The Comparability of the 16PF Form A and the 16PF5: some observations on the 16PF5 Test

Paul Barrett
University of Canterbury
Department of Psychology
Private Bag 4800
Christchurch
New Zealand

Laurence Paltiel
Psytech International Ltd.
Icknield House
Eastcheap, Letchworth
Hertfordshire
UK

Summary

The comparability of the 16PF form A and 16PF form 5 tests was examined at two levels of analysis, scale scores and factor structures. Using UK normative data, correlations between the scales of the two tests were seen to be less than 0.7 except for 3 out of the 15 personality scales. Correcting for unreliability of measurement, theoretical "best estimate" correlations of above 0.70 were seen in only 10 out of the 15 scales. The 2nd-order factor patterns between the two tests were also compared using normative manual data, in addition to other 16PF data and the normative dataset of the 15FQ personality test (an alternative to the 16PF). Only two factors were found to be highly comparable between tests. Anxiety and Extraversion. It was further noted that the 15FQ factor structure was more comparable to that of the 16PF form A than was the 16PF5. It was concluded that the 16PF form 5 is not comparable across all scales to the original 16PF form A. Further, attention was drawn to the fact that the increased alphas in the 16PF5 somewhat undermined the arguments made for the previous 16PF's lowered alphas, based upon Cattell's arguments concerning breadth of measurement.

The Direct Comparison between 16PF Form A and Form 5 Scales.

The 16PF5 (16PF Form 5) is being marketed as an "evolution" of the 16PFA (16PF Form A). Users of the 16PF5 are nevertheless informed (in the UK manual for the test, p.13, technical and norm addendum) ...***"users would be unwise to assume that scores (for some scales) from the 16PF form A and the 16PF5 are interchange-able".*** Interestingly, on page 3 of the US test manual, paragraph 1, it is stated that ...***"The 16PF Fifth Edition, although updated and revised, continues to measure the same 16 primary personality factor scales identified by Cattell over 45 years ago."*** It is apparent that some confusion exists between the US developers of the test and the ASE Ltd., the UK distributors and author of the UK manual!

One way to quantitatively assess the actual comparability is to examine the scores provided by individuals on both tests. A generally acceptable minimum bound for scale comparability computed using correlation coefficients is about 0.71 (the square root of a coefficient of this size can be interpreted as showing that 50% of the variation in responses in a 16PFA scale is accounted for by that in the corresponding 16PF5 scale, excluding the effect of the unreliability of measurement of both scales). Given this criterion, From Table D, also on page 13 in the UK manual, only **3** 16PF5 scales may be considered comparable to those in the 16PFA **(F and H and I).** It might reasonably be pointed out, however, that these raw correlations underestimate the real level of relationship between each scale pair. That is, the observed relationship may be adjusted for the amount of "random" measurement error

associated with each scale. Thus, If we correct each of the UK between-form correlations for unreliability of measurement, using the conventional formula:

$$\overline{R_{A5}} = \frac{R_{A5}}{\sqrt{R_{AA} R_{55}}}$$

where R_{A5} = the correlation between any form A scale and the corresponding 16PF5 scale

$\overline{R_{A5}}$ = the corrected correlation

R_{AA} = the estimate of reliability of measurement for the Form A scale (in this case the alpha coefficient reported for the 16PF form A standardisation sample (Saville and Blinkhorn))

R_{55} = the estimate of reliability of measurement for the 16PF5 scale (in this case the alpha coefficient reported for the 16PF5 UK standardisation sample in Table C, p.12 of the Technical Addendum)

we obtain the results as presented in Table 1 below. **10 out of the 15** scales so corrected may be considered acceptable in that at least 50% of the trait measured in a 16PF Form A scale is accounted for by the corresponding 16PF5 scale (four of the corrections exceed 1.0, indicating the fragility of this method of correction. However, for scales **F, H**, and **I**, the uncorrected correlation is already high).

Table 1: **Correcting the 16PF5 vs 16PF Form A scale correlations for unreliability of measurement in each scale.**

SCALE	ORIGINAL R	CORRECTED R
A	0.59	1.17
C	0.57	**0.92**
E	0.55	**0.86**
F	**0.80**	1.13
G	0.46	**0.76**
H	**0.85**	1.04
I	**0.71**	1.09
L	0.15	0.29
M	0.21	0.54
N	0.19	0.43
O	0.60	**0.91**
Q1	0.15	0.30
Q2	0.51	**0.90**
Q3	0.52	**0.87**
Q4	0.60	**0.85**

From this brief analysis above, it is clear that the warning message to users regarding problems with scale interchangeability is good advice. However, since one third of the test scales in the 16PF5 test are not comparable to those in the 16PFA, why is the test still being called the 16PF at all? This is highly misleading and confusing for users. The simple demonstration above indicates that many previous results obtained with the 16PFA will **NOT** be valid when the 16PF5 is used in place of the 16PFA. Primary scale profiles, second order scores, and other norms will not be comparable **except** where specific, unique, use is made of the 10 scales identified in Table 1 above.

The Evolution of a Revolution

As the marketing slogan would have it, *"the 16PF5 is an evolution of a revolution"*. The 16PF test, from its inception in the 50s, was indeed a revolution, as were the entire philosophy and psychometric viewpoints that accompanied the test. There can be no doubting the impact of Raymond Cattell on modern psychometrics. However, one feature of the 16PF always caused some misgivings among other psychometricians and informed test users, that was the fact that some scales had extremely low alpha coefficients. Cattell's views on this property of some of his scales were that low alphas were in fact a desirable feature of a scale, indicating a breadth of measurement that could not be achieved by a higher-alpha scale. The UK distributors of the test and various training companies all used this rather idiosyncratic statement of Cattell's as a major selling point for the test, isolating it from other tests on the market and positioning it as the elite amongst tests. The published quantitative

evidence indicating that the 16PF did not measure 16 primary/first order factors (and that the scales with low alphas were the very ones that could never be recovered via item analysis or factor analysis) was, however, totally ignored by both distributors, trainers, and users alike. Putting aside this situation, we are presented with a new 16PF version that now has reasonable to high alphas across all scales. **Why?** Do we conclude that the test is now of the same limited measurement breadth as those others that were being labelled in this way not so long ago?

The 16PF5 shares the same 2nd order factor structure as that of the 16PFA

If we accept that the primary (first order) scales may not be too comparable between tests, might we not reasonably ask whether the tests tend to converge at the second order level? From the USA 16PF5 test manual, p.76, the reported evidence is based upon a factor analysis of scale scores computed using 3498 individuals. However, above this matrix (on page 75 of the manual) is a matrix of scale intercorrelations from 2,500 individuals. In order to quantitatively compare the 16PF form A factors with the 16PF5 factors, we factored the scale intercorrelation matrix given in Table 12, extracting 5 factors as specified by the test manual (**not by the tests of factor extraction quantity**), and rotating them via hyperplane maximised direct oblimin rotation. We also used 5 other sets of 16PF form A data, and data from a sample of 84 UK volunteers who had completed the 16PF5. Finally, We also used a Psytech International 15FQ normative factor matrix as a comparison test for the 16PF and 16PFA. Factor comparisons were undertaken using the Kaiser-Hunka-Bianchini congruential fit procedure and Burt/Tucker congruence coefficients computed over the factor patterns. The congruential fit procedure yields a single correlation parameter that indicates how similar the two entire factor solutions are to one another, irrespective of any rotational procedure that has been previously applied to them. The conventional congruence coefficients reported for each factor are a measure of how similar the loadings on a specific factor are to one another, across any pair of factor solutions, **after** conventional rotation to simple structure. These individual factor coefficients indicate more precisely where divergences are occuring within the factor space. For both types of coefficient, a value of +1.0 indicates identity between the comparison datasets. A value of about 0.90 is considered an acceptable minimum in order to assert equivalence of measurement . This value indicates that over 80% of the variance in the first set of loadings (16PFA) can be explained by the second set (16PF5). However, it is a firm convention that male and female datasets on a single occasion should always correlate above about 0.95 for every individual factor in a well-designed test.

Factor comparison coefficients are required to be very high in order for a user to claim that the factors are essentially identical. As the coefficient size drops, so will the factor loadings on each pair of factors begin to diverge from one another. It is important to discriminate here between wishing to state that factors are **similar** to one another from stating that factors are **equivalent** to one another. As stated in the 16PF5 USA manual on page 3 ...*"The broad personality domains under which primary factors cluster are now called 'Global Factors' instead of 'Second-Order Factors'; however, these domains still exhibit an underlying factor structure **similar** to that*

found previously, reaffirming Cattell's original findings." **It is clear from this statement that the test developers do not consider the factors within the two tests equivalent to one another.**

The factor comparison results indicate that for each respective sample of 16PFA data, the male and female factor patterns are similar to one another, with the student volunteer sample demonstrating the weakest of the relationships. It is only the 15FQ that demonstrates unambiguous similarity. The UK vs US 16PF5 joint-sex datasets also show a fairly high relationship to one another. Bearing in mind the minimum acceptable value of 0.90 as being indicative of equivalence of factors, in only one set of data was this value exceeded in the 16PF5 and 16PFA comparisons, specifically for Anxiety (the UK sample of 16PF5 data). Overall, only the 16PF5 factors of Anxiety and Extraversion could be said to be consistently similar across the various 16PFA datasets. The remaining factors of Independence, Control, and Tough-Poise are **not** considered equivalent between versions.

Summarising all these results, we can conclude the following:

☹ ...The 16PF5 scales and factor solution are only partially equivalent to those from the 16PF form A, using both UK or US data. This is partly due to factorial instability of the 16PFA between sexes.

☺ ...Overall, the 15FQ factors seem to be more comparable to the 16PFA factors than do the 16PF5 factors. This is unusual since the 16PF5 is meant to be the "natural" successor to the 16PFA.

In conclusion:

From our reading of the US and UK manuals, we would conclude that the 16PF5 test is a psychometrically better test than the 16PFA. However, it is sufficiently different from the 16PFA at both the primary and second order level that it **cannot** be used as a straightforward replacement for the 16PFA. Thus, the body of occupational evidence that supported the old 16PFA test cannot easily be used to support this new test. Selection of this test against competitor products will have to be made by noting arguments other than those based upon the "pedigree" of the 16PF or the name of "Cattell". Rather, the information contained in the test manuals is now all that supports the use of this test. While not inconsiderable, it is not up to the standard or breadth of the previous version.

It is also considered unusual that given the resources available, the test developers did not seriously attempt to equate the two products (or those portions of the old 16PF that were crucial with regard to occupational validity). This is either a tacit admission that the 16PF form A was too full of measurement problems to be sufficiently equated, or that they simply overlooked the body of test users who would face major problems in introducing a new test into their selection procedures. It is the opinion of the authors, based upon the analyses reported above (and other previously published

analyses of the 16PFA) that the global changes to the 16PFA indicate that many of the original measurement constructs themselves were badly flawed, and needed major psychometric attention. It is perhaps interesting to note that no-one to date has been able to replicate the factor structure of the original 16 primary factor traits.

Finally, for those readers of this article who question these analyses as mere parlour game psychometrics, we would ask you to consider the purpose and meaning of a psychometric test. The names and descriptions assigned to scales are given validity by recourse to external, real-world behaviour predictions and relationships. However, a requirement of achieving any such validity is that scales of items are homogenous and unidimensional to some degree. Where this is not the case, criterion validity will be compromised by the instability of the measures used (rather like using a ruler that changes it's shape each time you use it to measure length). Secondly, unless the test items are to be used as no more than semi-structured questions for an interview, then the user will be making assumptions that the test measures what it claims to measure, that the profiles can be used quantitatively, and that the measures themselves are psychometrically valid and reliable to some quantifiable degree. The only means of assessing these latter psychometric properties is via quantitative indices. Relegating these to the box marked "trivia" also relegates psychometric testing to the same box. The point here is that a test publisher cannot introduce a psychometric test into the market, call it the same name as a previous version of the test, and yet fail to **quantitatively** demonstrate that little of the new test actually appears to measure the same concepts as the previous version. Wishful thinking is no answer to psychometric evidence.