl. psychol. Measur.* **2**, 595–607.

COMREY A. L. (1973) *A First Course in Factor Analysis.* Academic Press, New York.

CRONBACH L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–334.

DEVOOGD T. and CATTELL R. B. (1973) *The Seven Factor Supplement to the 16PF.* IPAT, Champaign, Illinois.

EYSENCK H. J. and EYSENCK S. B. G. (1969) *Personality Structure and Measurement.* Routledge & Kegan Paul, London.

EYSENCK H. J. and EYSENCK S. B. G. (1975) *Manual of the Eysenck Personality Questionnaire.* Hodder & Stoughton, London.

GORSUCH R. L. and CATTELL R. B. (1967) Second stratum personality factors defined in the questionnaire realm by the 16PF. *Multivar. behav. Res.* **2**, 211–223.

GUTTMAN L. (1953) Image theory for the structure of quantitative variates. *Psychometrika* **18**, 277–296.

HAKSTIAN A. R. and MULLER V. J. (1973) Some notes on the number of factors problem. *Multivar. behav. Res.* **8**, 461–475.

HOWARTH E. (1976) Were Cattell's 'Personality Sphere' factors correctly identified in the first instance? *Br. J. Psychol.* **67**, 213–230.

HOWARTH E. and BROWNE J. A. (1971) An item factor analysis of the 16PF. *Person. int. J.* **2**, 117–139.

JENNRICH R. I. and SAMPSON P. F. (1966) Rotation for simple loadings. *Psychometrika* **31**, 313–323.

KAISER H. F. (1960) The application of electronic computers to factor analysis. *Educ. psychol. Measur.* **20**, 141–151.

KAISER H. F. (1963) Image analysis. In *Problems in Measuring Change.* Chapter 9. (Edited by HARRIS C. W.). Univ. of Wisconsin Press.

KAISER H. F. and CAFFREY J. (1965) Alpha factor analysis. *Psychometrika* **30**, 1–14.

KLINE P., BARRETT P. T. and SVASTE-XUTO B. (1980) Personality traits of Thai students. *J. soc. Psychol.* In press.

LEVONIAN E. (1961a) A statistical analysis of the 16 Personality Factor questionnaire. *Educ. psychol. Measur.* **21**, 589–596.

LEVONIAN E. (1961b) Personality measurement with items selected from the 16PF questionnaire. *Educ. psychol. Measur.* **21**, 937–946.

NUNNALLY J. C. (1978) *Psychometric Theory*, 2nd edn. McGraw-Hill, New York.

SAVILLE P. and BLINKHORN S. (1976) *Undergraduate Personality by Factored Scales*. National Foundation for Educational Research.

VELICER W. F. (1976a) The relation between factor score estimates, image scores, and principal component scores. *Educ. psychol. Measur.* **36**, 149–159.

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

VELICER W. F. (1977) An empirical comparison of the similarity of principal component, image, and factor patterns. *Multivar. behav. Res.* **12**, 3–22.