

The Distortion of Meaning and Measurement in Applicant Sample Personality Questionnaire Responses

Paul Barrett and Rosalie Hutton

email: p.barrett@liv.ac.uk & rb@psytechnology.softnet.co.uk

Website: <http://www.liv.ac.uk/~pbarrett/paulhome.htm>

Affiliations:

Paul Barrett: The State Hospital, Carstairs and
Dept. of Clinical Psychology, Univ. Of Liverpool

Rosalie Hutton: Psychometric Technology Ltd

The Issue

Can we assume that when using personality questionnaires, the only difference between applicants for jobs, and non-applicants, is one of score level on any particular scale?

The issue revolves around the kind of distortion of responses that may be taking place. It may be **systematic, non-systematic**, or a mixture.

Systematic Distortion

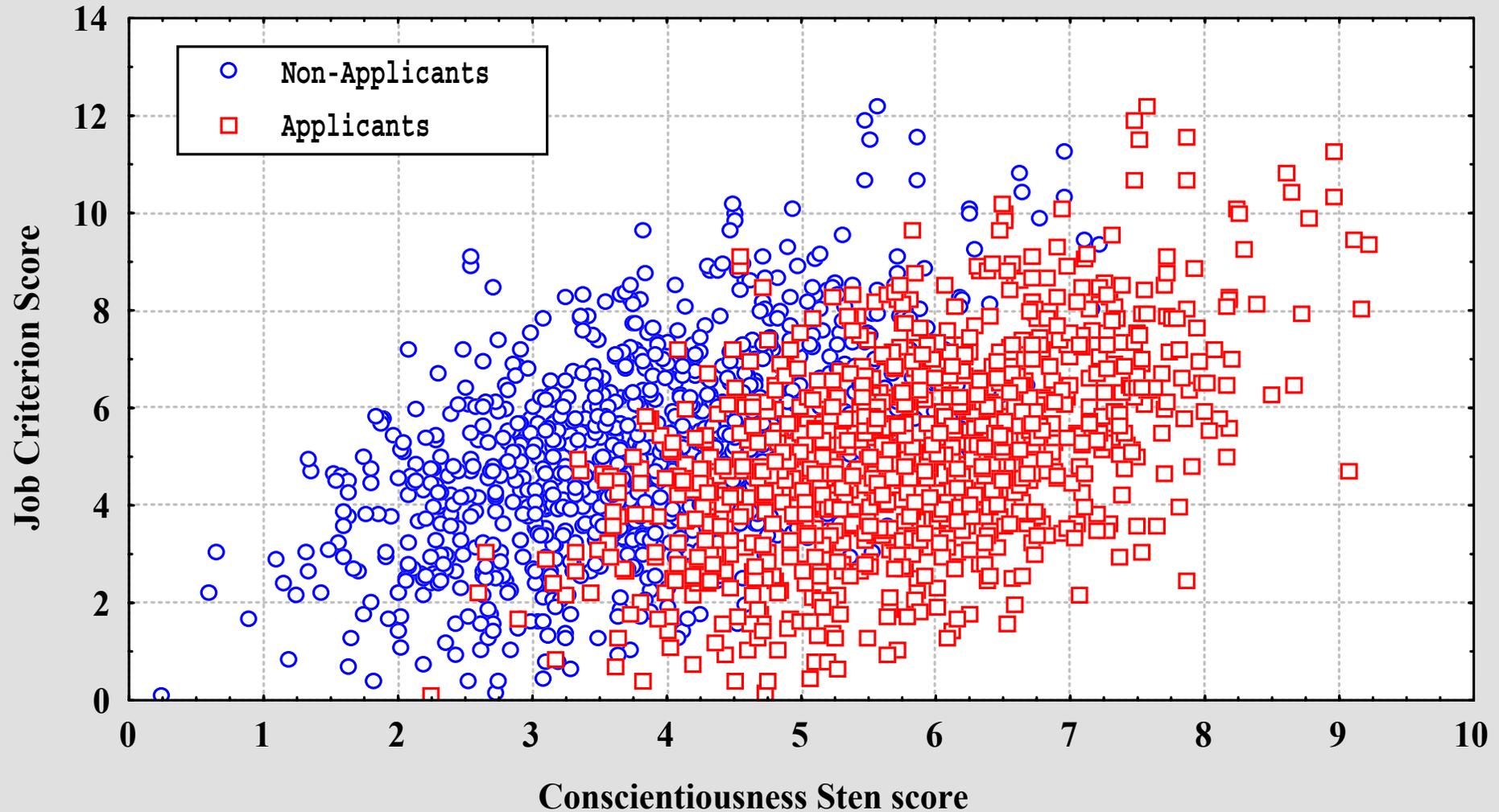
When candidates distort their responses, this can be **systematic** in that scale scores are elevated by some constant across all candidates (everybody tends to increase their scores say on conscientiousness). This kind of distortion has no effect upon the affected trait scale score and some criterion - and is a possible explanation for the Barrick and Mount (1996) result (distortion yet equivalent predictive validity).

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 2 applicant samples.

Results from structural equation modelling indicated that scores on both dimensions were distorted by both response styles. However, neither type of distortion attenuated the predictive validity of either personality construct.

Systematic Variation: Hypothetical Applicant and Non-Applicant Data

Correlation between Job Criterion score in each group = 0.52

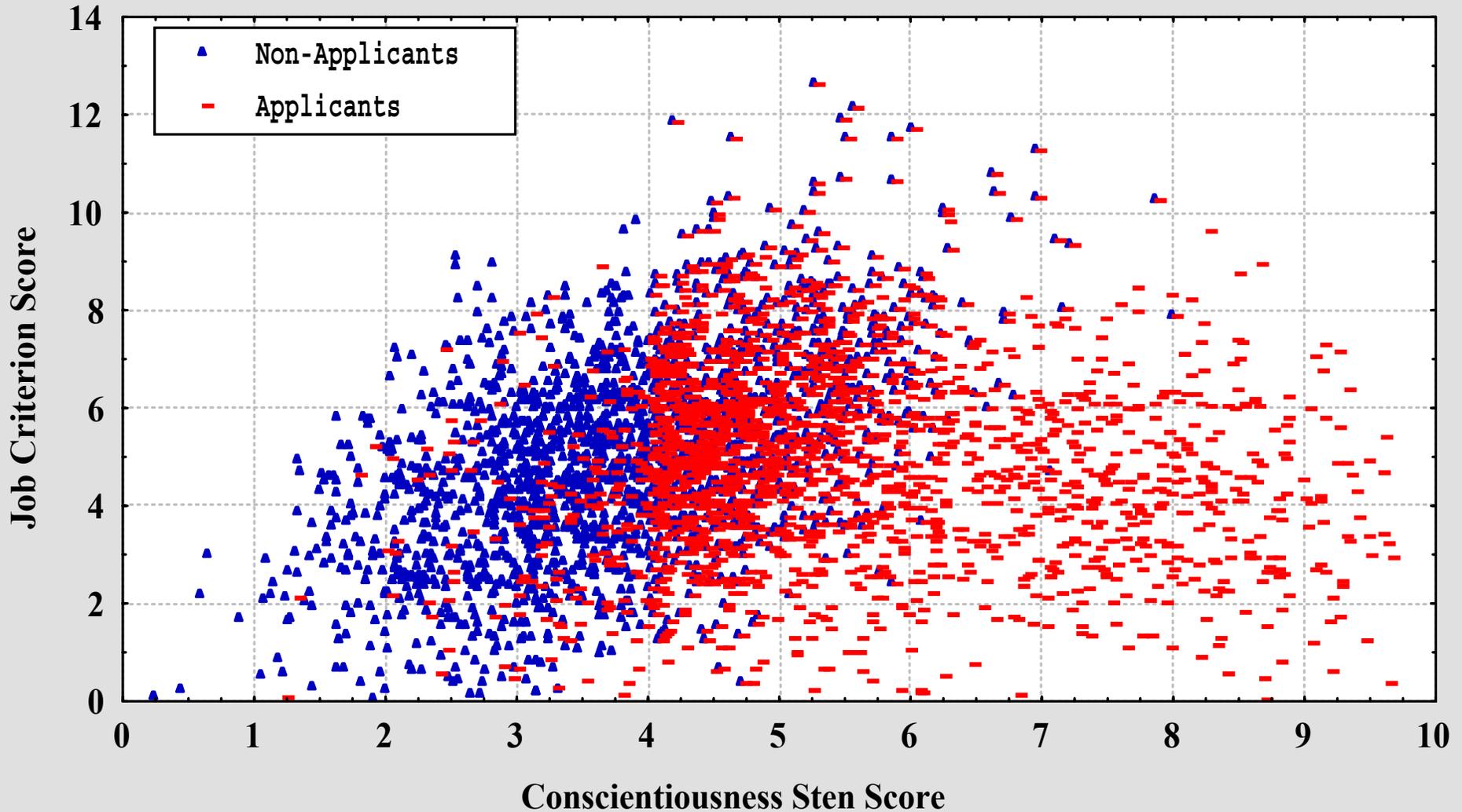


Non-Systematic Distortion

Alternatively, the distortion may be **non-systematic**, with certain candidates obtaining elevated scores whilst others remain static. This kind of distortion has unpredictable consequences upon trait-criterion correlations. An example below shows what happens when “true low-scorers on conscientiousness” tend to fake-good at a rate relative to the size of their low scores, whilst average to high scorers maintain their “true” score

Non-Systematic Distortion of Conscientiousness scores (Faking Good)

Non-Applicant correlation = 0.52, Applicant Correlation = -0.08



A further problem

What happens if response patterning to items in a test causes a change in the expected psychometric structure for a test?

Schmit and Ryan (1993) - an examination of the NEO 5 Factor Inventory structure in applicant and non-applicant (student volunteers) populations. They used Structural Equation Modelling to ascertain fit of each sample's data to the expected 5-factor model. The expected factor structure did fit student volunteers, but failed to fit the applicant data.

So...

Brown (now Hutton!) and Barrett (1999) examined the 16PF5, looking specifically for structural changes in the 2nd-order factor pattern (applicant vs non-applicant data), using structural equation modelling....

Sample Data for the 16PF5 Analyses

- ◆ **UK ASE N=1575 Non-Applicants mixed-gender correlation matrix**
- ◆ **US IPAT N=2500 Non-Applicants mixed-gender correlation matrix**
- ◆ **UK N=589 Non-Applicants mixed-gender correlation matrix**
- ◆ **UK N=506 Applicants mixed-gender correlation matrix**

The Key Result...

P. 14, 1st para., US Technical manual, “global factor equations were developed using only those primary scales having a loading of .30 or greater ...”

From Table 1.4,
p.16, US
Technical manual

$$\begin{aligned}\text{Extraversion} &= .3A + .3F + .2H - .3N - .3Q2 \\ \text{Anxiety} &= -.4C + .3L + .4O + .4Q4 \\ \text{Tough Minded} &= -.2A - .5I - .3M - .5Q1 \\ \text{Independence} &= .6E + .3H + .2L + .3Q1 \\ \text{Self-Control} &= -.2F + .4G - .3M + .4Q3\end{aligned}$$

From the UK
applicant dataset
SEM analysis

$$\begin{aligned}\text{Extraversion} &= .3A + .3F + .2H - .3N - .3Q2 \\ \text{Anxiety} &= -.4C + .3L + .4O + .4Q4 \\ \text{Tough Minded} &= -.2A - .5I - .3M - .5Q1 \\ \text{Independence} &= .6E + .3H + .2L + .3Q1 \\ \text{Self-Control} &= -.2F + .4G - .3M + .4Q3\end{aligned}$$

* Greyed scales fail to be identified

F= Lively, **L**= Vigilance, **M**= Abstractedness, **Q1**= Openness to Change

So - we looked at two other questionnaires...

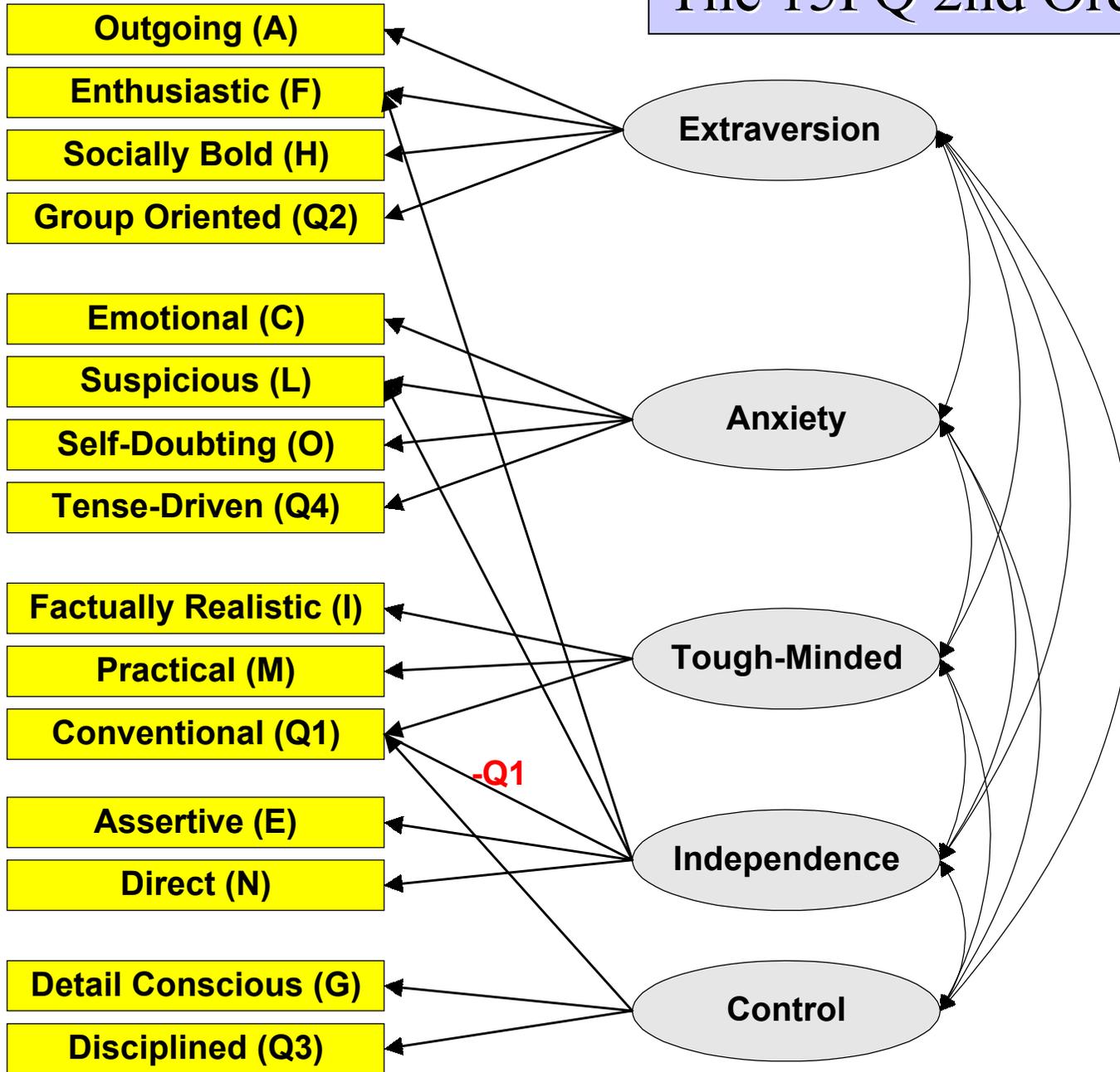
Sample Data for the Psytech 15FQ Analyses

- ◆ **UK N=355 Non-Applicants (#1) mixed-gender item data**
- ◆ **UK N=459 Non-Applicants (#2) mixed-gender scale data**
- ◆ **UK N=416 Applicants (#1) mixed-gender scale data**

Sample Data for the Saville and Holdsworth Concept 5.2 OPQ

- ◆ **UK N=279 Applicants mixed-gender item data**
- ◆ **UK N=392 Non-Applicants mixed-gender scale data**
- ◆ **UK N=291 Applicants mixed-gender scale data**

The 15FQ 2nd Order factor model



The 15FQ 2nd Order - **Extraversion**

	Non-Applicant #1	Non-Applicant #2	Applicant #1
FA	.75	.76	.78
FC			
FE			
FF	.58	.56	.55
FG			
FH	.84	.82	.80
FI			
FL			
FM			
FN			
FO			
FQ1			
FQ2	.70	.71	.68
FQ3			
FQ4			

The 15FQ 2nd Order - Anxiety

	Non-Applicant #1	Non-Applicant #2	Applicant #1
FA			
FC	.69	.68	.74
FE			
FF			
FG			
FH			
FI			
FL	.42	.39	.53
FM			
FN			
FO	.84	.87	.82
FQ1			
FQ2			
FQ3			
FQ4	.71	.70	.77

The 15FQ 2nd Order - Tough Minded

	Non-Applicant #1	Non-Applicant #2	Applicant #1
FA			
FC			
FE			
FF			
FG			
FH			
FI	.66	.66	.68
FL			
FM	.87	.85	.78
FN			
FO			
FQ1	.41	.42	.37
FQ2			
FQ3			
FQ4			

The 15FQ 2nd Order - Independence

	Non-Applicant #1	Non-Applicant #2	Applicant #1
FA			
FC			
FE	.70	.74	.62
FF	.39	.42	.37
FG			
FH			
FI			
FL	.28	.29	.27
FM			
FN	.67	.72	.67
FO			
FQ1	.33	.32	.33
FQ2			
FQ3			
FQ4			

The 15FQ 2nd Order - Control

	Non-Applicant #1	Non-Applicant #2	Applicant #1
FA			
FC			
FE			
FF			
FG	.59	.60	.54
FH			
FI			
FL			
FM			
FN			
FO			
FQ1	.24	.25	.01
FQ2			
FQ3	.71	.72	.67
FQ4			

It is here that the whole approach adopted so far collapses

Helen Baron (1996) ... in response to the Barrett, Kline, Paltiel, and Eysenck paper in JOOP ...p.22, 3rd para..

“The attempt at confirmatory factor analysis is also misguided. OPQ scales are divided into three broad domains: relationships with people, thinking style, and feelings and emotions. There is no claim that these domains are unidimensional or even that they represent higher order factors. They are merely collections of scales which relate to different aspects of behaviour...”

Questionnaire Construction - a 2 stage process

★ **MANDATORY** ... Choose (which tends to mean **assume**) a **model for your measurement** (e.g. extensive, additive unit concatenation, equal interval units, ordinal etc.). Your test theory (classical, IRT, or ordinal) will be embedded within the properties of measurement you have assumed for your variables.

★ **OPTIONAL** ... Choose whether you wish to formally **structure your measures** (e.g. factor analysis, multidimensional scaling, cluster analysis) - basically, specify particular covariance relations between variables that will define a specific structure (e.g. the first and second order factor models of the 16PF5 and 15FQ).

So ...

The 15FQ and 16PF5 (and NEO) are tests which assume additive unit concatenation, equal-interval measurement, a classical true-score model, and impose a structure on the covariances between scales (the second order/global factor patterns).

The OPQ also assumes additive unit concatenation, equal-interval measurement, a classical true-score model, **BUT**, **imposes no structure or particular covariance patterning amongst these scales.**

Thus, it is pointless looking for structural distortions in the OPQ data, as there is no *a priori* structure to be tested

Now what?

It seems the only summary indices left to us in this situation - indicating whether or not a scale of items retains its measurement properties - is either the internal consistency/mean inter-item correlation for a scale, or the criterion predictability of the scale scores. Since we have no criterion evidence, we can only evaluate the OPQ internal consistencies (alpha reliabilities) for evidence of response distortion.

For this comparative analysis, we can use the OPQ BMRB volunteer UK standardisation sample data, compared with our applicant item-level dataset.

OPQ Alphas - Applicants vs Non-Applicants (BMRB data)

Scale	UK	Apps	Scale	UK	Apps
Persuasive (R1)	.74	.74	Conceptual (T7)	.75	.73
Controlling (R2)	.88	.85	Innovative (T8)	.84	.87
Independent (R3)	.63	.45	Forw. Planning (T9)	.57	.63
Outgoing (R4)	.86	.85	Detail Consc. (T10)	.74	.81
Affiliative (R5)	.75	.71	Conscientious (T11)	.80	.74
Soc.Confid. (R6)	.86	.80	Relaxed (F1)	.83	.82
Modest (R7)	.75	.89	Worrying (F2)	.73	.72
Democratic (R8)	.65	.62	Tough Minded (F3)	.83	.81
Caring (R9)	.77	.69	Emot. Control (F4)	.76	.82
Practical (T1)	.87	.92	Optimistic (F5)	.73	.72
Data Rational (T2)	.88	.90	Critical (F6)	.60	.61
Artistic (T3)	.83	.86	Active (F7)	.79	.80
Behavioural (T4)	.73	.73	Competitive (F8)	.71	.77
Traditional (T5)	.74	.73	Achieving (F9)	.63	.65
Change Orient. (T6)	.62	.58	Decisive (F10)	.76	.71

What do we have so far?

- ★ **NEO** - some distortion to 2nd Order factor structure
- ★ **16PF5** - some distortion to 2nd Order factor structure
- ★ **15FQ** - No distortion to 2nd Order factor structure
- ★ **OPQ Concept 5.2** - no substantive alpha discrepancy

What do we conclude?

- ★ Even where some distortion of the 2nd Order factor pattern is occurring in the 16PF5, it seems to have had no discernable effect on the practical utility of the test. The same is apparent for the NEO.
- ★ When looking at tests such as the Concept OPQ, that have no defined structure, but are merely proposed as sets of useful scales (with some redundancy of measurement), then, the latitude of interpretation for these scales is so wide as to make irrelevant any sizeable perturbations in a minority of the scale alphas.
- ★ So, on this basis, we conclude that applicant distortion of responses on the questionnaires analysed, even when it is detectable, actually makes no difference to their practical use.

How can this be?

- ★ There are no units of measurement for any personality test, yet all tests assume additive concatenation and equal-interval units. **In reality, all tests are making ordinal measurement until proven otherwise.**
- ★ Subjective interpretation of test scores is given greater priority by the area than is accuracy of construct measurement. This is evidenced in part by the introduction of the UK BPS Level B accreditation procedure, and by last year's conference debate on the use of test scores (actuarial vs clinical/subjective-expert interpretation).

The Reality?

The fact is, personality questionnaires are not designed to make accurate measurement of *a priori*, well-specified meaning (*for which theory and rules for instantiation of constructs is required*). Instead, they make approximate ordinal measurement of approximate meaning. **This can still be useful.** However, it does mean that it is quite wrong to use quantitative methods of analysis (*and evaluation of theory and measurement from the perspective of science*) that far exceed the measurement expectations of these kinds of tests (*it is for this reason that I now find myself agreeing with Helen Baron's response to the 1996 OPQ analysis paper!*).

Of course, this begs the question (*hinted at by Drs. Hogan and McHenry*)... **why use equal-interval test theory any more?**

References

Baron, H. (1996) An evaluation of some psychometric parameters: A response to Barrett, Kline, Paltiel, and Eysenck. *Journal of Occupational and Organizational Psychology*, 69, 21-23

Barrick, M.R., and 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.

Barrett, P., Kline, P., Paltiel, L., & Eysenck, H. J. (1996) An evaluation of the psychometric properties of the concept 5.2 Occupational Personality Questionnaire. *Journal of Occupational and Organizational Psychology*, 69, 1-19

Brown (Hutton), R. and Barrett, P. (1999) Differences between Applicant and Non-Applicant Personality Questionnaire Data: some implications for the creation and use of norm tables. *BPS Test User Conference, Scarborough*. In published Conference Proceedings, pp. 76-86. *Leicester: British Psychological Society*. **Presentation** available for download from: <http://www.liv.ac.uk/~pbarrett/present.htm>

Schmit, M.J., Ryan, A.M. (1993) The big five in personnel selection: factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, 6, 966-974