

SELF-APPRAISAL PERSONALITY INVENTORIES

Laurence Paltiel

In the past four years the use of self-appraisal personality inventories in personnel selection has risen at an accelerated rate. If one is to believe distributors, over a million candidates have been asked to complete one of a range of such instruments within this period. Whilst competition between the various distributors of these inventories is increasing, there is no sign that the demand has reached a plateau. The British Psychological Society's (BPS) Test Standards Committee has not concealed its concern. In response to numerous letters from members, it recently wrote to personnel journals denying Thomas International claims that their Personal Profile Analysis (PPA) was BPS approved. A letter also appeared in the *BPS Bulletin* to this effect. NFER-NELSON, a major publisher and distributor of test materials in the UK, have started distributing the very similar PAL Personality Profile System (PPS) under licence. In response to the reaction from many of their clients, they have initiated a low profile study into the system. SHL also considered it worthwhile conducting a comparative study of their Occupational Personality Questionnaire and the PPS, distributed by NFER-NELSON to add to their validity section.

We have obtained seven self-appraisal personality inventories, each claiming to measure the same four personality categories: Dominance, Influence, Steadiness and Compliance. The trait names are not always the same but the descriptions remain almost identical. In addition to the PPS and PPA above, Cleaver, Performax, McQuaig all produce a similar self-descriptive questionnaire. A recent addition to this series which we have not yet seen is BEST. In the latter the personality traits measured are Bold, Expressive, Sympathic and Technical. In addition there also exists a computer program, Life-Line, directed at the games market, available for Sinclair Spectrums and BBC micros (at £14.99) which administers the test and provides an interpretation.

Each of the above share a common ancestry. Examination of the respective materials reveals that, in addition to the categories, they share most of the adjectives which form the basis of the self-appraisal. These it is claimed are based on the root words of William Marston (1928). Publication dates on the materials provide a clue to their origins. Cleaver's Self-description was first made commercially available in 1959, McQuaig in 1961, Performax 1977, Thomas International PPA in 1981 and PAL PPS in 1985. Each claims copyright of their materials. The Cleaver, Thomas PPA and PAL PPS also share the same technical manual. Thomas has copied it into their own format whereas PAL use the Cleaver original with the author's permission. Performax is identical to the Thomas PPA in order of presentation and 94 out of 96 items but has its own 'technical' manual as does McQuaig which is least like the others. The Life-Line program's items are closest to those in Performax, PPA and Cleaver although they too maintain copyright.

The claims and uses made by each of the distributors vary considerably. HAY/MSL, whose US operation acquired the Cleaver business in the US use it as a counselling and training device in limited situations. Life Skills Associates market Performax for management development or team building and play down the selection element. One of their manuals, however, claims the test as a 'splendid management tool' for selection as well as placement, career guidance and resolving interpersonal conflict. McQuaig's sole UK distributor, Svend Holst emphasizes its utility in selection as an instrument 'designed by businessmen for use by businessmen'. It is recommended by a variety of companies, one of which found it to be 'the best investment (this company) has ever made'. Thomas International say their PPA is 87 per cent accurate although it is unclear exactly what they mean by this. According to NFER, PAL's PPS is 'an effective system for improving selection decisions'. The director of PAL feels that for some high-ranking positions, it may not be sufficient and is presently examining alternative methods such as handwriting analysis to supplement this appraisal.

PPA and PPS

The aim of this report is to examine two widely used self-descriptive instruments, the Thomas International Personal Profile Analysis (PPA), and the PAL Personality Profile System (PPS).

One- and two-day courses are sufficient to become a 'qualified' user of PPA and PPS. BPS guidelines for personality inventories require a basic course on psychometric testing of five days duration plus five days on the specific test. The Thomas courses do not have psychometric content. Validity is described as 'how good a respondent considers his/her own result'. This is indeed a very powerful message for most untrained recruiters and has been effectively incorporated into their high-profile selling technique. Franchisees of these materials will offer to visit a potential customer and administer and interpret the test on the spot. If the customer likes the analysis, which is often the case, he/she is sold on the concept. With a one-day course (and the option of 'advanced user status'), the customer becomes a 'fully trained' test-user. In NFER-NELSON's two-day courses basic psychometrics is not covered, as according to NFER-NELSON the PPS is a system, not a test. The academic criteria adopted by the BPS in approving courses apply to 'tests' not to 'systems'.

The emphasis is clearly on the ease of use. The respondent is presented with 24 sets of four adjectives in a row and has to select the most descriptive and the least descriptive. The adjectives within each set usually represent the four categories (Dominance, Influence, Steadiness and Compliance). The most descriptive score is given a weight of +1 and the least descriptive -1. So for instance, if the



respondent selects 'attractive' as the most descriptive this will count as +1 towards 'Influence' and if least descriptive -1. Some items have only a +1 or -1 weighting and in the case of the PPA one item 'Trusting' contributes to two categories, positively on Steadiness and negatively on Influence. In the PAL the former weighting has been removed. In the PPA one item 'Eager' does not contribute to any category whereas in the PAL it is a 'Dominance' item. This unusual scoring procedure has considerable bearing on the structure of the test, which will be discussed later.

The scoring mechanism on the PPA is based on colour coding the items on the reverse page as either purple, red, black or green with combinations of fully enclosed box (for +1 and -1), line above (+1) and line below (-1). This results in considerable confusion and possible scoring errors. The author of the PAL, who, as a previous distributor of PPA, experienced this problem, revised the scoring by having a single carbonized sheet for each category. Responses are transposed by carbon on to the appropriate sheet below the question page. This works well for the first two scoring sheets, but unless the respondent has pressed very firmly indeed, his/her responses are scarcely visible on the third and fourth. We were told that this had been greatly improved in more recent copies of the test.

Validity and reliability

A question we have to ask about a particular test is whether it is a useful measure of the psychological characteristic it is attempting to assess? This question is usually asked in two parts: Firstly, is the test reliable and secondly, is the test valid. Reliability is the extent to which a person's score on a test is affected by factors which are extraneous to it, and thus introduce measurement error. Validity is the appropriateness or relevance of inferences drawn from test scores or other forms of assessment.

The issue of reliability needs to be addressed first. If a test is not reliable, it cannot be valid. Clearly, if a person's test score is mostly due to measurement error, the test cannot be measuring what it is attempting to measure and thus cannot be valid.

We can test reliability by using statistical techniques to measure internal consistency of the items, i.e. to see if items which are designed to measure the same dimension show a high degree of correlation.

Validity, on the other hand, needs to be examined in a different way. The construct validity of a test is the extent to which it actually measures what it purports to measure. One way of demonstrating this is to correlate the test with other validated tests which measure the same or closely related dimensions. Obviously this method of demonstrating construct validity is only as good as the existing psychological tests against which they are measured.

To demonstrate predictive validity of a test is different from construct validity. In a selection situation we may wish to know whether a test is a good predictor of job performance or that it discriminates between criterion groups such as more or less effective job incumbents. These questions will be examined in the following section.

Evaluating the PPA and PPS

The claims made on behalf of the PPA and PPS are empirical by their nature. Maintaining 87 per cent accuracy or suggesting that 40 per cent of the selection decision should be based on test scores constitute quantitative empirical claims. Even suggesting that an instrument (test or not) is 'an effective system for improving selection decisions' is a claim which requires at least substantive if not direct empirical support. The best means of evaluating empirical claims is by empirical methods. A 'technical manual' was written in 1958 in the US for the Cleaver test. Although Thomas and PAL provide this manual with their tests, it is not legitimate to use the data reported in this manual to evaluate the PPS and PPA which are different in many respects. The present report is based on PPA and PAL results on 919 job applicants.

One of the first problems to be mentioned concerning the PPA and PPS is that the items used to measure the four dimensions are not independent. Testees are required to indicate which four adjectives most and least describe themselves. The decision that one of the four dimensions most describes themselves, implies that one of the remaining three dimensions least describes themselves. The respondent is thus forced to indicate that he/she has a lower score on the remaining three dimensions, when in reality the testee could rate high on each of the four dimensions. This results in the dimensions being totally interdependent in a way which is impossible to disentangle.

The PPS and PPA tests assume that scores, derived from those items which respondents see as least and most describing themselves, together with the composite of these two scores (most-least), provide measures of different characteristics. In the case of PPA this is 'how others see you', 'behaviour under pressure' and 'how you see yourself'. PPS describes 'how others see you' as 'personality outward (mask)'. The hypothesis that least and most measure different characteristics is fairly implausible. Nowhere do Thomas International or PPS provide evidence to support it. What seems more plausible is to assume that 'least scores' measure the reverse end of the dimension measured by 'most scores' which is confirmed by the scoring. So being least decisive is the same as being most indecisive. The two test scores (most and least) obtained for each of the characteristics can be viewed as alternative halves of a test which

SELF-APPRAISAL PERSONALITY INVENTORIES *continued*

measures the opposite ends of these bipolar dimensions.

This is demonstrated, in part, by the correlations presented in Table 1 between the most and least scores for the few factors in the PPS ($n=452$) and the most and least scores in the PPA ($n=467$) tests. The results suggest that the most and least scores factors D, I and S in the PPS and PPA tests are measuring opposite ends of the same dimension as would be predicted from the above argument.

This does not seem to be the case for factor C as this result indicates that the most and least scores may not be treated as alternate halves of the same test. The Spearman—Brown split-half reliability coefficient was calculated for each of the four factors of these two tests. These results are presented in Table 2 and suggest that factors D and S are measured reliably by the PPS and PPA and I approaches an acceptable level of reliability for the PPS. Factor C is on the other hand not measured reliably by either PPS or the PPA and I on the PPA is below the acceptable level of reliability.

Table 1. Correlations between the most and least scores on the four factors of the PPS and PPA tests

	D	I	S	C
PPS ($n=226$)	-0.748	-0.57	-0.646	-0.35
PPA ($n=314$)	-0.696	-0.45	-0.564	-0.14

Table 2. Split-half reliability coefficients for the four factors in the PPS and PPA tests

	D	I	S	C
PPS ($n=100$)	0.84	0.64	0.73	0.55
PPA ($n=100$)	0.70	0.46	0.71	0.46

Tables 3 and 4 present the correlations between these four factors for the PPS and PPA respectively. These correlations suggest that for both the PPS and

the PPA factors S and D are measuring opposite ends of the same dimension. The correlation between Influence and Compliance on both the PPA and PPS are highly elevated, suggesting that these two categories have much in common.

Table 3. Intercorrelation matrix PPS ($n=452$)

	D	I	S	C
D	1	-0.11	-0.73	-0.46
I		1	-0.30	-0.50
S			1	-0.34
C				1

Table 4. Intercorrelation matrix PPA ($n=467$)

	D	I	S	C
D	1	-0.14	-0.71	-0.39
I		1	-0.26	-0.41
S			1	-0.27
C				1

In order to explore the validity of the present tests, the D, I, S and C of the PPA and PPS were correlated with the dimensions measured by the 16PF and OPQ respectively. Any associations with compliance must be viewed with extreme caution due to the low reliability of this factor on both the PPA and PPS.

Table 5 presents the correlations between factors D, I, S and C on the PPA and the subscales of the 16PF ($n=77$). As Table 5 demonstrates factor D (Dominance) in the PPA is most highly correlated with factor E (Assertive) on the 16PF. This finding is consistent with the hypothesis that factor D measures something akin to dominance.

Factor S (Steadiness) which appears to measure the reverse of factor D, is negatively (and more significantly) correlated with factor E of the 16PF providing further evidence to suggest that factor S is in fact measuring the reverse of factor D, something

Table 5. Relationship PPA and 16PF ($n=77$)

	Dominance	Influence	Steadiness	Compliance	MR
A Participating	-0.06	0.28	-0.16	0.04	0.40
B Bright	0.23	-0.16	-0.12	-0.01	0.19
C Calm Stable	0.18	0.27	0.03	0.00	0.28
E Assertive	0.55	0.26	-0.60	-0.33	0.63
F Enthusiastic	0.24	0.53	-0.41	-0.41	0.61
G Conforming	-0.17	0.06	0.03	0.20	0.26
H Socially Bold	0.31	0.65	-0.60	-0.35	0.77
I Tender-minded	-0.03	0.04	-0.09	0.10	0.19
L Sceptical	-0.24	-0.05	-0.20	-0.10	0.26
M Imaginative	0.32	-0.19	-0.22	0.02	0.36
N Diplomatic	-0.23	0.08	0.23	0.16	0.32
O Apprehensive	-0.17	-0.43	0.32	0.14	0.51
Q1 Experimenting	0.45	0.13	-0.36	-0.35	0.45
Q2 Self-sufficient	-0.06	-0.30	0.07	0.08	0.30
Q3 Controlled	-0.20	0.04	0.10	0.27	0.31
Q4 Tense Driven	0.02	-0.34	0.12	0.10	0.32

Table 6. Relationship PPS and OPQ (OCTAGON) ($n=286$)

	Dominance	Influence	Steadiness	Compliance	MR
Assertive	0.43	0.28	-0.43	-0.28	0.50
Empathy	-0.43	-0.11	0.48	0.14	0.49
Gregarious	-0.06	0.54	-0.12	-0.31	0.56
Abstract	0.11	-0.03	-0.10	-0.02	0.13
Methodical	-0.24	-0.37	0.22	0.44	0.48
Relaxed	-0.09	-0.03	0.17	-0.05	0.21
Self-control	-0.28	-0.33	0.33	0.29	0.42
Vigour	0.38	-0.04	-0.31	-0.14	0.39

more akin to submissiveness than steadiness. In addition to the above correlation factors D and S were also correlated to a lesser extent with factors Q1, (Experimenting) and H (Socially Bold) on the 16PF. While these traits are different from dominance, the notion of an assertive, experimenting and venturesome person is broadly consistent with the idea of a dominant, socially bold personality. Factor I (Influence) was highly correlated with factors H (Socially Bold), F (Enthusiastic) and O (Self-assured) on the 16PF, suggesting that factor I is in fact measuring something akin to a self-assured, enthusiastic, venturesome approach to life.

If we turn to the OPQ relationships with the PPS, we find that again Dominance and Steadiness correlate with the same factors; in this case with Assertive and Empathy. The correlation with Assertive could be seen as supportive of the validity of D although the equally high relationship with Empathy is more in line with the high S description. I correlates with Gregarious and negatively with Methodical.

Multiple correlations from PPA and PPS with each of the 16PF and OPQ factors reveal the extent to which factors measured by the OPQ and 16PF are measured by the best weighted combination of D, I, S and C. These are listed in Tables 5 and 6 respectively. 16PF factors that are well predicted include Socially Bold (H), Assertive (E), Enthusiastic (F), Self-assured (O) and Experimenting (Q1) which has about 20 per cent in common with PPA categories. 16PF factors C (Calm Stable), Q4 (Tense Driven), B (Bright), G (Conforming), I (Tender-minded), L (Sceptical), M (Imaginative) and N (Diplomatic) have less than 10 per cent in common with the PPA factors.

OPQ factors Assertive, Empathy, Gregarious and Methodical have over 20 per cent in common with PPS factors whereas Abstract and Relaxed share less than 5 per cent. This is broadly in agreement with the 16PF result suggesting that neither thinking style nor emotional stability are measured by the PPS or PPA.

Conclusions

The present results suggest that there is no reason to believe that the most, least and combined scores on the PPA and PPS describe respectively how a person is seen by others or his behaviour under pressure. A logical analysis of the test suggests that the most and the least scores for each of the four factors are in fact measuring the opposite ends of each of these factors. Correlations between the most and least scores suggest that this is the case for the factors D, I and S and that the most and least scores are best seen as alternate halves of a composite test. The Spearman–Brown split-half reliability coefficients suggested that the composite scores for factors D and S on both tests and I on the PPS formed fairly reliable subscales. Factor C on both the PPS and PPA is unreliable as is factor I on the PPA.

Correlations between the four factors suggested that factors D (Dominance) and S (Steadiness) were measuring opposite ends of the same scale. This finding was further supported by a validity study which correlated factors D, S and I of the PPA and PPS tests with the factors of the 16PF and OPQ respectively. These results suggested that factor D was measuring a dominant, venturesome and experimenting attitude towards life and that factor S (Steadiness) was measuring a passive, timid and conservative approach to life. These results therefore suggest that factor S should be reconceptualized as measuring something more akin to submissiveness than steadiness. The validity data further suggested that factor I (Influence) was measuring something more akin to a bold, enthusiastic and confident, approach rather than social influence.

The present results have therefore demonstrated that the PPA and PPS fail to live up to their own claims or the demands of empirical method. Factor C has proved to be very unreliable for both the PPA and the PPS and factors I and particularly S appear to be measuring constructs which are different from those described by the authors. This finding suggests that the analysis which is used to obtain thumb-nail sketches of the candidate's personality is probably invalid. This procedure places great emphasis upon the importance of interactions between the test's four factors, suggesting that various combinations of these four factors produce characteristic personality orientations which are more than the simple combination of each of the factors. If the test's four factors do not, however, have the psychometric properties they purport to have, then it seems highly unlikely that personality descriptions or 'personality types', which rely upon profiles based upon psychometrically invalid dimensions, could possibly be valid.

Consequently there seems to be little evidence to suggest that either the PPA, or the PPS can provide the detailed analysis of personality which the authors claim of their system of profile analysis. In reality, the PPA and the PPS approximate the 16PF and OPQ with regard to two of its factors by offering an alternative measure of dominance and sociability. It is worth noting that sociability is one of the most reliably rated personal attributes from interview (Ulrich & Trumbo, 1965). It is left to the recruiter or counsellor to decide whether these two dimensions are of sufficient relevance to make the exercise worthwhile.

References

- Cronbach, L.J. (1949). *Essentials of Psychological Testing*: London: Harper & Row
- Cleaver, J.P. (1965). *Self-description: A Technical Manual*. Princeton, NJ: Cleaver.
- Ulrich, L. & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin*, **63** (2), 100–116.
- PAL, PPA, OPQ, 16PF, Performax, McQuaig, Life-Line, Cleaver. Slough, Bucks: NFER.

Laurence Paltiel is an Occupational Psychologist with Psytech International Ltd