The Fourth Test User Conference

Book of Proceedings

9 - 11 June 1999

as presented by

Division of
Occupational Psychology
of the British Psychological Society

in conjunction with
Selection & Development Review

DIFFERENCES BETWEEN APPLICANT AND NON-APPLICANT PERSONALITY QUESTIONNAIRE DATA: SOME IMPLICATIONS FOR THE CREATION AND USE OF NORM TABLES.

Authors: [1] **Rosalie Brown and** [2]**Dr Paul Barrett**

Affiliations: [1]Psychometric Technology Limited and Psytech International Ltd
[2]The State Hospital, Carstairs and University of Liverpool

Abstract

There is a curious reticence amongst test publishers concerning the possible distortion of item responses within samples of individuals who stand to gain or lose some reward, depending upon their responses to a questionnaire. If the distortion is hypothesised as resulting in changed scores on a set of questionnaire scales, then it is understandable that this effect is not of critical importance, as the user can interpret the normative scores from applicant groups accordingly. However, what if the distortion causes substantial disturbances in the psychometric structure of the questionnaire? That is, what if it can be shown that such response distortion causes the factor scale composition to be degraded to such an extent that it can be concluded that we are no longer making measurement of the same factors within both an applicant, and non-applicant sample? If this is the case, then we have a problem, for how can we produce norm tables of scores for now "non-existent" factors? These questions are the focus of this paper. By reviewing available empirical evidence in the literature that addresses these questions, and by some psychometric analysis of scale-level data drawn from the US and UK, we were able to demonstrate that there is little structural distortion of the second order factor structure in the 16PF-5 questionnaire. However, two scales did not load as expected, Q1 (openness to change) on the Tough-Minded factor, and L (Vigilance) on the Anxiety second order. This overall result is in disagreement with the results reported by Schmit and Ryan (1993) using the NEO questionnaire, where more severe distortion of the expected factor structure was found within an applicant sample. A by-product of our analyses is the clear evidence that structural equation modelling is a more optimal methodology for achieving these forms of constrained procrustes solutions for target and sample datasets than exploratory congruential methods, regardless of model fit issues.

Introduction

People regularly ask on personality test training courses and prior to or after completing personality questionnaires if these measures can be 'faked'. This question has several interesting facets:

Raising or enhancing of scores:
Whether people can consciously raise their scores to enhance their image on personality measures has been extensively studied and the conclusion to this question appears to be a resounding yes. A study using the Edward's (1959) Personal Preference Schedule (EPPS), which was designed to minimise social desirability, demonstrated individuals' ability to portray a more favourable self-image in line with a specific purpose (Borislow, 1958; Dicken, 1959; Stollack, 1965; and Orpen, 1971). Dicken (1960), Canter (1963) and Johnson (1986) using the California Psychological Inventory (CPI), and Hough et al (1985) with a customised personality inventory designed for the army, all demonstrated that people could consciously enhance self-image.

Conscious dissembling in the selection context
There is also further considerable research evidence to support the belief that respondents can consciously dissemble on personality inventories. A study by Wesman (1952) tested students twice on a personality inventory with a one-week interval. On one occasion they were instructed to pretend that they were applying for a sales role in a large industrial organization and on the second testing they were instructed to pretend that they were applying to be a librarian. The self-confidence scores for the simulated sales role were significantly higher than for the corresponding librarian scores. A further study by Orpen (1971) used the EPPS to test 30 male and female applicants for clerical positions within a large insurance company. He tested the same group several weeks later, as part of a research project, and acquired data on a matched sample of under-graduates who completed the questionnaire under normal and fake good conditions. He found that the under-graduates scores changed significantly over the two conditions but that the applicants' scores changed only slightly.

Hough et al (1985) conducted a more rigorous study, which was a repeated-measure, counterbalanced design where faking good was trying to get into the army and faking bad was attempting to fail selection. 245 soldiers were randomly allocated across four groups: (1) Fake good-Honest; (2) Honest-Fake good; (3) Fake bad-Honest; (4) Honest-Fake bad. 'Honest' was an instruction to respond honestly. Scores were compared to 276 army applicants who had not been informed of the selection decision. Hough et al found that (a) when instructed, soldiers could distort their responses; (b) the distorted protocols could be detected with a validity key; but (c) regular army applicants did not distort their responses. Hough et al (1985) conclude that, "intentional distortion does not appear to be a problem in an applicant setting".

Self-deception and impression management

According to Paulhus (1984, 1986) faking can occur when a respondent answers in a positively biased way and actually through self-deception believes their responses are true. Paulhus draws a distinction between this and impression management, which is, defined as conscious dissimulation in answering questionnaire items. Several studies have demonstrated that self-deception and impression management are two quite distinct factors of social desirability (Jackson & Messick 1962, Paulhus 1984, Wiggins 1964). A recent study by Barrick & Mount (1996) examined whether self-deception and impression management affects the predictive validity of two of the "Big Five" personality dimensions, conscientiousness and emotional stability, in two applicant samples. Results from structural equation modelling indicated that applicants did distort their scores on both dimensions through both self-deception and impression management. However, neither type of distortion attenuated the predictive validity of either personality construct. It should be noted that this is one of the few studies found from the literature search that used actual job-applicants and not newly hired employees or experimental groups, whereby participants are asked to respond positively.

The Key Argument within this paper

When a personality test is used within an "applicant-oriented" setting, an implicit assumption is made that the questionnaire scales retain the same measurement properties and psychometric relations as initially generated using volunteer datasets. So, in all the studies above (apart from Barrick and Mount), the aim has been to immediately detect score profile differences across the scales rather than first examine the extent to which the measurement properties of the scales within the questionnaire are being distorted. But what if the factors found from within the volunteer datasets are no longer identifiable in an applicant sample? This was a question initially examined by Schmit and Ryan (1993) in their examination of the NEO Five-Factor Inventory structure in applicant and non-applicant populations. Using structural equation modelling, the expected 5-factor structure did fit a student volunteer data sample, but did not fit the applicant data. What this means is that to continue scoring questionnaire data using a score-key generated from volunteer data might no longer be considered a valid activity; a state of affairs which could have implications for the legal defensibility of such a procedure. If Schmit and Ryan's results are replicated and are seen to hold across other questionnaires, then the test user would be placed in an invidious position of having to use a score key for measures which cannot empirically be shown to exist within the sample for which scoring is to be undertaken.

The Question to be Answered

What degree of factor similarity exists between applicant and non-applicant data? That is, if the differences between samples are only ones of "score level", and not of psychometric structure, then we should be seeing virtual identity between factor solutions within each sample.

Respondent Datasets

Our own datasets comprised a sample of 589 non-applicant (N=403 training course delegates), mixed gender individuals who completed the 16PF-5. Our applicant sample consisted of 506 mixed gender participants, the majority of whom were graduate applicants to a merchant bank. We were also given access to the ASE UK-standardisation (N=1575) 16PF-5 scale correlation matrix.

Statistical Analyses

The analyses of these datasets used exploratory MINRES common factor analysis of each correlation matrix, with forced extraction of 5 factors, rotation using direct oblimin, and factor comparison analysis using the modified congruential Kaiser-Hunka-Bianchini methodology of Barrett, Petrides, and Eysenck (1998). Full details of all the procedures are given in Brown (1998). We were unable to use structural equation modelling (SEM) for a conventional group comparison strategy as the 16PF-5 second order model fails to fit either the UK or US normative datasets (see Brown, 1998). This is not surprising as SEM fails to fit almost any existing personality questionnaire, including the NEO 5 factor model (McCRae, Zonderman, Costa, Bond, and Paunonen (1996).

In answer to question 1 above, Table 1 below shows the factor congruences (measures of similarity akin to a correlation coefficient that varies between –1.0 through to +1.0) between the *ASE non-applicant reference sample dataset and the non-applicant sample data collected by us* (N=589). The purpose of this initial analysis was to show that our data was in no way different in structure from the expected UK factor structure.

Table 1: Tucker Congruence Factor Similarity Coefficients – Kaiser Orthogonal Factor Matrices, UK ASE standardisation matrix (British Adults N=1575) Varimax target compared with our sample of UK Non-Applicants (N=589)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	**.9905**	-.0622	-.1151	.4009	-.3575
Factor 2	-.0617	**.9388**	-.1318	.1977	.3587
Factor 3	-.1100	-.1269	**.9821**	-.1218	.1343
Factor 4	.3961	.1968	-.1260	**.9731**	-.0619
Factor 5	-.3613	.3653	.1420	-.0633	**.9623**

The Mean Solution Congruence for this comparison was 0.97. This measure is an indicator of global similarity between the two solutions. What we see here is that our sample data possess virtually the same structure as the ASE reference sample data. The use of a Varimax target solution is required in order to ensure identifiability of the target solution, prior to comparison matrix rotation. Thus, in order to ensure that the Varimax target was a valid rotation solution for the data, the factor patterns from a Varimax and Direct Oblimin rotation solution for the ASE matrix data were compared. No factor congruence dropped below 0.98 – indicating that the overall 5 factor solution was essentially orthogonal.

Table 2 below shows the factor comparison results between *ASE non-applicant reference sample dataset and the applicant sample data collected by us (N=506).* Here we see that for the comparison between factor 5, we are losing factor comparability. The mean solution congruence of 0.93 reflects this single factor disparity.

Table 2: Tucker Congruence Factor Similarity Coefficients – Kaiser Orthogonal Factor Matrices, UK ASE standardisation matrix (British Adults N=1575) Varimax target compared to UK Applicants (N=506).

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	**.9493**	.0025	-.1352	.4397	-.4656
Factor 2	.0027	**.9527**	-.0458	.2025	.3497
Factor 3	-.1438	-.0453	**.9618**	-.2052	.1498
Factor 4	.4075	.1745	-.1788	**.9471**	.0740
Factor 5	-.4942	.3451	.1495	.0847	**.8075**

In order to see just what this disruption to loadings on factor 5 looks like, Table 3 below provides the comparison matrices/loadings respectively for the UK ASE non-applicant reference sample, our non-applicant, and applicant datasets respectively. Note that the non-applicant UK ASE reference sample dataset factor solution is the "target", against which the non-applicant and applicant factor matrices are rotated toward maximum congruence. All these solutions are orthogonal factor matrices.

In Table 3, the factor loadings for the US sample of N=3498 (presented in Table 1.3, p. 14 of the US Technical manual) are listed in italics next to the UK MINRES solution loadings. *The reason for including these data (which are from an intermediate 14-item per scale development version of the 16PF-5) is to show that the UK and US solutions are apparently less comparable than the applicant and non-applicant data.* We stress the word "apparently" because we have to consider that our solutions are orthogonal whereas the US matrix was rotated obliquely. Further, the US matrix data is based upon the development 224 item questionnaire rather than the final 185 item questionnaire that comprises the currently retailing 16PF-5. So, in order to make *maximally optimal comparisons* between the 16PF-5 second order model solution, the UK ASE data, and our two data samples of UK applicants and non-applicants, we returned to a structural equation modelling approach. This is not because we are interested in model fit (which it doesn't), but rather for the use of the methodology as means of estimating factor loadings within a series of loading and oblique factor correlation constraints. Using SEPATH, the loadings above |0.3| in table 1.4 of the US manual were specified as paths to be estimated. All other loadings were defaulted to 0.0. The factor correlations from the US Technical manual Appendix 1B were specified as a series of constraints between the 5 second order factors. Completely standardized solutions were computed from the various scale intercorrelation matrices. Table 4 provides the factor loadings for each of the 5 factors, for the US data in Table 1.4, and four SEPATH solutions. Out of interest, we analysed the US normative sample scale correlation matrix (Appendix 5B, p. 94, N=2500) for the final 16PF-5 scales. This was done in order to ensure that the results from Table 1.4 were not in some way affected by the inclusion of extra items in each of the 16PF-5 scales.

The result of these analyses indicates that there are just two factors which seem to be substantially affected by the test-taking mode, Anxiety and Tough Minded. These data in Table 4 clarify to some extent those in Table 3, by constraining the solutions to a fixed target which consists only of salient loadings, whilst constraining the factor intercorrelations to be exactly the same as those in the US normative data of N=3498). The lesson learned from the above analyses with regard to future work in this area, is that exploratory procrustes analysis using congruential or simple factor pattern congruence matching is not as efficient or powerful as structural equation modelling. This is especially the case when interest is focused on the technique as a form of orthogonal or oblique procrustes solution.

Conclusions

So, where does this leave us with regard to our question ... *What degree of factor similarity exists between applicant and non-applicant data*? From all the results in Tables 3 and 4, we would conclude that a high degree of similarity exists overall in the 16PF-5, given our particular dataset. However, the results for Anxiety and Tough Minded in Table 4 are nevertheless awkward. The scale Q1 shares the largest 2nd order factor weight with scale I on Tough Minded (0.5) in the factor scale equations given on page 16 of the US Technical manual (p.16), yet is barely identifiable as a significant variable in our applicant data. Likewise scale L on the Anxiety second order, although this is not so severely affected. On the basis of these results, and those of Schmit and Ryan, there does seem to be some evidence that applicant factor structures of questionnaires are different in certain respects to those of volunteers. Although for the 16PF-5, our data suggest that the difference is marginal, *it is still the case that two scales used in the second order factor equations would not have been identified under the rules for identification used in the US Technical manual (loadings > 0.30), and therefore not weighted at all in the equations*. Further, one wonders to what extent the meaning attributed to these second orders might change? With NEO data, this problem was substantially worse. Whether or not the effects extend to other questionnaires is an empirical question, and one that should be required to be answered by every publisher of commercial questionnaires. As we have shown above, this is not difficult, but it does require that a questionnaire possesses an *a priori* known and reasonably clear psychometric structure, as in the case of the 16PF-5 and NEO.

References:

Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: an examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences*. **25**, **5**, 805-819

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, **81**,**3**,261-272.

Borislow, B. (1958). The Edwards Personal Preference Schedule and fakability. *Journal of Applied Psychology*, **42**, 22-27.

Brown, R. (1998) A study investigating the differences between applicants and non-applicants on the sixteen personality factor questionnaire – Fifth edition. *Thesis submitted in partial fulfilment of the MSc degree in Occupational Psychology, University of Sheffield*

Canter, F. M. (1963). Simulation on the California Psychological Inventory and the adjustment of the simulator. *Journal of Consulting Psychology*, **27**, 253-256

Dicken, C. F. (1959). Simulated patterns on the Edwards Personal Preference Schedule. *Journal of Applied Psychology*, **43**, 372-377.

Dicken, C. F. (1960). Simulated patterns on the California Psychological Inventory. *Journal of Counseling Psychology*, **7**, 24-31.

Edwards, A. L. (1959). *Edwards Personal Preference Schedule*. New York: Psychological

Corporation

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., and McCloy, R. A. (1990) Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 5, 581-595

Jackson, D. N., & Messick, S. (1962). Response styles on the MMPI. *Journal of Abnormal and Social Psychology*, **65,** 285-299

Johnson, J. A. (1986). *Can applicants dissimulate on personality tests?* Paper presented at the annual meeting of the American Psychological Association, Washington, DC

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology,* **70, 3,** 552-566

Orpen, C. (1971). The fakability of the Edwards Personal Preference Schedule in personnel selection. *Personnel Psychology,* **24,** 1-4.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology,* **46,** 598-609.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & Js. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement (pp. 143-165).* New York: Springer-Verlag.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in Personnel Selection: Factor Structure in Applicant and Nonapplicant Populations. *Journal of Applied Psychology,* **Vol. 78, No. 6,** 966-974.

Stollack, G. E. (1965). EPPS performance under social desirability instructions. *Psychological Reports,* **16,** 119-122

Wiggins, J. S. (1964). Convergences among stylistic measures from objective personality tests. *Education and Psychological Measurement,* **24,** 551-562

Wesman, A. G. (1952). Faking personality test scores in a simulated employment situation. *Journal of Applied Psychology,* **36,** 112-113

Table 3: Factor Pattern Loadings from exploratory MINRES common factor analyses and Orthogonal Procustes rotations

UK: UK 16PF-5 reference sample, MINRES factors, Varimax rotated, 5 factor solution

Non-App: the non-applicant sample (N=589) MINRES factors, maximally congruent to the UK reference sample

App: the applicant sample (N=506) MINRES factors, maximally congruent to the UK reference sample.

- Figures in brackets are the *US matrix loadings reported in Table 1.3, p.14, of the US Technical Manual*
- Loadings less than |0.3| are left blank.

	Fac1			Fac2			Fac3			Fac4			Fac5		
Scale	**UK/*US***	**Non App**	**App**	**UK/*US***	**Non App**	**App**	**UK/*US***	**Non App**	**App**	**UK/*US***	**Non App**	**App**	**UK/*US***	**Non App**	**App**
A	**.62** ***(.74)***	.57	.69				**-.33** ***(-.35)***		-.33						
C			.30				**.52**	.46	.51	**-.49** ***(-.70)***	-.52				
E							**.50**	.44	.40				**.49** ***(.87)***	.35	.54
F	**.66** ***(.70)***	.71	..47	***(-.39)***											
G				**.64** ***(.78)***	.65	.64									
H	**.58** ***(.44)***	.59	.46				**.32**	.30					**.36** ***(.43)***	.33	
I							**-.61** ***(-.75)***	-.43	-.39						
L										**.57** ***(.57)***	.63	.62	***(.31)***		
M				**-.50** ***(-.58)***	-.44	-.46	***(-.39)***			**.35**	.31		**.34**	.37	
N	**-.51** ***(-.67)***	-.51													
O							**-.67**	-.65	-.58	**.36** ***(.76)***	.39				
Q1							***(-.68)***						**.60** ***(.49)***	.65	.50
Q2	**-.69** ***(-.81)***	-.67	-.37												
Q3				**.64** ***(.82)***	.58	.71									
Q4										**.50** ***(.86)***	.39	.37			

Fac 1: Extraversion **Fac 2:** Self-Control **Fac 3:** Tough-Minded **Fac 4:** Anxiety **Fac 5:** Independence

Table 4: Factor Pattern Loadings from SEPATH Structural Equation Modelling solutions

US1: US (N=3498) matrix loadings reported in Table 1.3, p.14, of the US Technical Manual
US2: **US (N=2500) 16PF-5 normative data correlation matrix sample (Appendix 5B, p. 94 US Technical Manual)**
UK: UK 16PF-5 reference sample
Non-App: the non-applicant sample (N=589)
App: the applicant sample (N=506)

	Fac 1					Fac 2					Fac 3					Fac 4					Fac 5				
Scale	**US1**	**US2**	**UK**	Non App	App	US1	US2	UK	Non App	App	US1	US2	UK	Non App	App	US1	US2	UK	Non App	App	US1	US2	UK	Non App	App
A	**.74**	.61	.56	.56	.49						**-.35**	-.11	-.49	-.40	-.37										
C																**-.70**	-.86	-.80	-.78	-.68					
E																					**.87**	.78	.79	.58	.65
F	**.70**	.57	.63	.65	.53	**-.39**	-.39	-.40	-.29	-.25															
G						**.78**	.77	.69	.79	.66															
H	**.44**	.51	.53	.51	.46																**.43**	.37	.43	.46	.40
I											**-.75**	-.30	-.72	-.61	-.85										
L																**.57**	.54	.47	.45	.24	**.31**	.22	.32	.19	.17
M						**-.58**	-.49	-.53	-.41	-.49	**-.39**	-.24	-.21	-.16	-.15										
N	**-.67**	-.51	-.54	-.55	-.36																				
O																**.76**	.67	.62	.63	.60					
Q1											**-.68**	-.93	-.34	-.45	-.15						**.49**	.39	.43	.45	.56
Q2	**-.81**	-.69	-.67	-.68	-.60																				
Q3						**.82**	.47	.56	.46	.70															
Q4																**.86**	.51	.44	.41	.43					

Fac 1: Extraversion **Fac 2:** Self-Control **Fac 3:** Tough-Minded **Fac 4:** Anxiety **Fac 5:** Independence

* The two shaded loadings represent serious departures from the expected values.