Edition 64
Spring 2012

# Psyche

## The newsletter of The Psychometrics Forum

Visit our website at **www.psychometricsforum.org**

You will find information about forthcoming events, speakers, the origins of the group and much more.

Want to add something to the website? Contact our Administrator – **Caro Leitzell: admin@leitzell.com**

Also keep up to date with developments by signing up to **The Psychometrics Forum Affiliates Group** on LinkedIn

# Interpreting Commercial Personality Test Scores: Separating the Subjective from the Objective

*Paul Barrett – MD Advanced Projects R&D Ltd.*

*Honorary Professor of Psychology, University of Auckland, NZ.*

*Associate Professor of Psychology, University of Canterbury, NZ.*

I'm being asked a question by an increasing number of psychologist-practitioners whose practice might be described as Individual Psychological Assessment. That is, they specialize in the psychological assessment of one or more individuals for a client, using a variety of assessment tools, including psychometric personality assessment via a typical self-report personality test. There is a considerable amount of subjective judgement involved in how they arrive at a series of statements about an individual's psychological make-up and how the mix of information they have acquired from an individual will translate into likely future performance. Some practitioners even avoid using the usual narrative normative-data test reports produced by many test publishers, preferring instead to develop interpretations using the questionnaire item responses and/or raw test scores.

Usually, the question is asked of me after a practitioner has attended a conference, where they have found themselves criticized by some for being so 'subjective', for not being 'scientific', or for being a 'dinosaur' in the face of modern commercial psychometrics test practices.

The question, put simply, is: "Compared to those who use publisher test reports to inform their judgments about an individual's personality, am I being unscientific in how I arrive at my decisions"? My short answer is: "You are being no more scientific or unscientific than someone who uses an automated report derived from any of the main personality measures".

Clearly, at a mechanical level, a computer-generated narrative test report is utterly objective, in that given a set of input test scores, fixed algorithms are applied which will produce the same or very similar text-based output and interpretation. So, it is a form of standardized reporting. But, it can be very misleading when based upon normed scores using highly skewed norms, as interpretation is based not upon the number of actual behaviours endorsed, but on the comparison between a group whose scores might nearly all be located across the top quartile range of raw scores. So, an individual who scores say 8 out of 12 with a percentile score of 30 might then be interpreted as 'low" on say "emotional Intelligence", which actually makes no sense at all if the score is meant to convey a preference for behaving in a certain way.

The design, construction, calibration, and validation of psychometric tests is where serious attention is paid to the objectivity of assessment. The score-key is the end-result of the 'appliance of science'. Now, if we wish to continue as scientists, we would we cautious about making the assumption that any psychological attribute varies according to the axioms of quantity that underpin the SI unit measures within physics (*for that is what all who use conventional 'metric' methods of analysis*

*assume*). So, what we would do is acquire evidence of frequencies of occurrence of certain behaviours associated with the magnitudes of test scores. If for example we propose that the higher scores on an attribute lead to greater prevalence of derailer behaviour, we would make observations of derailer behaviours associated with assessed employees within organizations, thus enabling us to state a factual probability of occurrence of specific or generic derailer behaviours associated with each magnitude attribute score.

Within a science of personality assessment, when a statement is made about the likelihood of occurrence associated with a particular test score, one might expect it to be accompanied by a quantified probability of that occurrence. But, with personality testing at least, the reality is that the science grinds to a halt after the score-key development. If you do not possess facts about the frequencies of occurrence of certain behaviours associated with a particular test score, you have to move to a more subjective interpretation of what those scores might indicate for future outcomes. You may embed someone's personal expert interpretation in a computer program or rely upon the judgment of an experienced psychologist to interpret the scores. The end result is the same, the objective 'measurement' embedded in the item responses, raw test scores, and correlational 'validity' evidence, has been augmented by subjective interpretation.

There are ways of designing and analyzing tests such that a cumulative response function can be computed, based upon actual frequencies of occurrence of certain behaviours, with those scoring very high on such a scale showing greater quantified prevalence of the construct criterion behaviours. The reporting information for such tests is in terms of the likelihood of occurrence of one or more particular events occurring. There is no subjective interpretation of a test score; its 'meaning' is provided by the cumulative probabilities of occurrence of the behaviours which form the cumulative scale. However, subjectivity can and does enter into how to weight such information in a decision-making process, unless more formal empirical work is undertaken to develop a classifier function in which the test scores may be embedded and used mechanically.

Ultimately, It is not my intention to argue the pros and cons of individual psychological assessment (*although that is a topic worthy of close examination as has taken place recently in a target article and series of peer commentaries in the APA SIOP journal Industrial and Organizational Psychology, Vol. 4, 2011*). I simply want to suggest that a similar degree of subjectivity is prevalent in how practitioners prefer to use the results of applying a score-key to a set of self-report questionnaire items. In my view, to claim 'standardized test reporting' is more scientific merely because one or more individuals have embedded their subjective expert judgments in a computer program is mistaken. Standardization does bring 'regularity', but such regularity is invariably overridden in practice by the 'optional' interpretative judgments of users, who like the individual expert psychologist-practitioner, find that they often need to take context and other information into account in order to arrive at a particular decision.

And, there is a new concern that as test publishers compete to produce more reports based upon 'plausible reasoning' rescoring, re-interpreting the same set of items in a test and changing attribute names and descriptions to make them more commercially attractive, the degree of subjectivity inherent in producing such narrative reports eventually overwhelms the objectivity of the psychometrics. Paradoxically, under such conditions, an individual psychologist-practitioner's more consistent use of item responses and psychometric test results may well prove more accurate for a client in the long run.

**Paul Barrett**
**MD Advanced Projects R&D Ltd.**

*Honorary Professor of Psychology, University of Auckland, NZ.*
*Associate Professor of Psychology, University of Canterbury, NZ.*