

TESTING THE TEST: HOW ADEQUATE IS THE 16 PF WITH A N.Z. STUDENT SAMPLE ?

NGAIRE V. ADCOCK

Victoria University of Wellington

In view of doubts raised about the validity of the 16 PF test a simple procedure was proposed to test whether grave weaknesses existed when this battery was used with a New Zealand sample of university students. The results indicated that five factors (N, Q2, Q1, L and M) failed to achieve better than chance discrimination. It was concluded that, at least with this sample, many items were measuring related factors rather than those for which they were designed.

The 16 PF has a considerable international reputation, and has been claimed (cf. Cattell, Pichot and Rennes, 1961; Cattell, 1965) to be valid despite translation difficulties. The studies on which such claims have been made, however, seem to have been usually confined to comparisons of factor scores or second-order factors generated from them. Until the last few years few factorial studies or item analyses have been carried out, but these have raised some doubts about the applicability of the test. Eysenck's (1969) studies in England suggest that the factorial structure does not hold at all, Howarth and Browne (1971) find different factor patterns in Canada, and factor analyses in New Zealand (Adcock, Adcock and Walkey, 1971) have raised similar doubts.

Nevertheless, clinical and vocational guidance use has been claimed to be effective and research studies have suggested meaningful outcomes. For example a massive normative study carried out in Great Britain by the N.F.E.R. (Saville, 1972) shows sex differences in personality on a number of factors, with t-values running as high as 31.78.

In the light of this confusion it was felt that some simple assessment should be made of the factor scores of the 16 PF. The procedure adopted was that of sectioning the variables according to their scoring allocation as given by the 16 PF Manual so that the correlations between the items relating to each factor (intra-correlations) could be compared with those relating to items scored for other factors (inter-correlations). Clearly the value of a factor score depends upon its component variables being more highly saturated by that factor than are the variables assigned to another factor. By setting a cutting point at the 5 percent level of significance and counting the number of significant intra-correlations as compared to the number of significant inter-correlations one obtains a measure of the discrimination capacity in regard to the factor. Since on a chance basis there are more opportunities for significant inter-correlations to arise, an appropriate correction

has to be made before a ratio is calculated, but the ratio calculated after such correction should obviously favour the intra-correlations if the test shows any discriminative capacity. A ratio of unity would indicate no discrimination for the sample tested. A ratio less than unity would indicate that the factor score was positively misleading because the variance of its variables was more highly related to other factors.

It might be argued that, when submitted to such a criterion, a multi-factor test ought to have as a minimum requirement a greater absolute number of significant intra-correlations than inter-correlations since such a state of affairs would indicate a very high contamination of the factor scored by variances from other sources, presumably largely from other factors. This is just what a multi-factorial test should aim to avoid, and, although the use of suppressor variables may to some extent be cancelling out this contamination, the likelihood of their succeeding substantially is very doubtful.

The particular approach made here, however, would seem most appropriate where it is suspected that some factor-scores may have completely lost discriminative capacity with the population being studied and it was this state of affairs which seemed possible here. A factor analysis of the data (Adcock, Adcock and Walkey, 1971) had already been carried out and the results had indicated that only a few of the factors obtained corresponded substantially with the factors which the test was designed to measure, while some factors which did appear to correspond in terms of meaning obtained little variance from the items which were supposed to contribute.

BASIC DATA

The sample tested consisted of 164 (74 females and 90 males) First-year psychology students who took the test as part of their practical work for the course. The results were confidential and supplied only to the student concerned after computer scoring, and the significance of the scores was not discussed with students until after the test had been taken.

The factor analysis provided difficulties because the available computer programmes were unable to cope with more than eighty variables. To meet this difficulty the test items had to be divided into three sections. This was done in the first place by simply taking each third item so that all factors would have an opportunity to appear in each analysis, but the poor correspondence of the factors so obtained to those expected in terms of the Manual led to a new tactic being required. It was decided to section in terms of the factors indicated in the Manual. Factors were chosen for inclusion in a section on the basis of their similarity: a group related to Anxiety (O, C, Q3, Q4, L), a group related to Exvia (A, F, H, Q2, M), and the remaining five factors (omitting B). Mr F. H. Walkey wrote computer programmes for re-sorting the variables on this basis.

The analysis was carried out on an Elliott 503 computer using a standard programme written by Mr D. Kinneburgh. This provides for an initial principal components analysis and then proceeds to an iterative solution based upon the number of factors having latent roots above unity. The principal factors thus obtained are then rotated to varimax simple structure and this orthogonal rotation converted to the equivalent oblique solution by Promax (Hendrickson and White, 1966).

The correlation matrix provided by the second form of sectioning was used for the significance test here described. The correlation coefficients were checked by direct inspection and counting the number significant at the .05 level. The results are set out in Table 1, which

TABLE 1
Relative Discrimination of Factors

16 PF Factor	Discrimination Index (Intra-r) / (Inter-r)	Percentage of Intra-r's Significant at $\rho < .05$	Number of Significant Intra-r's Over Number of Significant Inter-r's
I	3.69	42	19/32
H	3.65	69	54/58
F	3.10	54	42/61
G	2.46	39	14/39
Q4	1.69	52	41/103
A	1.68	20	9/32
E	1.46	20	18/50
O	1.22	32	25/87
Q3	1.12	18	8/43
C	1.11	28	22/84
L	.90	13	6/40
Q1	.89	9	4/27
M	.87	9	7/34
Q2	.78	13	6/42
N	.62	9	4/39

shows also the percentage of intra-correlations which reach significance for each factor and the ratio of significant intra-correlations to the rate of occurrence in the rest of the matrix. It should be noted that some factors have ten and some thirteen variables while, due to an omission in the original data, G has only nine.

Two striking things are to be noted here: the very low number of significant correlations for some of the factors (six factors with less than 20 percent of their intra-correlations significant at the .05 level), and the fact that five factors fail to achieve even the minimum degree of discrimination. In no case is the number of intra-correlations greater than the number of inter-correlations. Detailed examination of these results in relation to the factor analysis carried out earlier (Adcock, Adcock and Walkey, 1971) follows.

FACTOR N. Naivete versus Shrewdness

The discrimination index as defined previously is a mere .62 and there are only four significant correlations at the .05 level in the factor cluster. In the factor analysis only one factor had more than two loadings above .30 on N-variables and this is quite obviously the G factor. The failure of factor N is doubtless considerably influenced by attenuation in a university student sample but it is quite clear that it is useless with such a sample.

FACTOR Q2. Group Dependency versus Self Sufficiency

With a discrimination index of .78 this factor also definitely fails and attenuation can hardly be offered as an explanation. In the factor analysis no factor picks up more than two loadings greater than .20 and no factor is clearly equivalent to Q2. The nearest approach is factor 3 in this section. It has a loading of -.485 on Variable V.22 and -.352 on V.122 but its highest loading is on V.3 from factor A (.518) and it has three other high loadings outside the Q2-variables. The central feature of this factor seems to be a need for the company of others, something in the nature of gregariousness. This may be an essential part of what Q2 purports to measure, but the fact that in this analysis only two of the defining variables are from Q2 while the remaining four came from A, F, H and M certainly discredits the factor score as given by the test and explains why the discrimination index should be so damning.

FACTOR Q1. Conservative versus Radical

This is a tricky trait to measure since radical goals are often very specifically motivated. These 16 PF variables obviously fail with our students. Attenuation may again significantly affect the results, but the fact that only 41 significant correlations turned up in the matrix and yet the variables got loadings greater than .30 on seven factors, suggests that the items badly fail in their aim. The corresponding factor (8) in the New Zealand study has three out of its five high loadings on these variables and is certainly closest to what they aim to measure. Its highest loadings are on "My decisions are governed more by my heart" (.56), "I have (not) the instructions for this test clearly in mind" (V.1) (.43), "Rather have job with higher salary which depended on constantly persuading people" (.40), "Money will buy anything" (.35). The trait looks almost like cynicism.

FACTOR L. Alaxia versus Protension

The discrimination index dictates failure as a practical measure but the factor analysis indicates that three of the variables (V.38, V.88 and V.13) give rise to a factor which has no high loadings outside and which must be regarded as indicative of the major trait influencing the student subjects in responding to these variables. The highest loading is

on V.113 ("If I am quite sure that a person is unjust or behaving selfishly, I show him up even if it takes some trouble"). V.63 and 64 seem to indicate a similar tendency to righteous anger. One has the suspicion that, not only is the factor as so measured poorly defined because of contamination, but it is also rather distorted.

FACTOR M. Praxernia versus Autia

Here again the discrimination index indicates a complete failure to discriminate from other factors. The factor analysis, however, picks up a factor which has appreciable loadings on three of its variables together with another from Factor A. This last (V.101, "Talking to customers" versus "keeping accounts") has a loading of .542, and had an obvious parallel in the next highest variable (V.40: "In a group task would rather try to improve arrangements than keep records"). V.116 and V.5 reflect a similar interest in working with people rather than alone. This suggests Guilford's "Thinking introversion" and is the central feature of Cattell's "Autia." One must conclude that the majority of the items for this factor are unsuitable for New Zealanders. It is difficult to understand, for example, how the item, "You can always notice on a man's face when he is dishonest" could discriminate this trait.

FACTOR H. Threctia versus Parmia

This is the most promising of the factors. The discrimination index (3.65) is slightly lower than that of factor I (3.69), but it is based on 54 significant correlations as compared to the latter's 19 so should have much higher reliability. In fact 69 percent of its intra-correlations are significant, indicative of quite respectable internal consistency. Nevertheless it is still to be noted that there are *more* significant correlations outside the factor cluster and an examination of the factor analysis throws light on this. Three factors accounted for the major variance of H: Parmia (4 variables), sociability (5 variables), inhibition (3 variables). The New Zealand parmia factor has only one substantial loading outside H and would appear to be quite a pure measure of the essential H factor. Sociability, however, which has the most H-variables loading on it, has even more loadings on F-variables (7) and this doubtless explains where most of the contamination comes from. It is worthy of observation that the three factors which thus shared the variance of the H-variables do so without overlap. None of these variables has a loading above .30 on more than one of the factors so that re-allocation of the items would place twelve of the thirteen with appropriate factors.

FACTOR F. Surgency

This has the next highest discrimination index with over half of its intra-correlations significant, but the factor analysis, as already mentioned, indicates that the factor which loads on the largest number

of its variables (seven) also loads on five of the H-variables. This factor seems quite clearly to be concerned with a need for social enjoyment as opposed to mere social contact (gregariousness) and quite distinct from *Parmia*. Five of the F-variables, however, define (with two Q-variables) another factor whose highest loading is on a liking for change and variety. This indicates a splitting of the surgency factor rather than confusion with another 16 PF factor. It looks as though surgency, as originally conceived, may manifest itself in both social enjoyment and need for change. If the five items from factor H were moved into this factor and its own items culled, its effectiveness should be very good.

FACTOR A. Schizothymia versus Affectothymia

Only 20 percent of the intra-correlations are significant and the discrimination index of 1.68 indicates quite a degree of confusion with other factors. Less than a quarter of the significant correlations fall within the factor cluster. In the analysis these variables loaded on six of the twelve factors, but only one factor has loadings on a minimum of three. This factor, which has five high loadings in all, appears at first sight to indicate self-sufficiency, but none of its loadings are on Q2-variables. Its highest loading is on "being a forester rather than a teacher" and another is on "living alone in the deep woods," but none of the sociability variables get appreciable loadings. "Forgetting trivial things" and "liking popular tunes" does not add much to the clues. Inspection of the A-variable matrix confirms the factor analysis. There is no consistent cluster beyond that of the three variables in this nuclear factor and one is left to conclude that the factor-score here is quite useless.

Looking at the 16 PF factors in this introversion/extraversion area it might be concluded that F is the only one which is sufficiently homogeneous as it stands, but its capacity to discriminate is badly reduced by many of the H-variables which are closely related to these. M and Q2 are useless as they stand, but there is evidence in both cases of such a factor existing although not measured by the factor scores. For A the present data offer no clear evidence and the measure provided by the factor score must be considered highly suspect for this sample.

The Anxiety Factors

The major primaries in this group are C, O, and Q4. Q3 was included with these because high control is often confused with low anxiety, and L because it has something in common with the pathological significance of Q4. Both of these latter get small loadings on the second-order Anxiety factor.

The poor performance of L has already been noted. Q3 and C are both marginal in discrimination (1.12 and 1.11), and O is only a little better at 1.22. Q4 has the highest discrimination index at 1.69 but, when one notes that this involves more than three times as many significant

inter-correlations than intra-correlations, it is obvious that the factor scores even here may be highly contaminated. The actual position as reflected by the factor analysis may be conveniently studied in Table 2, where all the variables with loadings .30 or higher on one of the sample factors best matching the relevant test factor are indicated by their variable reference. F.1 is included because it is the largest factor found.

TABLE 2
Comparison of 16 PF and N.Z. Factors

16 PF Factors	N.Z. Study Factors					
	F.1 Emotionality	F.5 O/Stolid	F.3 C/Docile	F.7 Q3/Ego- Control	F.12 Q4/Social Sensitivity	F.2 L/Aggressive
L	1 (139)	1 (114)	2 (64, 63)	—	—	3 (38, 88, 113)
O	1 (19)	2 (18, 168)	1 (93)	1 (68)	1 (43)	—
Q3	3 (147, 23, 123)	—	—	4 (48, 173, 172, 148)	—	—
Q4	2 (149, 174)	1 (100)	—	1 (124)	3 (74, 150, 50)	—
C	3 (5, 179, 129)	1 (154)	4 (104, 79, 129, 83)	—	—	—

Numbers of relevant 16 PF variables are shown in brackets.

It is interesting that the N.Z. F.1, which appears to be equivalent to a second-order factor of Emotionality, is directly measured by variables which in only one case overlap the variables defining the other factors. This suggests that the primary factors might be rather clearer if their items were more specific and Emotionality measured independently.

F.3 best represents the C factor, but three of its seven variables come from L and O and the docility which they seem to define would give rather a different meaning from that of Ego-strength. The original C-variables spread over three of the new factors and confirm the situation indicated by the discrimination index.

F.5 best represents factor O but not very obviously, and the trait of stolidity which seems to emerge here is hardly a very specific negative pole for the original "guilt proneness".

F.7 is rather more clearly related to Q3 since it gets two-thirds of its loadings from this area and all the variables concerned agree with the

concept of a personality under good control. The fact that none of these loadings overlaps the Emotionality factor suggests that here is a possibility of a lead for a factor which will disentangle the common effects of low emotionality and high control.

F.12 picks up an aspect of Q4 other than the general tendency to over-emotional response. This appears to be an undue sensitivity to social disapproval. The Q4 variables as they stand, however, do not constitute a homogenous group for the N.Z. student sample. Three get no appreciable loading on any of our twelve factors and the remaining ten spread their loadings over seven factors, including F.12.

The L-variables, as indicated earlier, obtain loadings on only F.2 and are insufficient to provide a useful factor. Moreover the factor indicated stresses an aggressive attitude rather than suspiciousness. In this connection the comment of a vocational selection expert to the effect that he associated factor L with the aggressive competitiveness which distinguishes the successful businessman is illuminating and suggestive.

FACTOR G. Super-ego Strength

This comes in fourth place according to the discrimination index. The factor analysis gives an even more favourable picture. Factor 1 has six of its eleven loadings on G-variables which have only five loadings on other factors and there is no major confusion with another factor in its group. The factor-score should therefore be quite meaningful and judicious culling of the items could lead to a good, reliable measure.

FACTOR I. Harria versus Premsia

This has the highest discrimination index and the factorial evidence is very favourable. Factor 3 has six out of its eight loadings on this factor so that the factor confirmation is excellent. There are also only two I-variable loadings on other factors so that confusion should be small. It is worth noting also that the four variables which do not reach the cutting point of .30 for loadings, nevertheless have low loadings from .158 to .197. With a more general sample this factor might give an even better showing.

FACTOR E. Dominance versus Submissiveness

The discrimination index is down to 1.46 and only 20 percent of the intra-correlations are significant. The factorial evidence is even more unfavourable, the eleven loadings being dispersed over eight factors. The factor which seems best to correspond has only two E-variable loadings with two from I and one from N. It would appear that no faith could be placed on the factor score for these subjects and, at least for New Zealanders, a new look needs to be taken at the problem of item writing for this factor.

CONCLUSION

In summary, the following conclusions may be noted: (1) With a New Zealand university sample only a few 16 PF factors are measured with satisfying reliability (H, I, F, G). (2) Five factors fail to achieve minimum discrimination by the index used (N, Q2, Q1, L and M), and, on the combined evidence presented, E and A have also to be rejected. In the case of N and Q1 this may be explained by the degree of selection in such a sample. (3) Factors in the anxiety group overlap very much in their variables and do not differentiate very adequately at the first-order level, but provide a stable second-order measure. (4) Despite the failure of so many specific factors when scored according to the test key, there is some evidence for majority of the 16 PF factors and reallocation of variables would apparently greatly strengthen some of the factors. Such changes, however, would have to be based on a larger and more representative sample of the general population.

REFERENCES

- Adcock, Ngaire V., Adcock, C. J., and Walkey, F. H. Basic dimensions of personality. Paper presented to I.A.A.P. Congress, Liège, 1971. *International Review of Applied Psychology*. (In press.)
- Cattell, R. B. A cross-cultural check on second-stratum personality factor structure—notably of anxiety and exvia. *Australian Journal of Psychology*, 1965, 17, 12-23.
- Cattell, R. B., Pichot, P., and Rennes, P. Constance interculturelle des facteurs de personnalité mesurés par le Test 16 PF. II: Comparaison Franco-Américaine. *Revue de Psychologie Appliquée*, 1961, 11, 165-196.
- Eysenck, H. J. and Eysenck, Sybil B. G. *Personality Structure and Measurement*. London: Routledge and Kegan Paul, 1969.
- Hendrickson, A. E. and White, P. O. A method for the rotation of higher-order factors. *British Journal of Mathematical and Statistical Psychology*, 1966, 19, 97-104.
- Howarth, E. and Browne, J. A. An item factor-analysis of the 16 PF. *Personality: An International Journal*, 1971, 2(2), 119-139.
- Saville, P. *The British Standardisation of the 16 PF*. Windsor, Berks: N.F.E.R. Publishing Co. Ltd., 1972.