Person. individ. Diff. Vol. 7, No. 3, pp. 327–340, 1986
Printed in Great Britain.

0191-8869/86 $3.00 + 0.00

FACTOR COMPARISON: AN EXAMINATION OF THREE METHODS

PAUL BARRETT
Department of Psychology, Institute of Psychiatry, Denmark Hill, De Crespigny Park, London SE5 8AF, England

(*Received* 9 *July* 1985)

Summary—Three coefficients of factor similarity were examined with regard to their behaviour within four sets of data. Two simple methods using Pearson *r* correlations and Tucker Congruence coefficients were compared with a more complex method given by Kaiser, Hunka and Bianchini (1971). Three of the data sets involved the use of 100 random data matrices, the fourth was that provided by the Eysencks' work on cross-cultural differences in personality using the EPQ. Drawbacks with each coefficient were apparent from the results, with the Kaiser *et al.* coefficient being capable of the most misleading results overall. However, use of the mean solution cosine in addition to the variable pair cosines was suggested as a method of validating the Kaiser *et al.* coefficient. It was concluded that using the three coefficients simultaneously as a multiple indicator yielded the best solution to the problem. In addition, it was suggested that other psychometric indicators should be employed to increase the degree of certainty of factor similarity.

INTRODUCTION

A common problem in factor-analytic studies is that concerned with comparing factors found in one study with those from another. Given the same variables but different observations on each variable, it is possible to compare the factors found from such studies in an effort to determine their similarity. This is important for ascertaining evidence of replication of results and, in the instance of repeated observations over time, evidence of change or factor modification.

There are four methodologies that may be said to encompass factor-comparison techniques. Firstly, there is the confirmatory maximum likelihood procedure put forward by Joreskog *et al.* (Joreskog, 1966, 1969, 1971; Joreskog and Gruvaeus, 1967; Joreskog, Gruvaeus and Van Thillo, 1970; Joreskog and Sorbom, 1982). This procedure extracts factors with predetermined characteristics from a set of data and then tests to determine if the residual correlation matrix still contains significant variance. A target solution is hypothesized that can specify the number of factors, the relationships between the factors and the sizes of the factor loadings themselves. The confirmatory solution will be a best fit to this target. One major benefit of maximum likelihood confirmatory analysis is that fit may be assessed using the χ^2-distribution to determine the statistical significance of the solution. Using this technique for factor-comparison purposes is perhaps the most sophisticated statistical method that could be adopted. Given that the target solution is exactly specified using all available parameter data, then the best-fit solution may be said to be statistically equivalent or not (as the case may be) at a specified level of significance. Probably, exact loading specification is not too important so long as all salient factor variables are specified to about the first decimal place as with the factor correlations (non-salient variables can be set to zero or near zero). Unfortunately, there are two major drawbacks to the routine use of confirmatory analysis. Firstly, the adequacy of the χ^2-test to indicate significance of fit is doubtful (Harris and Harris, 1971; Schonemann, 1981). While the test appears to give an upper bound to the number of factors that could be extracted from a set of data, this is not particularly helpful in the use of the test to ascertain factor similarity for a set of specific factors. Secondly, the program for confirmatory analysis (LISREL V; Joreskog and Sorbom, 1982) requires an inordinate amount of computer storage and time for data sets larger than about 40 variables. This precludes the general use of the technique with most sets of questionnaire data.

Finally, there are a series of procedures that may also be subsumed under the heading of confirmatory analysis, but do not have the benefit of an underlying distribution theory enabling statistical tests of fit. These procedures rotate a factor structure or pattern to a best-fit solution in comparison with a target factor structure or pattern, bearing in mind that the rotated factors must still adequately reproduce the original correlation matrix. The rotational algorithms are known as Procrustes solutions (Hurley and Cattell, 1962; Borg, 1978; Browne and Kristoff, 1969; Hakstian, 1970). If the target matrix is not specified in a thorough manner, then these methods are quite likely to produce best-fit solutions that may fit even random data target factor matrices (Horn and Knapp, 1973). Given a best-fit matrix is achieved, the problem still remains as to how similar the resultant solution is to the target, in other words, how 'best' is 'best'! Coefficients for determining this similarity are given below. Working with the residual correlation matrix by expressing variance unaccounted for by the rotated solution as a mean value etc. is not very useful. This form of parameter is too gross to allow sufficient accuracy of interpretation of similarity.

The second methodology of factor comparison is that typified by the coefficient presented by Kaiser, Hunka and Bianchini (1971). In this procedure, the variables defined by a set of factors are rigidly rotated against the target set of variables defined by the target factors. The variables are projected into the target factor space and rotated so that the cosine between each variable's vector in the 'experimental' matrix and the target matrix is maximized. The mean cosine from all such variable pairs provides some information on the suitability of assessing factor similarity. If it is low, then we may conclude that the measurement error (whether random or systematic) on most of the variables is so high that capitalization on chance in the computation of the factor similarities would be excessive. Given a high mean solution cosine via the rigid rotation of the variables in the target factor space, the cosines between all factor vectors in the 'experimental' data and those in the target data may be measured. These cosines are then interpreted as correlation coefficients that indicate the similarity between factors in two studies. This method effectively 'undoes' both target and 'experimental' oblique factor matrices and reduces them both to orthogonal form. The rigid rotation then takes place within this Cartesian framework. As such, the technique is generally insensitive to any prior rotational transformation carried out on the data so long as such a rotation was reasonably optimized to some criterion (e.g. Varimax, Promax, Maxplane, Direct Oblimin, Harris–Kaiser Orthoblique). Of course, the technique could be used to compare unrotated matrices or one rotated and one unrotated matrix. In addition, this method does not require that all variables are identical between two studies; however, there are severe conceptual problems to be overcome in this particular use of the method. The work by the Eysencks (Eysenck and Eysenck, 1983) on the cross-cultural similarity of personality as defined by the EPQ (Eysenck and Eysenck, 1975) questionnaire factors has made sole use of this coefficient to adjudge factor similarity.

The third methodology encountered in the area of factor comparison is that based upon the correlation of factor scores computed from two or more studies. Given the problems of computing factor scores from common factor models, it is not surprising that several alternative techniques exist to perform the basically simple task (Ahmavaara, 1954; Hamilton, 1967; Hundelby, Pawlik and Cattell, 1965; Hurley and Cattell, 1962; Mosier, 1938, 1939). The methodology is very simple when applied to component factor analysis, although there is no published research to date that has fully compared the behaviour of similarity coefficients computed in this way with other coefficients of similarity.

Lastly, the fourth methodology is concerned with calculations made directly on the factor loading vectors alone, immaterial of the relationships between the factors within any single study. Typical procedures are the use of Pearson correlations calculated between all pairs of factor vectors between studies. Also, Tucker's (1951) Congruence coefficient (see also Burt, 1948; Wigley and Neuhaus, 1955) is computed in likewise manner. A problem associated with the Pearson coefficient is that it is insensitive to the size of the factor loadings being compared, rather it assesses similarity as a function of the proportionality existing between two vectors. Likewise, the Congruence coefficient tends to overestimate similarity when the sign of all loadings is predominantly the same and the level of the loadings is approximately the same (Pinneau and Newhouse, 1964). The salient variable similarity index (Cattell, 1949, 1978), is a test of significance for determining whether or not two factors match in the sense that they have the same salient variables. Several variables could load the same factor by chance, so the question arises as to whether or not a sufficient number of the identical variables load the two factors for it to be assumed that such parallel loadings could have occurred by chance. Significance tables for the coefficient are provided by Cattell *et al.* (Cattell, Balcar, Horn and Nesselroade, 1969; Cattell, 1978). One drawback with this coefficient is that the factor loadings are treated as ordinal values rather than ratio measure values; hence some information is lost in the computation of the index. However, it does have a good rationale that approximates the way investigators tend to regard factor vectors, i.e. the tendency to only regard loadings >0.3 as being salient. Finally, Cattell, Coulter and Tsujioka (1966) introduced the pattern similarity coefficient that attempted to take into account the size of the two factors, as well as the rank order and pattern of loadings. However, this coefficient has been rarely used due to the difficulty of adequately approximating the difference between a variable's loadings on two factors in terms of the standard deviation of loadings on a given variable. Although Horn (1966) provided some data for ascertaining significance, the approximation of the loading standard deviations is crucial to the accuracy of the coefficient.

When an investigator uses any one of the above techniques, it is under the assumption that the resultant similarity coefficients are an accurate reflection of similarity between the factors. Given the known 'weaknesses' associated with some of the coefficients detailed above, this may not always be the case. It is with this point in mind that a recent paper by Bijnen, Van Der Net and Poortinga (1985) is of special importance. Using a matrix of 40 variables loading on eight factors, with six unique variables loading on the first four factors, and four unique items loading on the last four factors, they randomly permuted the variable loadings within each factor vector yielding 16 permuted factor structures. The initial structure as well as the 16 randomized structures were then rotated to approximate simple structure using the Promax algorithm (Hendrickson and White, 1964). The 16 factor patterns were then compared to the initial matrix rotated factor pattern using the KHB (Kaiser, Hunka, Bianchini) procedures. The results indicated that similarity indices ranged from 0.57 to 0.98 with an average factor comparison value of 0.81. Six of the 64 indices reached a value $\geqslant 0.95$. Using these results, the authors discuss their impact on the cross-cultural work of the Eysencks with regard to the similarity of the factors P, E, N and L (Psychoticism, Extraverion, Neuroticism and Social Desirability) measured by the EPQ. They conclude that the confidence the Eysencks have shown in their KHB coefficients as indicating factor invariance may be somewhat generous. If coefficients as high as 0.95 can be generated by pseudo-random permuted factor loading matrices, then what guarantee is there that at least some of the Eysenck coefficients are not due simply to chance? Bijnen *et al.* (1985) also highlight the results from a study reported by Nesselroade, Baltes and LaBouvie (1971) on the size and range of congruence coefficients calculated using another variant of congruential rotation applied to a series of random data matrices. These results indicates that for some matrices, the congruence coefficients could be as high as 0.96. Thus, it would appear that the KHB coefficients may not be adequate indicators of dimensional similarity. Certainly it seems that observing high values for such coefficients does not guarantee a high degree of similarity in all cases.

However, Bijnen *et al.* failed to calculate the means cosine between the variable pairs for each factor matrix comparison prior to calculating the KHB coefficients. If this value was low, then any comparison made between the factors would be quite artifactual. If the variable cosines are still low even when optimized by the KHB algorithm, then factor comparison is logically meaningless. The factor vector positions may be similar in each set of data, but the variable loading pattern would be quite different; this situation would yield some high KHB coefficients regardless.

Bijnen *et al.* also failed to completely specify the Nesselroade *et al.* results. The high coefficients observed tended to occur when the number of factors extracted from a matrix approached the number of variables in the matrix. In fact Tucker (1951) points out that the use of factor-matching procedures, in cases where the number of factors involved approaches the number of variables common to the two studies being matched, can lead to striking but quite nonsensical results. The EPQ matrices analysed by the Eysencks contained 90 variables and only four factors. Those analysed by Bijnen *et al.* contained 40 variables and eight factors.

Lastly, the Meredith (1964) algorithm used by Nesselroade *et al.* (1971) differs substantially from the Kaiser *et al.* procedure. Whereas the Kaiser algorithm simply rotates the 'experimental' variable space into maximum congruity with a targer factor space and calculates the factor vector cosines within that rigid framework, the Meredith procedure involves the mathematical determination of a matrix of best fit for the factor matrices of concern, then each factor matrix is rotated obliquely to maximize its agreement to the best-fit matrix.

Given the above and its potential impact on a major area within cross-cultural psychology, it was apparent that a more comprehensive analysis of the KHB coefficient, than that provided by Bijnen, was required. To this end the investigation below is an attempt to provide the basis for a more accurate interpretation of the KHB coefficients, in addition to confirming the validity or otherwise of the Eysencks' use of the KHB coefficients within their cross-cultural work. The primary focus of the investigation is to establish whether coefficients of the size consistently reported by the Eysencks could indeed occur by chance, and if so, to develop a systematic analysis framework for the use of the KHB coefficient that will allow the user to cross-validate any results using this particular method of factor comparison.

METHOD

Data

Four sets of data were used as part of the investigation. The first set of data involved the creation of 100 random data matrices using a pseudo-random number generator GO5CAF (yielding values between 0 and 1, uniformly distributed) provided in the NAG Mk. 10 subroutine library (NAG, 1983). The data were created using the same values as those used by the Eysencks for their unscored data (a 1 or 3). All matrices were of the order of 90 variables with 500 observations on each variable, they were of a completely random nature, with no bias introduced for targeting variable clusters or a predetermined number of factors. The random series were generated for each of the 100 matrices using a unique seed. The second data set was the English reference sample data for the 90-item EPQ held by the Eysencks. Both the male (1396 *S*s) and the female (1546 *S*s) data matrices were used. The third data set was a 44-variable subset of the 90-item scored EPQ, corresponding to those items measuring E and N. The data was extracted from the English reference sample data described above; the number of observations remained the same. The fourth data set was that of 71 KHB factor-comparison matrices that had been computed within the Eysencks' cross-cultural work and subsequently used in a demonstration of the clarity of the factor-comparison results (Eysenck, Barrett and Eysenck, 1985).

Statistics

Three coefficients of factor similarity were used in all analyses, the Pearson r, the Congruence r and the KHB r. The reason for choosing the Pearson and the Congruence r's was that these are the two most popular alternative methods of factor similarity generally employed in factor loading matrices. Since part of the aim of this investigation is to examine the behaviour of the KHB coefficient, these other two were included as comparative statistics against which the KHB results might be assessed. In addition, valuable information would also be obtained with regard to their future use as partial cross-validators of the KHB values. Both Pearson and Congruence r's were computed across oblique factor pattern matrices.

The computing formula used for the Pearson r was:

$$r_P = (N\Sigma a_{ij}a_{ik} - \Sigma a_{ij}\Sigma a_{ik})/\sqrt{\{[N\Sigma a_{ij}^2 - (\Sigma a_{ij})^2]*[N\Sigma a_{ik}^2 - (\Sigma a_{ik})^2]\}}, \quad (1)$$

where

a_{ij} = loading of the ith variable on the jth factor in the 'experimental' matrix,
a_{ik} = loading of the ith variable on the kth factor in the target matrix

and

N = the number of variables being correlated.

The use of the terms 'experimental' and target simply reflects the convention that one is comparing one matrix with another that has been previously defined, published or is of particular interest with regard to replication etc.; all three coefficients are symmetric with regard to order (experiment vs target, target vs experiment).

The computing formula used for the Congruence coefficient was:

$$r_C = \Sigma a_{ij}a_{ik}/\sqrt{[(\Sigma a_{ij}^2)*(\Sigma a_{ik}^2)]}. \quad (2)$$

The computing procedures and computing formula for the KHB coefficient were implemented on obliquely rotated factor pattern matrices as contrasted with reference vector and transformation matrices. The computing formula given below will use the notation provided by Kaiser *et al.* (1971) adopting the subscripts of 1 and 2 as indicating the target and 'experimental' matrices. Eight matrices are required initially:

$\mathbb{F}_1$ = the N variable $\times$ k factor orthogonal target loading matrix;
$\mathbb{H}_1^2 = \text{diag}(\mathbb{F}_1 \mathbb{F}_1')$, the diagonal matrix of reproduced communalities;
$\mathbb{T}_1$ = a transformation matrix that transforms the factors of $\mathbb{F}_1$ into an oblique pattern;
$\mathbb{L}_{11}$ = the factor intercorrelation matrix.

Likewise, the four matrices $\mathbb{F}_2$, $\mathbb{H}_1^2$, $\mathbb{T}_2$ and $\mathbb{L}_{22}$ for the 'experimental' matrix.

Given the use of Direct Oblimin rotation, which does not readily yield a transformation matrix in the manner say of Promax, $\mathbb{F}_1$, $\mathbb{F}_2$, $\mathbb{T}_1$ and $\mathbb{T}_2$ matrices had to be generated by the multiple group method (Guttman, 1952; Gorsuch, 1983; Harman, 1976). That is factor $\mathbb{L}_{11}$ into a $\mathbb{T}_1$ so that

$$\mathbb{L}_{11} = \mathbb{T}_1' \mathbb{T}_1 \tag{3}$$

and

$$\mathbb{F}_1 = \mathbb{A}_1 \mathbb{T}_1', \tag{4}$$

where $\mathbb{A}_1$ = the oblique factor pattern matrix, then compute $\mathbb{H}_1^{-1}\mathbb{F}_1$ and the corresponding $\mathbb{L}_{22}$, $\mathbb{F}_2$ and $\mathbb{H}_2^{-1}\mathbb{F}_2$, with

$$\mathbb{C} = (\mathbb{H}_1^{-1}\mathbb{F}_1)'(\mathbb{H}_2^{-1}\mathbb{F}_2) \tag{5}$$

and

$$\mathbb{G} = \mathbb{C}\mathbb{C}' = \mathbb{W}\mathbb{M}^2\mathbb{W}', \tag{6}$$

where $\mathbb{W}$ = matrix of unit-length column eigenvectors (loadings) of $\mathbb{G}$, and $\mathbb{M}^2$ is the diagonal matrix of positive eigenvalues of $\mathbb{G}$;

$$\mathbb{K}' = \mathbb{C}'\mathbb{W}\mathbb{M}^{-1}\mathbb{W}' \tag{7}$$

and

$$\mathbb{L}_{12} = \mathbb{T}_1'\mathbb{K}\mathbb{T}_2, \tag{8}$$

where $\mathbb{L}_{12}$ is the matrix containing the cosines of the angles between the two sets of factor vectors defined by matrices $\mathbb{F}$, $\mathbb{A}$ and $\mathbb{T}$. These angles may be interpreted as correlation coefficients much in the same way as the Pearson correlation coefficient.

Finally, equation (9) provides the mean solution cosine, that is, the mean cosine between all variable pairs for the optimal congruential rotation. It is this value which is indicative of variable similarity prior to the interpretation of factor similarity from matrix $\mathbb{L}_{12}$ above:

$$\mu = (1/N)*\text{trace}(\mathbb{H}_1^{-1}\mathbb{F}_1\mathbb{K}\mathbb{F}_2'\mathbb{H}_2^{-1}). \tag{9}$$

All three coefficients, the Pearson, Congruence and KHB, have a value range between ± 1.00, with 0.0 indicating no relationship at all (90° angular separation in terms of the KHB coefficient).

Procedure

The 100 random observation, 90-variable, data matrices were generated as indicated. These data matrices were then submitted to principal-components analyses. The first four components were extracted and rotated via hyperplane maximized Direct Oblimin rotation (Jennrich and Sampson, 1966; Barrett, 1985) with δ swept from -10.5 to $+0.5$ in steps of 0.5. The English reference sample EPQ data and the 44-variable E and N subset variable matrices were analysed in the same manner. The 71 KHB factor similarity matrices had been previously analysed by the Eysencks, with a description of the principal-components analysis and Promax rotation procedures given in Eysenck and Eysenck (1983). Three main analyses were undertaken using the various data matrices. In all analyses absolute value coefficients were used throughout; the size of the coefficients was the important variable under examination, hence using absolute values, this property would be preserved accordingly as when computing mean coefficients etc.

Analysis 1. From the possible 4950 unique random factor matrix similarity comparisons, 2000 were used here. This subset was not selected for any other reason other than computing time constraints; the matrix comparisons were selected over the full range of possible pairings. Given each comparison matrix has 16 cells, a total of 32,000 KHB, Pearson and Congruence coefficients were generated. However, for the purposes of this analysis, cell identity was preserved in order that factor position coefficient bias might be assessed.

For the KHB coefficient, μ the mean solution cosine (see above) was also computed as part of the comparison procedure. For each of the three coefficients, the mean, standard deviation, minimum and maximum value coefficient within each cell of the 4×4 comparison matrix was computed. In addition, the histogram of observed values within the range 0.0–1.0 in steps of 0.05 was also computed for each cell.

The aim of this analysis was to establish the behaviour of the three coefficients over a large range of random data matrices. If the coefficients behave as an investigator might prefer, all coefficients should tend to zero with none above about 0.5.

With regard to the unscored EPQ English reference data, the 100 random data factor pattern matrices were compared to the coresponding male and female EPQ matrices. This was a check against the possibility that using 'real' data, it might be possible to observe high coefficient values when comparing this data against totally random factor patterns. The same statistics as those computed above were utilized.

Analysis 2. This analysis was restricted to the 'ranging' properties of the coefficients within two sets of real data; the *scored* EPQ reference sample data and the 44-variable E and N subset data. The aim of this analysis was to examine the coefficient for a particular factor when salient E items/variables were systematically 'erased' from the pattern. To fully explain the procedure, I will take the case of the male EPQ 90-item reference data. Beginning with the factor pattern matrix, this was initially compared with itself yielding 1.0 down the main diagonal for each of the three comparison coefficient matrices. Then, returning back to the scored (0 or 1) data matrix, the first of the 21 E items had its responses replaced by either a 0 or 1, the sequence of which was pseudo-random (using the NAG algorithm detailed above). This 'degraded' data was then refactored using the procedures described above, and the resulting factor pattern matrix compared with the original 'master'. The degrading and comparison loop was computed until all 21 E items had been 'erased'. The use of the word 'erased' represents the effect that such an operation has on a variable that previously loads the E factor above 0.3. This value is reduced radically in all cases to below ± 0.2 and in most cases to below ± 0.1. Certainly, the variable is no longer salient in the accepted meaning of the term.

The reason for variable 'erasure' at the data matrix level was to preserve the overall measurement properties of the data as a whole and to allow any matrix disturbance to be reflected in both the PCA extraction and Direct Oblimin rotation phase. In addition, it provided an excellent opportunity to observe what might happen should a group of *S*s guess at random on some EPQ items. The 'ranging' of the resultant comparison coefficients could then be observed with regard to how sensitive each was with regard to salient variable degradation holding all other variable loadings more or less constant (the non-salient values would be expected to change slightly due to the disturbance within the hyperplanes caused by the randomized variables).

An investigator commonly interprets a factor by looking initially at the salient variable loadings, if half or more of the expected salients are 'missing' then he would obviously prefer to use a comparison coefficient that suitably indicated this. The results from Bijnen *et al.* (1985) suggest that the KHB coefficient might not be this sensitive.

Analysis 3. The final analysis used four sets of 4×4 factor-comparison matrices: the 2000 random data matrices, the EPQ male and the EPQ female reference factor pattern comparison with the 100 random data factor patterns and the 71 KHB cross-cultural factor-comparison matrices. Two histograms were computed detailing the percentage values for observing up to four KHB coefficients $\geqslant 0.90$ and $\geqslant 0.95$ within the four sets of matrices. The pertinent aim here is to establish the likelihood of observing up to four values $\geqslant 0.90$ (at least) within a 4×4 factor-comparison matrix where both or either of the comparison matrices reflect totally random variable relationships. The 71 KHB real data matrix comparisons provide the likelihood within the area of the Eysencks' cross-cultural work.

RESULTS

Analysis 1

For each cell in the 2000 random 4 × 4 comparison coefficient matrices the frequency of occurrence of values within the range 0–1 was computed (see the *Procedure* section). Table 1 presents the results from this analysis for the KHB, Pearson and Congruence coefficients. Given all cells in the 4 × 4 matrix possessed equivalent frequency distributions, the data presented is from just 1 of the cells and may be considered typical of the remaining 15. This equivalence demonstrates unequivocally that there is no particular positional bias occurring as part of the coefficient calculations, e.g. first factor comparisons will not tend to be higher than remaining second, third or fourth factor comparisons.

As can be seen from this table, almost 11% of the KHB coefficients have values ⩾0.80, with 4% occurring within the range 0.90–1.00. Out of the 16,000 possible Pearson coefficients, the maximum value was 0.4028, while for the Congruence coefficients, the maximum was 0.4016. A typical set of means, standard deviations and minimum and maximum values for the three coefficients is presented in Table 2.

Given the frequency distributions in Table 1, the values in Table 2 are not surprising. The results from comparing the EPQ reference data with each of the 100 random factor pattern matrices were extremely similar to the above results, therefore, separate data will not be presented here.

With regard to the properties of the random data matrices, all matrices were gramian, with positive eigenvalues. Typical eigenvalues for the first four unrotated components were 2.0178, 1.9672, 1.8396 and 1.8219. The mean solution cosine for the 2000 comparisons was 0.1742 with an SD of 0.0327 and maximum value of 0.3083. These values were similar to those observed in the male and female EPQ vs 100 random matrix comparisons except that the maximum values were slightly lower. Thus, all of the above comparisons are logically invalid and would not have taken place at all within a typical study using two sets of 'real' data.

Table 1. Typical frequency of observing Kaiser, Pearson and Congruence coefficients between 0 and 1 within 2000 random factor matrix comparisons

Coefficient range	Kaiser freq.	Pearson freq.	Congruence freq.
0.00–0.0499	117	725	762
0.05–0.0999	153	550	580
0.10–0.1499	118	404	357
0.15–0.1999	126	210	195
0.20–0.2499	136	73	79
0.25–0.2999	121	29	21
0.30–0.3499	123	7	6
0.35–0.3999	121	2	0
0.40–0.4499	118	0	0
0.45–0.4999	95	0	0
0.50–0.5499	115	0	0
0.55–0.5999	91	0	0
0.60–0.6499	97	0	0
0.65–0.6999	89	0	0
0.70–0.7499	84	0	0
0.75–0.7999	82	0	0
0.80–0.8499	79	0	0
0.85–0.8999	55	0	0
0.90–0.9499	45	0	0
0.95–1.000	35	0	0
	2000	2000	2000

Using absolute value coefficients.

Table 2. Typical mean, standard deviation and minimum and maximum absolute value Kaiser, Pearson and Congruence coefficients within 2000 random factor matrix comparisons

Coefficient	$\bar{X}$	SD	Minimum value	Maximum value
Kaiser	0.4210	0.2658	0.0007	0.9987
Pearson	0.0848	0.0626	0.0000	0.3575
Congruence	0.0850	0.0632	0.0001	0.3834

Using absolute value coefficients.

Finally, it was surmised that perhaps the higher KHB values would be observed where the mean solution cosine was also higher. To this end, the mean solution cosine was correlated with the maximum KHB value over the 2000 comparisons; this correlation was 0.0166. Similarly, this correlation was computed for the EPQ male and female data yielding the values −0.0065 and −0.0227, respectively. We may conclude that KHB size and mean solution cosine are not linearly related.

Analysis 2

The results computed from the degradation methodology (see the *Procedure* section) are given in Fig. 1–4. The graph notation employed is: FK = female EPQ KHB coefficient; FP = female EPQ Pearson coefficient; FC = female EPQ Congruence coefficient (the MK, MP and MC abbreviations similarly apply to the male EPQ data).

The correlation (r) values plotted are for the E vs E factor comparison only. The remaining three factors in the four-factor EPQ solution, and the remaining one in the two-factor EPQ solution all held their similarity values within the range 0.90–1.00. However, when 17 or so out of the 21 E items had been erased, the E factor tended to correlate as high with one or two of the other factors as with its counterpart in the reference 'target' matrix. In addition, the other factor/factor comparisons tended toward a more drastic reduction in size. This is not surprising as the Direct Oblimin rotation would be spreading variance across the larger factors to the reduced size 'E' factor.

Figures 1 and 3 provide plots of the mean solution cosine for each of the 21 comparisons against the KHB coefficient. These figures clearly indicate that the mean solution cosine is not correlated with size of the KHB coefficient in any consistent manner. In addition, for the two-factor 44-variable data, it would appear that over half the E items could be removed from the factor while still maintaining a mean solution cosine of around 0.70. For the four-factor 90-variable data presented in Fig. 3, up to 21 items could be removed while still maintaining a mean solution cosine above 0.70. This result indicates that the use of the mean solution cosine as an indicator of coefficient validity will be a function of the number of variables in the matrices being compared. The larger the number, the less effect that small numbers of variable pair dissimilarities will have on the KHB coefficients. Although the erased variable matrices are somewhat too perfect in terms of the method of variable erasure, they nevertheless do show the relative insensitivity of the KHB

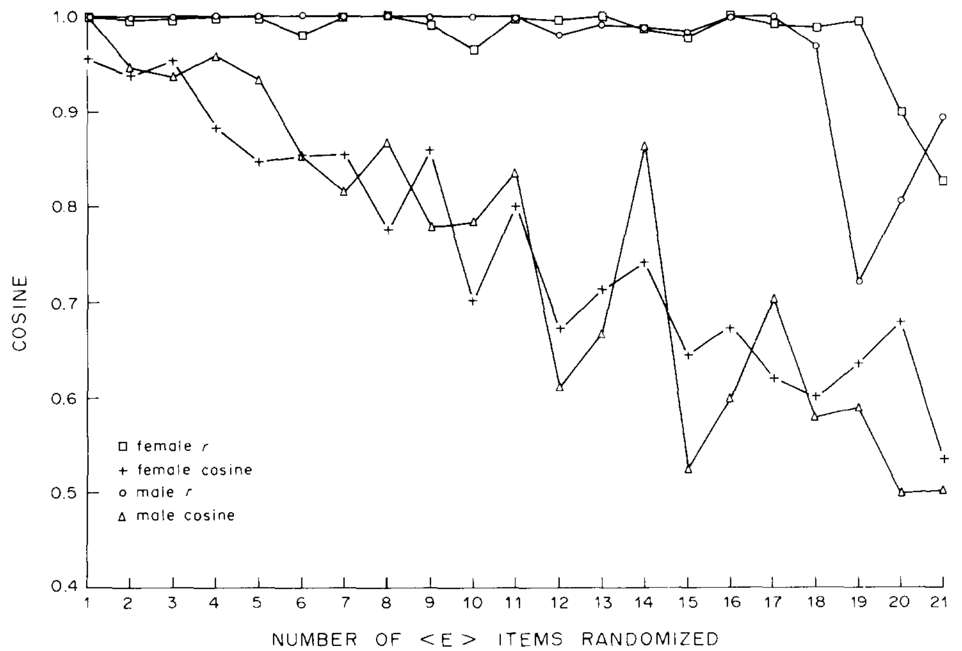

Fig. 1. Comparison between degraded scale Kaiser coefficient and mean solution cosine for the two-factor E and N item data.

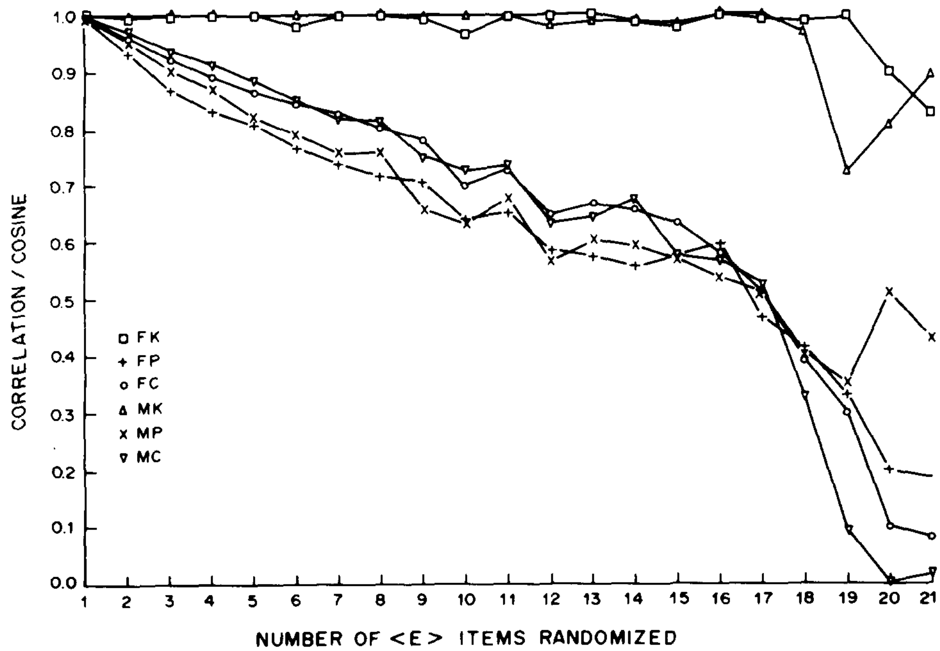

Fig. 2. Comparison between degraded scale Kaiser, Pearson and Congruence coefficients for the two-factor E and N item data.

coefficients to the loss of salient variables. Figures 2 and 4, however, indicate strikingly that the Pearson and Congruence coefficients are sensitive to salient variable loss.

Finally, as a matter of interest, the sensitivity of the Pearson coefficient to the proportionality of relationship between two sets of numbers was demonstrated inadvertently by the two-factor male EPQ reference targer solution (containing real data observed on all 44 variables). Comparing the matrix with itself yielded a Factor 1 vs Factor 2 Pearson r of 0.91. Although 21 E salient variables

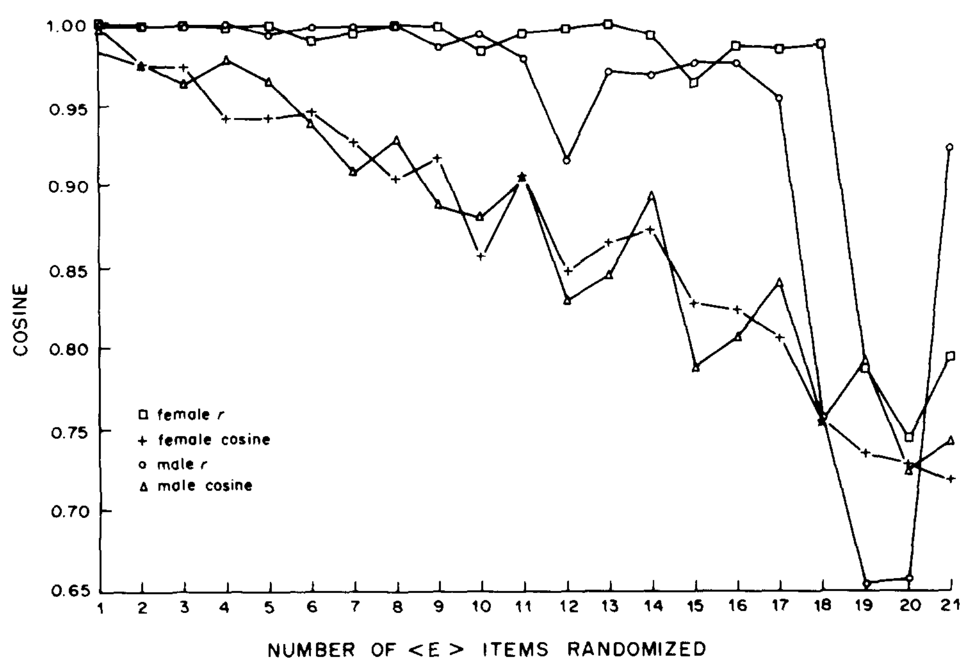

Fig. 3. Comparison between degraded scale Kaiser coefficient and mean solution cosine for the four-factor, 90-item EPQ data.

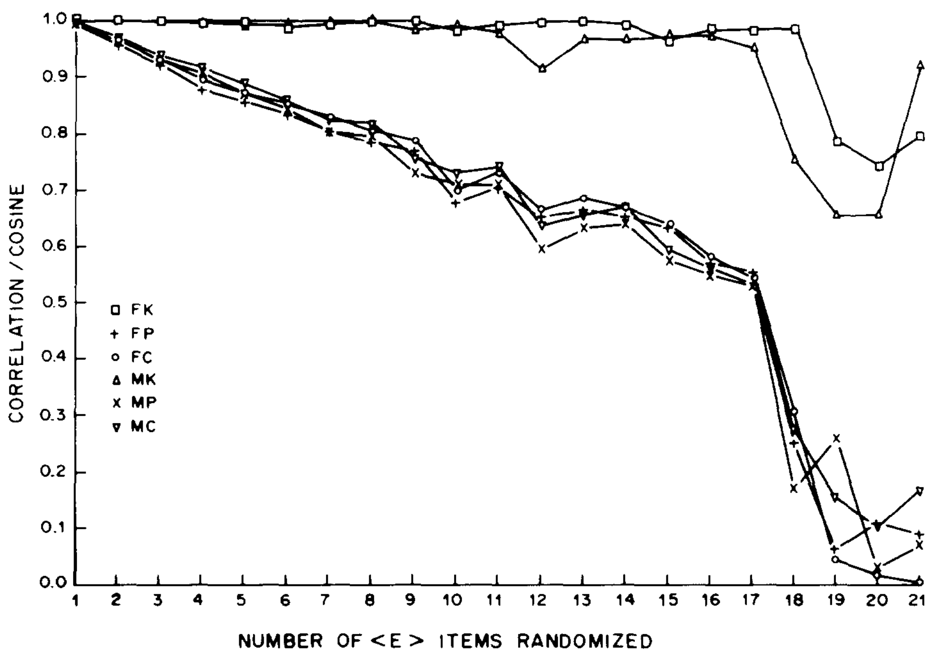

Fig. 4. Comparison between degraded scale Kaiser, Pearson and Congruence coefficients for the four-factor, 90-item EPQ data.

loaded the first factor ⩾0.30, and 23 N salient variables loaded the second factor likewise with perfect factorial simplicity, the proportionality between the two sets of loadings was such that the Pearson r was very high. Neither the KHB nor the congruence coefficients were affected by this proportionality (KHB ≈ 0.12, r_c ≈ 0.10).

Analysis 3

The histograms of percentage values for observing 0, 1, 2, 3 and 4 KHB coefficients in any one comparison matrix, within the four datasets described in the *Procedure* section, are presented in Fig. 5. The abbreviations in the figure describe the results from the 2000 random matrix comparisons (2000 RMs), the female EPQ vs 100 random matrices (F EPQ), the male EPQ vs 100 random matrices (M EPQ) and the 71 cross-cultural matrix comparisons (71 MATS). As can be seen from this figure, the likelihood of observing up to four values ⩾0.90 in any of the random matrix comparisons is practically nil. However, the Eysenck data demonstrates that almost 95% of their matrices have this property; this simply could not have occurred by chance.

DISCUSSION

From the above, it can be concluded that it is possible to obtain KHB values ⩾0.90 using totally random data matrices although the mean solution cosine for every comparison indicates that these results are invalid. In addition, it appears that the KHB coefficients are insensitive to the removal of salient variable loadings from a factor, unlike the Pearson or Congruence coefficients. These results beg the question of suitability of the KHB methodology for assessing factor similarity. Superficially it would appear that the Pearson and Congruence are far better indicators of factor similarity than the KHB coefficients. However, although the KHB methodology is insensitive to salient variable distribution on a factor, it is capable of correcting any 'apparent' factor pattern dissimilarities induced by the application of various rotational strategies. That is, if a matrix of unrotated factor loadings is compared to the same matrix but forced into an oblique factor pattern, the KHB procedure will 'undo' the rotation and show that the factors are in fact identical. However, because the Pearson and Congruence coefficients work directly on the factor patterns, they will show less than optimal agreement and in fact might indicate a lack of similarity (given

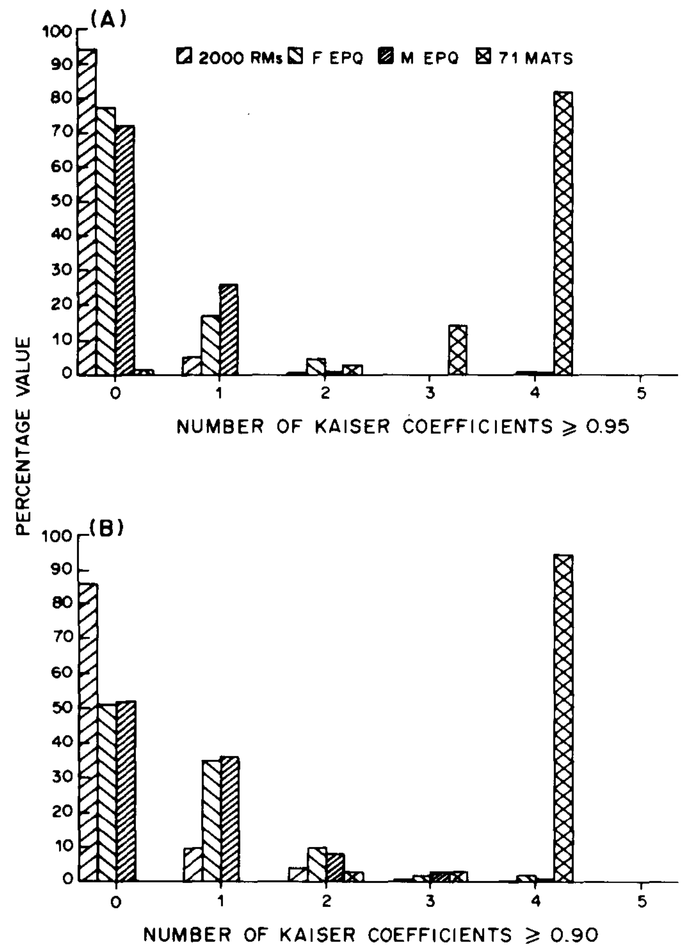

Fig. 5. (A) Percentage values for observing up to four Kaiser coefficients ≥ 0.95 within four types of matrix comparisons. (B) Percentage values for observing up to four Kaiser coefficients ≥ 0.90 within four types of matrix comparisons.

suitable distortion of the unrotated matrix). Thus the KHB methodology does have a powerful justification for its use. Note also that the Pearson coefficient is insensitive to the size of variables; relationship is expressed in terms of proportional similarity. In addition, the Congruence coefficient tends to overestimation when the sign of all variable pairs is predominantly positive or negative. It is apparent that all three coefficients have their respective drawbacks and advantages.

While the above may be relatively straightforward with regard to the theoretical and somewhat especial properties of the matrices used within this paper, an awkward question remains to be answered. What is the lower bound for a coefficient that may be said to determine factor similarity? For example, should the investigator accept Pearson or Congruence values ≥ 0.90 as accepting essential similarity or is that too low? Obviously, this is a question that has no simple mathematical answer except where a coefficient equals unity. Factor similarity is essentially a subjective concept unless defined within the framework of statistical inference as in the use of maximum likelihood factor analysis. However, even here some subjectivity has to be used in the setting of an 'appropriate' level of significance for the test statistic. For all practical purposes, the lower bound of coefficient size for factor similarity may be said to be defined by consensus opinion of investigators in the field. Generally, for the Pearson and Congruence coefficients, values ranging from 0.95 to 0.80 will be used as lower bounds. Although significance tests may be made on the Congruence coefficients (Korth and Tucker, 1975; Schneewind and Cattell, 1970), they are quite irrelevant to most factor pattern matrices; the values required for significance are normally below the 0.80 minimum conceptual lower bound. The significance test adopted is of the form 'is the coefficient significantly greater than zero?', *not* 'are the factors so similar as to be considered invariant?'. There is a great difference in these two questions, the latter being the one most asked by the investigator.

For the KHB coefficients, the concept of a minimum bound value (as may be used for the Pearson and Congruence coefficients) is quite simply inappropriate. It is possible to observe values $\geqslant 0.95$ when there are only 3 out of 21 salient variables left on a factor. Of course, it might be surmised that the KHB procedure has corrected for rotational 'error' within the matrices used in Analysis 2 above and is able to demonstrate higher similarities than the other pattern-dependent measures. The argument against this supposition is that the mean solution cosine decreases with each variable 'erasure' in addition to the loss of communality for each erased item (indicating that the loadings of that variable on all unrotated factors is so low that it could never assume salient status). Admittedly, although the angular separation between the low 'random loadings' and their salient counterparts in the target matrix could be very close, this would be extremely unlikely due to the randomness of the comparison matrix loadings. Within the four-factor 90-variable EPQ matrices, up to 21 factor-pure variables had their responses randomized; given the factor retains at least 75% of its low loading and hyperplane variables as defined within the target matrix, it is not surprising that its rotational position remains fairly stable as the salient items are erased. This stability is enough to ensure that the factor–factor correlation as assessed by the KHB metholodogy will remain very high. The key to 'validating' the size of the KHB coefficients must therefore lie in the use of the mean solution cosine as a bound/validity coefficient. In addition, it is possible to examine the main diagonal of equation (9) above, yielding the cosine separation for each variable pair; in this way, individual variable dissimilarity may be assessed. Thus for the KHB values, a two-stage validation procedure is probably required. Firstly, the mean solution cosine should be greater than about 0.90. The justification for this arbitrary value is that it does represent a fair degree of similarity between two-factor solution configurations. If the value is less than 0.90 then it is essential to examine the individual variable pair cosines to determine whether the fit is bad overall or confined primarily to a subset of variables. On the basis of this information, decisions can be taken as to the validity of some or all of the KHB coefficients. The degree to which a factor comparison may be considered as approaching invariance is then a matter for the subjectivity of the investigator; at least all the information required to assist in that judgement will now be available.

If we were to use these rules of thumb on the data above, using the values from all three coefficients to guide the decisions concerning levels of similarity, then all the random data matrix comparisons would be considered as either invalid or a complete failure as regards factor similarity between solutions. With regard to the 'ranging' procedures in Analysis 2, we would accept solutions with up to about six missing salient variables on the E factor (there are 21 items loading this factor >0.3). For some investigators this might appear reasonable, for others, too conservative. This data exemplifies the dilemma facing an investigator who wishes to make a meaningful statement concerning factor similarity using a single-valued index. In the present case, knowing the factor is relatively unidimensional [the greatest majority of the E items in the EPQ assess Sociability (Barrett and Kline, 1980, 1982)], it could be argued that the factor can be considered invariant even with up to 10 items missing; also there are no alternative item salients loading the factor >0.30. If alternative item salients were present then it would be best to assume that the factors being compared were not so similar as to be considered identical; at least, evidence to the contrary would be impossible to obtain within the psychometric framework of factor-comparison methodology alone.

With regard to the cross-cultural work of the Eysencks and their dependence on the KHB coefficient for the determination of factor invariance, the results from Analysis 3 demonstrate that their results could not be significantly affected by chance. In addition, the scale length of P, E, N and L within the 25 countries so far assessed (Barrett and Eysenck, 1984) differ by no more than six items on any one scale. Given scale construction has always been implemented by the selection of salient variable factor loadings and given the relative unidimensionality of the P, E, N and L factors (Eysenck, 1978; Barrett and Kline, 1980, 1982), it is highly improbable that the majority of the KHB coefficients could be >0.90 just by chance. It is possible that one or two of these country comparisons might considered to contain borderline matches, but this distinction is strictly subjective with many available arguments existing for choosing competing decisions. As Gorsuch (1983) and Cattell (1957, 1962, 1978) have indicated, the degree of certainty of match can and should be heightened through the use of indirect evidence such as the factor correlations within a study, the size (eigenvalues) of the factors to be matched, the αs and/or factor validities from any scales composed from the factor salients etc. Utilizing this information in conjunction with the KHB mean solution cosine and variable pair cosines, the Pearson and the Congruence coefficients jointly, the investigator is able to describe more completely and accurately the similarity between two or more factor patterns. While the Eysencks have not made use of the alternative coefficients, they have used the size of eigenvalues, salient loading pattern and scale αs to increase the confidence of the factor matches.

In conclusion, it is apparent that:

1. The probability of observing a spuriously high KHB coefficient is small but significant. The validity of any KHB coefficient may be enhanced by considering the mean solution cosine and the individual variable pair cosines. If the mean solution cosine is less than about 0.9, then the variable pair cosines should be examined to determine whether the bad fit is a generalized effect or confined to a particular subset of variables.
2. The use of the Pearson and Congruence coefficients is recommended, but only in conjunction with the KHB methodology. Unlike the latter, both coefficients are sensitive to any transformations made to one or both factor patterns. Although this may not be a problem if both matrices are rotated to a criterion such as simple structure or maximized factor variance (Comrey, 1967), it is better to maintain a conservative outlook and additionally compute the KHB coefficients.
3. The use of random data matrices or made up data, such as that used by Bijnen *et al.* (1985), is of no value in testing the validity of the KHB coefficients without the additional use of the mean solution cosine. If this cosine is less than about 0.9, then any similarity coefficients computed will be of dubious value unless further validated using the variable pair cosines. If the mean cosine is less than about 0.6, then no meaningful match can be made.
4. The pattern of KHB coefficients reported within the Eysencks' cross-cultural studies was shown to be impossible to generate using random vs random or real vs random data matrices. Given the other ancillary information used by the Eysencks such as the salient variable loading scale construction and the use of coefficient α and scale intercorrelations, the validity of the KHB coefficients is further increased.

Acknowledgements—The author would like to thank H. J. and S. B. G. Eysenck for the use of their EPQ data within the above data analyses.

The author is supported by a special grant from the Council for Tobacco Research.

REFERENCES

Ahmavaara Y. (1954) Transformation analysis of factorial data. *Annls Acad. scient. fenn., Ser. B* **88**(2), 54–59.

Barrett P. T. (1985) *The Psychometric Software Package—X Series Version 1.* Psychometric Research Agency, London.

Barrett P. T. and Eysenck S. B. G. (1984) The assessment of personality factors across 25 countries. *Person. individ. Diff.* **5**, 615–632.

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

Barrett P. T. and Kline P. (1982) The itemetric properties of the Eysenck Personality Questionnaire: a reply to Helmes. *Person. individ. Diff.* **3**, 73–80.

Bijnen E. J., Van Der Net T. Z. J. and Poortinga Y. H. (1985) On cross cultural comparative studies with the Eysenck Personality Questionnaire. *J. cross-cult. Psychol.*

Borg I. (1978) Procrustean analysis of matrices with different row order. *Psychometrika* **43**, 277–278.

Browne M. W. and Kristoff W. (1969) On the oblique rotation of a factor matrix to a specified pattern. *Psychometrika* **34**, 237–248.

Burt C. L. (1948) The factorial study of temperamental traits. *Br. J. Psychol.* (*Statist. Sect.*) **1**, 178–203.

Cattell R. B. (1949) r_p And other coefficients of pattern similarity. *Psychometrika* **14**, 279–298.

Cattell R. B. (1957) *Personality and Motivation Structure and Measurement.* World Books, New York.

Cattell R. B. (1962) The basis of recognition and interpretation of factors. *Educ. psychol. Measur.* **22**, 667–697.

Cattell R. B. (1969) Comparing factor trait and state scores across ages and cultures. *J. Geront.* **24**, 348–360.

Cattell R. B. (1978) *The Scientific Use of Factor Analysis in Behavioural and Life Sciences*. Plenum Press, New York.
Cattell R. B. (Ed.), Coulter M. A. and Tsujioka B. (1966) The taxonometric recognition of types and functional emergents. In *Handbook of Multivariate Experimental Psychology*. Rand-McNally, Chicago, Ill.
Cattell R. B., Balcar K. R., Horn J. L. and Nesselroade J. R. (1969) Factor matching procedures: an improvement of the S index with tables. *Educ. psychol. Measur.* **29,** 781–792.
Comrey A. L. (1967) Tandem criteria for analytic rotation in factor analysis. *Psychometrika* **32,** 143–154.
Eysenck H. J. (1978) Superfactors P, E and N in a comprehensive factor space. *Multivar. behav. Res.* **13,** 475–481.
Eysenck H. J. and Eysenck S. B. G. (1975) *Manual of the Eysenck Personality Questionnaire*. Hodder & Stoughton, London.
Eysenck H. J. and Eysenck S. B. G. (1983) Recent advances in the cross cultural study of personality. In *Advances in Personality Assessment* (Edited by Spielberger C. D. and Butcher J. N.). Erlbaum, Hillsdale, N.J.
Eysenck H. J., Barrett P. T. and Eysenck S. B. G. (1985) Indices of factor comparison for homologous and non-homologous personality scales in 24 different countries. *Person. individ. Diff.* **6,** 503–504.
Gorsuch R. L. (1983) *Factor Analysis*, 2nd edn. Erlbaum, Hillsdale, N.J.
Guttman L. (1952) Multiple group methods for common factor analysis: their basis, computation, and interpretation. *Psychometrika* **17,** 209–222.
Hakstian A. R. (1970) A computer program for generalised oblique procrustes factor transformation. *Educ. psychol. Measur.* **30,** 707–710.
Hamilton M. (1967) Notes and correspondence. *Br. J. math. statist. Psychol.* **20,** 107–110.
Harman H. H. (1976) *Modern Factor Analysis*, 3rd edn. Chicago Univ. Press, Chicago, Ill.
Harris M. L. and Harris C. W. (1971) A factor analytic interpretation strategy. *Educ. psychol. Measur.* **31,** 589–606.
Hendrickson A. E. and White P. O. (1964) PROMAX: a quick method for rotation to oblique simple structure. *Br. J. statist. Psychol.* **17,** 65–70.
Horn J. L. (1966) Integration of structural and developmental concepts in the theory of fluid and crystallised intelligence. In *Handbook of Multivariate Experimental Psychology* (Edited by Cattell R. B.). Rand-McNally, Chicago, Ill.
Horn J. L. and Knapp J. R. (1973) On the subjective character of the empirical base of the structure of intellect model. *Psychol. Bull.* **80,** 33–43.
Hundelby J. D., Pawlik K. and Cattell R. B. (1965) *Personality Factors in Objective Test Devices: a Critical Integration of a Quarter of a Century's Research*. Knapp, San Diego, Calif.
Hurley J. and Cattell R. B. (1962) The procrustes program: producing direct rotation to test a hypothesised factor structure. *Behavl Sci.* **7,** 258–262.
Jennrich R. I. and Sampson P. F. (1966) Rotation for simple loadings. *Psychometrika* **31,** 313–323.
Joreskog K. G. (1966) Testing a simple structure hypothesis in factor analysis. *Psychometrika* **31,** 165–178.
Joreskog K. G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34,** 183–202.
Joreskog K. G. (1971) Statistical analysis of sets of congeneric tests. *Psychometrika* **36,** 109–133.
Joreskog K. G. and Gruvaeus G. (1967) A computer program for restricted maximum likelihood factor analysis. *Res. Bull.* ETS, Princeton, N.J.
Joreskog K. G. and Sorbom D. (1982) *LISREL V Users Guide*. National Educational Research, Chicago, Ill.
Joreskog K. G., Gruvaeus G. and Van Thillo (1970) ACOUS: a general computer program for analysis of covariance structures. *Res. Bull.* ETS, Princeton, N.J.
Kaiser H. F., Hunka S. and Bianchini J. (1971) Relating factors between studies based upon different individuals. *Multivar. behav. Res.* **6,** 409–422.
Korth B. and Tucker L. (1975) The distribution of chance congruence coefficients from simulated data. *Psychometrika* **40,** 361–372.
Meredith W. (1964) Rotation to achieve factorial invariance. *Psychometrika* **29,** 187–206.
Mosier C. I. (1938) A note on Dwyer: the determination of the factor loadings of a given test. *Psychometrika* **3,** 297–299.
Mosier C. I. (1939) Determining a simple structure when loadings for certain tests are known. *Psychometrika* **4,** 149–162.
NAG Software Package Mk 10 (1983) Numerical Algorithms Group, Oxford.
Nesselroade J. R., Baltes P. B. and LaBouvie E. W. (1971) Evaluating factors invariance in oblique space: baseline data generated from random numbers. *Multivar. behav. Res.* **6,** 233–241.
Pinneau S. R. and Newhouse A. (1964) Measures of invariance and comparability in factor analysis for fixed variables. *Psychometrika* **29,** 271–281.
Schneewind K. and Cattell R. B. (1970) Zum Problem der faktoridentifikation: Verteilungen und vertranensintervalle von kongruentzkoeffizienten. *Psychol. beitr.* **12,** 214–226.
Schonemann P. H. (1981) Factorial definitions of intelligence: dubious legacy of dogma in data analysis. In *Multidimensional Data representations: When and Why* (Edited by Borg I.). Mathesis Press, Ann Arbor, Mich.
Tucker L. R. (1951) A method for synthesis of factor analysis studies. Personnel Research Section Report No. 984, Dept of the Army, Washington, D.C.
Wrigley C. S. and Neuhaus J. O. (1955) The matching of two sets of factors. *Am. Psychol.* **10,** 418–419.