
Can a single item replace an entire scale: POP vs the OPQ 5.2

Authors: Paul Barrett
Chief Psychologist
Mariner7 Ltd
640 Great South Road
Manukau
Auckland
New Zealand

Email: paul.barrett@mariner7.com
Fax: + 64 9 262 6290

Laurence Paltiel
CEO Psytech International
The Grange
Church Road
Pulloxhill, Nr. Bedford
Bedfordshire
MK45 5HE
United Kingdom

Complete Publication Reference

Barrett, P.T. and Paltiel, L. (1996). Can a single item replace an entire scale? POP vs the OPQ 5.2. *Selection and Development Review*, 12, 6, 1-4

CAN A SINGLE ITEM REPLACE AN ENTIRE SCALE? POP vs. the OPQ 5.2

Paul Barrett and Laurence Paltiel

The Occupational Personality Questionnaire (Concept 5.2) (Saville et al. 1993) contains 248 items measuring 31 scales. Each scale has 8 items. Responses to each item are on a normative 5 point rating scale. Reliability coefficients (alpha) range from 0.57 to 0.88, with a median alpha of 0.75. Since alpha is known to depend on scale length as well as internal consistency, scales of just 8 items may achieve high levels of reliability (greater than 0.7) due to item redundancy e.g. where the items within a scale are simply reworded counterparts of one another. Instead of measuring a broad dimension of behaviour, it is possible that just one rather specific behavioural item is being assessed – using 8 very similar items to achieve this.

If this is the case with the Occupational Personality Questionnaire (OPQ), then it should be possible to replace each of the 30 personality scales (excluding the social desirability 'validity' scale) with a single composite item that captures the essential meaning of the scale and its constituent items. Thus the 240 items (30 scales) of the OPQ could be replaced by just 30 items. If item redundancy is a significant factor in the reliability of the OPQ scales, the following 'hypotheses' should hold true:

- 1. The alpha coefficient should be strongly and positively related to the correlation between the single and the OPQ scale score.**
To justify replacing 8 item scale with a single item, it is first necessary to show that the single item correlates highly with the full scale score. The stronger the relationship between this correlation and the alpha coefficient, the greater the item redundancy and the more likely it is that just one item is sufficient to capture the behaviour measured by the scale.
- 2. The difference between a single item score and the full scale score should be negligible.**
Where both scores are transformed to comparable scales of measurement, we should observe only small differences between them, and the larger the alpha, the smaller the difference.

QUESTIONNAIRE GENERATION AND ADMINISTRATION

A 30 item questionnaire (the 'POP' questionnaire) was generated by the authors, based simply on a perusal of the items composing each scale of the OPQ. Each POP item attempted to capture the composite meaning of all 8 items in an OPQ scale. Actual OPQ items were not used. The POP questionnaire is presented below in Appendix 1. The items all used an 11-point response format ranging from *Strongly Disagree* to *Strongly Agree* (see Appendix 2). Both the OPQ concept 5.2 and the POP questionnaire were administered concurrently to 420 managerial applicants (359 male; 61 female) as part of the corporate selection exercise.

RESULTS

To test Hypothesis 1 above, the correlations between each POP item and its respective OPQ scale score were computed. Each correlation was then corrected for unreliability of measurement of both the OPQ scale score and the single item¹. (See Table 1). It is important to remember that the purpose of this correction is solely to indicate what the maximum relationship between the POP item and the OPQ scale might be, *given no measurement error*. Obviously, these corrected values could never be observed in practice but they do serve to demonstrate how close the comparisons are, when adjusting for the observed unreliability inherent in the data.

Table 1 here

The median correlation (uncorrected between the POP items and the OPQ scales is 0.64. The median of the corrected values is 0.89. Allowing for unreliability of measurement, half of the OPQ scales correlate above 0.90 with a single item response. Figure 1 below shows the plot of the alpha coefficients (in decreasing size order) against the POP vs. OPQ scale score correlations. A Spearman rank order correlation between the two sets of data (alpha coefficients and POP vs. OPQ scale correlations) was computed as a direct test of Hypothesis 1. The observed correlation was **0.69** ($P < 0.0001$, one tail).

Figure 1 here

To test Hypothesis 2, the OPQ scale scores were linearly transformed from their full-scale score to one that matched the 11 point POP response scale. Then the POP scores were subtracted from the OPQ rescaled score, and their absolute (unsigned) values noted. The mean difference score was then computed across all respondents. These results are also in Table 1.

It can be observed that the mean absolute differences between the POP and OPQ rescaled scores are fairly low, with the exception of scales F6, F8, F9, and F10 (which all had mean absolute differences greater than 2.0). The mean of these data is 1.41, with median = 1.30. Figure 2 compares the profiles of the mean POP scores and the corresponding OPQ rescaled scores. On the same graph is shown the difference between these two profile plots, scale by scale. This shows clearly where the similarities and dissimilarities occur over the length of the profile plot. To test the specific prediction in Hypothesis 2, that the size of the score differences should be negatively related to the size of the respective alpha coefficient, we again computed a Spearman rank order correlation between these two items of data, which resulted in a coefficient of **-0.35** ($P = 0.030$, one tail). This is not so compelling as the previous 'order-effect' coefficient, but the result does remain conceptually important.

¹The single item unreliability could only be approximated. This was achieved by calculating actual item test-retest indices for three sets of personality questionnaire data, where the test-retest interval was approximately the same (10-14 weeks). The three datasets were composed of the USA version of the Eysenck Personality Questionnaire - Revised (EPQR: N=165, 106 items, 2 choice response), the Psytech Occupational Personality Profile (OPP: N=337, 98 items, 5 choice response), and Gordon's Survey of Interpersonal Values (SIV: Normative version, N=96, 90 items, 5 choice response). Altogether, 20 scales and 294 items were used. The results of this analysis indicated that the EPQR median test-retest coefficient was 0.59, the OPP was 0.63, and the SIV 0.64. The minimum and maximum coefficients respectively for each test were EPQR: -0.02 to 0.87; OPP: 0.32 to 0.86, and SIV: 0.37 to 0.82. From these data it was decided to use the most liberal estimate of test-retest, hence we took the median value of 0.64 as a reasonable estimate of overall expected item test-retest for the POP items.

Figure 2 here

Next, we looked at the range and variability of the difference scores – as an aid in judging the distribution of error around the median value for each scale. Since absolute differences tell us nothing about the proportion of over/under-estimated scores, we used the actual differences, summarised in a box and whisker plot in figure 3. This plot shows the median raw error for each scale, a box around this where the middle 50% of errors lie (the interquartile range), and the two 'whiskers' which show the minimum and maximum value observed for that scale.

Figure 3 here

From this graph, it can be seen that the interquartile range of errors around the median is confined in most cases to ± 1.0 or less. Also there is remarkable little bias shown toward gross under-or-over-estimated full-scale scores.

CONCLUSIONS

The results above indicate that a single item can capture virtually all of the measurement 'breadth' of many of the OPQ concept 5.2 scales. For one third of the scales, corrected correlations of 0.95 or above between a single POP item and the 8 items of the OPQ scale were obtained, which suggest that these scales are redundant. This is not to recommend that in fact just one item is used in future, because the inherent measurement reliability of a single item is significant. However, it is a moot point whether up to 8 highly repetitive items are needed. In two recent papers (Barrett and Paltiel, 1995a,b) we have shown that we could reduce 17 OPQ scales from 8 to 4 or 5 items, while maintaining alpha reliabilities at a comparable level to those for the full scale, and achieving correlations of 0.92 or more between the long and short versions of the scales in every case bar one.

Saville and Sik (1995a,b) have responded to our previous observations, indicating that it is better to err on the side of redundancy than to miss out a key item or concept. There are many issues in the 1995b rejoinder which we would like to make comment upon. However, we think the issues can now best be resolved by the reader. If you have any OPQ5.2 nearby, look at the actual items of the OPQ scales that correlate 0.95 and above with a single POP item, then decide for yourself if the arguments of Saville and Sik still hold. We agree that some scales are broader than others; we also agree that more than one item is required to reliably assess any behavioural construct. However, we disagree that significant reductions in questionnaire length cannot be made, whilst preserving all the measurement properties in the OPQ to which Saville and Sik allude. This would be a trivial point to argue except that such reductions can result in many kinds of indirect financial benefits to the test user.

Bibliography

Barrett, P. & Paltiel, L. (1995a) The Concept 5.2: Insufficient Measurement or Broad Occupational Assessment? *Occupational Psychology Conference Book of Proceedings* Jan 1995, 45-50

Saville, P. and Sik, G. (1995a) Reductio Ad Absurdum? *Selection and Development Review*, 11, 3, 1-3.

Barrett, P. & Paltiel, L. (1995b) Reductio Ad Absurdum: A reply to Saville and Sik (1995) *Selection and Development Review*, 11, 4, 3-5.

Saville, P. and Sik, G. (1995b) Hitting the Target: A rejoinder to Barrett and Paltiel. *Selection and Development Review*, 11, 4, 6.

Saville, P., Holdsworth, R., Nyfield, G., Cramp, L., and Mabey, W. (1993) *Occupational Personality Questionnaires: Concept Model Manual and User's Guide*. Esher: Surrey, Saville and Holdsworth Ltd

Appendix 1: The 30 POP items

1. I am at ease in social settings (**R6**: Socially Confident)
2. Variety and change appeal to me (**T6**: Change Oriented)
3. I am usually critical of people's ideas (**F6**: Critical)
4. I am modest about my achievements (**R7**: Modest)
5. I enjoy the discussion of hypothetical issues (**T7**: Conceptual)
6. I like hard physical activity (**F7**: Active)
7. I like the group to participate in making decisions (**R8**: Democratic)
8. I generate creative and innovative ideas (**T8**: Innovative)
9. I play to win (**F8**: Competitive)
10. I am sensitive to other people's problems (**R9**: Caring)
11. I enjoy forward planning and setting priorities (**T9**: Forward Planning)
12. I have ambitious career targets (**F9**: Achieving)
13. I always check the full detail of my work (**T10**: Detail Conscious)
14. I am a decisive decision-maker (**F10**: Decisive)
15. Once started, I always see a job through to the end (**T11**: Conscientious)
16. I like selling – whether ideas or products (**R1**: Persuasive)
17. I enjoy repairing objects or devices (**T1**: Practical)
18. Trivial things sometimes worry me too much (**F1**: Relaxed)
19. I like organising and taking charge of people (**R2**: Controlling)
20. I enjoy working with numbers and statistics (**T2**: Data Rational)
21. I get anxious when things go wrong (**F2**: Worrying)
22. I speak my mind even if it's unpopular (**R3**: Independent)
23. I appreciate the performing and literary arts (**T3**: Artistic)
24. I get upset if people criticise me unfairly (**F3**: Tough-Minded)
25. I am an outgoing and sociable person (**R4**: Outgoing)
26. I like analysing other people's behaviours (**T4**: Behavioural)
27. I often show my feelings and emotions to others (**F4**: Emotional Control)
28. I enjoy being in the company of others (**R5**: Affiliative)
29. I am described as something of a traditionalist (**T5**: Traditional)
30. I am an optimist (**F5**: Optimist)

Appendix 2: The response scale used for the POP items

Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Table 1: The comparison between the single item POP scores, OPQ scale scores, and the mean absolute differences between the scores from the two assessment instruments.

Scale	Normative Alpha	POP vs OPQ scale Correlation	Corrected Correlation	Mean Absolute Difference
R1	0.74	0.64	0.93	1.05
R2	0.88	0.73	0.97	0.81
R3	0.63	0.48	0.76	1.25
R4	0.86	0.69	0.93	1.47
R5	0.75	0.58	0.84	1.02
R6	0.86	0.64	0.86	1.31
R7	0.75	0.68	0.98	1.55
R8	0.65	0.58	0.90	1.35
R9	0.77	0.51	0.73	0.92
T1	0.87	0.88	1.00	1.07
T2	0.88	0.84	1.00	0.95
T3	0.83	0.79	1.00	1.30
T4	0.73	0.64	0.94	1.15
T5	0.74	0.56	0.81	1.50
T6	0.62	0.48	0.76	1.15
T7	0.75	0.68	0.98	1.43
T8	0.84	0.74	1.00	0.93
T9	0.57	0.51	0.84	1.31
T10	0.74	0.65	0.94	1.01
T11	0.80	0.63	0.88	0.99
F1	0.83	0.57	0.78	1.52
F2	0.73	0.56	0.82	1.82
F3	0.83	0.62	0.85	1.88
F4	0.76	0.66	0.95	1.37
F5	0.73	0.60	0.88	1.04
F6	0.60	0.27	0.44	2.50
F7	0.79	0.82	1.00	1.21
F8	0.71	0.64	0.95	2.32
F9	0.63	0.54	0.85	2.16
F10	0.76	0.34	0.49	2.87





