Basic Psychometric Issues Surrounding Performance Prediction

JB Peterson

Professor of Psychology, University of Toronto

January 6, 2011

Definitions:

abstract reasoning: *the ability to generate dynamic low resolution mental models of a subset of experience, verbal or nonverbal*

complex job: *any job consisting of tasks that are too unpredictable to transform into an algorithm*

trait: *an enduring pattern of biologically-biased value, typically manifest in five dimensions among human beings, affecting motivation, perception, emotion, cognition, and behavior*

INTELLIGENCE:

Intelligence as the quantitative construct Intelligence Quotient or IQ has been studied since the time of Spearman, who noted that all tests requiring the use of abstract reasoning covaried positively, with an average of between .2 and .3 (a medium to large statistical relationship, by the standards of biomedical or social science research).[1] The situation is somewhat complicated by the fact that the average intercorrelation is higher for comparatively low IQ individuals (.44 for individual tests; .60 for individual tests and IQ) than for high (.23 and .26, respectively) (Detterman & Daniel, 1989).

Spearman noted that analysis of any sufficiently large, random sample of such tests would reveal a latent variable, g, which is roughly equivalent to the average of the test score performance (although factor analysis can produce a better estimate than the simple average).

Once g is derived, it correlates very highly with g derived from any other sample of tests of abstract reasoning (r > .9). Furthermore, g itself is generally a better predictor of future performance on tasks which require abstract reasoning than is any single test of abstract reasoning.

How intelligence should be partitioned at higher resolution levels than g is subject to debate. It is arguably reasonable, however, to consider g's differentiation in terms of nonverbal and verbal ability, which correspond, roughly, to what has been considered fluid and crystallized intelligence. Verbal ability can be tapped quickly and easily using measures of vocabulary, which are potent indicators of acquired knowledge (probably because vocabulary tests assess literacy efficiently and accurately). Nonverbal ability, which is strikingly similar to g, can be most efficiently and accurately measured with progressive matrices tests, which require the decomposition of a sequential complex

---

[1] One of the consequences of the fact that the typical single abstract reasoning test may be correlated at between .23 and .44 is that researchers will improperly claim divergent validation for their single tests of cognitive ability (often of low reliability) when they obtain such correlations with g.

pattern into its constituent elements and generalization from that analysis. It is possible that prefrontal ability of various sorts, including "working memory" may also constitute a usefully measured subcomponent of general mental ability or intelligence, particularly among those of high ability (Higgins, Peterson, Lee & Pihl, 2007). "Working memory" itself was equated with general reasoning ability (deductive and inductive) (r~.8) by Kylonnen and Christal (1990), although the separability of their "working memory" measure from g is debatable. It is also possible that tests of verbal or ideational fluency, which are usefully correlated with "creativity," can add additional useful information (Carson, Peterson & Higgins, 2005; Runco, Millar, Acar & Cramond, 2010), although tests of trait personality measuring "openness" may be as or more efficient in this regard, assuming valid testing.

How intelligence should be measured is a question that cannot be answered without considering (1) available time and (2) purpose. Progressive matrices tests provide the most efficient measurement/time payoff. The addition of a vocabulary test, secondarily, makes good sense, assuming equivalent familiarity with the testing language across the sample. Working memory or other prefrontal tests may add additional useful information, if time allows.

Construct Equivalents

Nonverbal IQ: g, fluid intelligence, working memory (may have additional predictive validity), reasoning ability (deductive and inductive), SAT, GMAT, LSAT, etc. (nonverbal portions)

Verbal IQ: crystallized intelligence, knowledge, vocabulary, verbal comprehension, SAT, GMAT, LSAT, etc. (verbal portions)

PERSONALITY

Personality is generally measured by self or other report questionnaire. Both provide valid sources of information. Such questionnaires have, at minimum, a five factor structure if they sample a sufficiently wide library of test items. Those factors have been termed extraversion, surgency, or dominance (and sometimes novelty seeking or impulsivity), with the opposite pole termed introversion; neuroticism or negative emotionality, with the opposite pole termed emotional stability or self-esteem; agreeableness (maternality and empathy at one pole, predatory aggression at the other); conscientiousness, an admixture of orderliness and industriousness; and, finally, openness to experience, which is decomposable into intellect (best measured by IQ rather than personality tests) and openness proper, which appears to be equivalent to ideational fluency and divergent thinking.

The strong Big Five advocate claim is that personality has a five factor structure. Whether or not this claim is true is dependent in part on what phenomena are deemed part of the personality realm. The weak Big Five claim is that human traits, as measured by descriptive adjective sets, phrases, or questions, have at minimum a five factor structure, that may be reduced to two (stability: conscientiousness, agreeableness, emotional stability; plasticity: extraversion, openness) at a lower level of resolution and increased to ten at a higher (extraversion: gregariousness and dominance; neuroticism: withdrawal and volatility; agreeableness: politeness and compassion; conscientiousness: orderliness and industriousness; openness: creativity and intelligence). The weak Big Five claim is all but proven, insofar as such things can be proven in psychological science.

What this means, essentially, is that a good Big Five measure reduces all other questionnaire measures to irrelevance. This is a bitter pill for experimental psychologists to swallow, contaminated as their thinking is with quasi-alchemical categorization schemes. Here is a brief breakdown of the relationship of some common psychological constructs into their big five categories:

Extraversion: incentive reward sensitivity, positive emotion, happiness, well-being, dominance, narcissism (with low agreeableness), optimism, gregariousness, sociability (along with agreeableness), positive illusions, impulsivity (of certain types, along with low agreeableness, high neuroticism, and low conscientiousness).

Neuroticism: self-esteem (reversed), self-efficacy (reversed), life satisfaction (reversed), job satisfaction (reversed) depression, anxiety, defensive aggression, irritability, low social status, sensitivity to stress (see Judge, Erez, Bono & Thoreson, 2002; Judge & Bono, 2001).

Agreeableness: maternality, empathy, sympathy, warmth, compliance (positively), tender-mindedness, femininity (along with higher neuroticism), "emotional intelligence" (Landy, 2005), altruism (short-term; kin-related); criminality, psychopathy, antisocial behaviour, assertiveness, tough-mindedness, masculinity, aggression, (along with lower neuroticism), competitiveness (along with higher extraversion).

Conscientiousness: dutifulness, orderliness, right-wing authoritarianism, industriousness, Protestant work ethic, disgust sensitivity, discipline, purity, frugality, obsessiveness and compulsiveness (along with higher neuroticism), conservatism.

Openness: intelligence, creativity, broad mindedness, mystical, artistic, poetic, literary, intellectual, philosophical, liberal (along with high agreeableness), ideational fluency (with IQ), insight problem solving, "thinking outside the box," inventiveness, entrepreneurial motivation.

Creativity and creative achievement are well predicted by openness proper and IQ. Academic and job performance are best predicted by IQ, conscientiousness and emotional stability, in that order (r > .6, in combination). Conscientiousness may be as good a predictor of job performance as IQ if (1) it is measured accurately and validly and (2) if industriousness is measured, rather than mere orderliness.

An efficient measurement battery assessing IQ (and, perhaps, working memory or other elements of prefrontal function) and personality can be administered, online, in 45-60 min.

PERFORMANCE PREDICTION

The potential economic benefits of developing more accurate tests for performance can hardly be overstated. Hunter and Hunter (1984) derived the following formula for determining the incremental dollar-value of test improvement in employee selection procedure (after an exhaustive analysis of the relevant literature):

\$Value= (employee N)*(tenure)*(wage)*(productivity SD)*(selection ratio standard score)*(procedure *r* increment)

Assume an organization with 10 employees, 5-yr. tenure, \$100 k salary, production SD of 1.00 [this is typical or even low for high-level sales positions (Hunter, Schmidt & Judiesch, 1990)] and the capacity to hire 1/10 job applicants (selection ratio standard score of 1.76[2]).

Now assume a validity improvement of a minimum of r= .35 by the expansion of a given selection algorithm (approximately equivalent to the addition of a good test of conscientiousness to a cognitive test or interview).

Use of the better tests would increase the productivity of the organization more than \$3,500,000 over the 5-year period – \$70,000 per year per employee.

What this means is that even small increments in predictive ability have very high potential for economic payoff. This is partly because of Price's Law (Price, 1963/1986): the square root of the number of people producing in a given domain produce half the output. This observation, which quantifies the massive difference in productivity that obtains between people, means that competence increases linearly as an organization grows, while incompetence increases exponentially.

Read alternatively, Price's Law means that a relatively small number of people will allows massively outperform the large majority of competitors. Thus, selecting higher up on the performance or ability distribution produces unexpectedly large economic gains.

---

[2] A selection ratio of 1.0 has a standard score of .00; .90– .2; .80 – .35; .70 – .50; .60 – .64; .50 – .80; .40 – .97; .30 – 1.17; .20 – 1.40; .10 – 1.76; .05 – 2.08 (Hunter & Hunter, 1984).

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. Creativity Research Journal, 17, 37–50.

Detterman, D.K. & Daniel, M.H. Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. Intelligence, 13, 349-359.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. Journal of Personality and Social Psychology, 91, 1138-1151.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. Journal of Personality and Social Psychology, 93, 880–896.

Digman, J. M. (1997). Higher-order factors of the Big Five. Journal of Personality and Social Psychology, 73, 1246–1256.

Higgins, D.M., Peterson, J.B., Lee, A. and Pihl, R.O. (2007). Prefrontal cognitive ability, intelligence, Big Five personality and the prediction of advanced academic and workplace performance. Journal of Personality and Social Psychology, 93, 298-319.

Hirsh, J. B., DeYoung, C. G., & Peterson, J. B. (2009). Metatraits of the Big Five differentially predict engagement and restraint of behavior. Journal of Personality, 77, 1085–1102.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits-self-esteem, generalized self-efficacy, locus of control, and emotional stability-with job satisfaction and job performance: A meta-analysis. Journal of Applied Psychology, 86, 80-92.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? Journal of Personality and Social Psychology, 83, 693-710.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! Intelligence, 14, 389-433.

Landy, F.J. (2005). Some historical and scientific issues related to research on emotional intelligence. Journal of Organizational Behavior, 26, 411-424.

Runco, M. A., Millar, G., Acar, S. & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty year follow-up. Creativity Research Journal, 22, 361–368.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75,* 28–42.

Price, D.J. De Solla. (1963/1986). Little science, big science and beyond. Columbia University Press. New York, NY.