Edition 74
Summer 2015

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@psychometricsforum.org

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

Follow us on Twitter @TPF_UK

The utility of psychometric tests for small organisations

Paul Barrett

Chief Research Scientist – Cognadev

Paul Barrett

Consider the evidence bases/validity coefficients for personality, EQ/EI, and other variants of psychometric tests presented in test manuals, academic articles, presentations at conferences; virtually all are drawn from large corporate organizations, government departments, the police and armed forces. For employee selection scenarios at smaller-size organizations, 'validity generalization' is invoked so as to assure these particular clients that even they can benefit from psychometric testing of prospective candidates. But can they really? The issue here is whether the expected positive benefit indicated by a validity coefficient computed over large numbers of cases within one or more corporates is at all noticeable by a smaller organization who may be employing fewer than 20 new employees each year.

What I want to look at is the likely real-world consequence of using psychometric tests for a small organization, in terms of whether any such organization would ever notice the claimed performance consequences which are meant to be accruing from the use of such tests.

For this exercise, in order to gain a clearer picture of the utility (or otherwise) of using psychometric tests in small organizations, I will need to model what will happen for thousands of such employers using a percentile cut-score on a test to screen-out or screen-in candidates.

Although it is possible to factor-in realistic constraints on sampling such as:

- how many people are actually available as serious potential 'candidates' for a job,
- what happens to that 'candidate availability' over time as more and more organizations seek to select from that subset while job-turnover puts some candidates back into the market,
- the variations in candidate 'quality' in terms of 'prospective job performance',
- whether discrete or composite 'profile' cut-scores are used, or just a single 'primary' scale,
- how test scores are used; for example, as the basis of subjective narrative-report interpretations,

- test-score and performance-outcome distributions not being 'normal', but rather skewed-beta or Pareto.

These constraints require more complex modelling that is beyond the focus of the current article.

The current modeling constraints and parameters

- Test validities are varied between 0.2, .03, and 0.4.
- Interview and other kinds of selection process validity is assumed to be a fixed constant at 0.2.
- Interview and other selection methods validity adds [0.1] to each level of test validity (because it is reasonable to assume that there will be some overlap between a test score and what can be judged about a person's test score from a variety of other information about a candidate).
- Number of employees to be selected: 5 or 10.
- Both test scores and performance are assumed to be normally distributed with sample values for both attributes are expressed as integers.
- Test scores are generated for a typical 15-item attribute scale, with Likert response range between 0 and 4 per item, giving a measurement range between 0 and 60, in integers.
- The mean population test-score is [38], with standard deviation of [7].
- Test cut-scores (select-in *at* or *above*): 30th [=34], 50th [=38] and 70th [=42] percentiles.
- Performance-outcome 'ratings' vary between 0 and 100, with the mean 'population' rating as [60], and standard deviation of [10].
- Candidates are grouped into five predicted performance groups based upon their performance rating:
    - 'Poor' [*below* 25th percentile = 53].
    - 'Below-Average' [*at or above* the 25th and *below* the 40th percentiles = 53, 57]
    - 'Average' [*at or above* the 40th and *at or below* the 60th percentiles = 57, 62]
    - 'Above-Average' [*above* the 60th and *at or below* the 75th percentile = 62, 67], and
    - 'Excellent' [*above* the 75th percentile = 67].
- 5,000 organization-samples of new employees are randomly drawn from the appropriate bivariate distribution for each of the 27 possible conditions (3x test validities x 3x cut-scores x 3x numbers of selected employees).
- A 'contrast' condition for each psychometric test-score validity condition is reported for a sample drawn with no prior knowledge of psychological test scores. This condition represents candidate selection using existing methods other than psychometric testing, with an implicit validity of 0.2.
- Each empirical bivariate data distribution consists of 100,000 cases of data (*rounded integers*) sampled randomly from the parameter-specific bivariate normal distributions.

The sampling sequence

1. Randomly select 5, 10, or 20 'new employees' from an empirical bivariate distribution with specified validity, where each candidate's test score meets or exceeds the specific test cut-score threshold. Do this 5,000 times, tallying the numbers of 'selected' employees classified within each performance group.
2. Express the numbers in each group as a percentage of the total employees selected. E.g. for selecting 5 employees, the total number selected is 25,000. For selecting 10 employees. The total number selected is 50,000.
3. Contrast these summary results with those generated from sampling within the 'contrast' bivariate distribution where cases are sampled entirely at random from the entire applicant distribution.

Model 1: Test score validity 0.2 and projected employee job-performance group-classification; three test cut-score thresholds.

Figure 1: Selection rates for Model 1

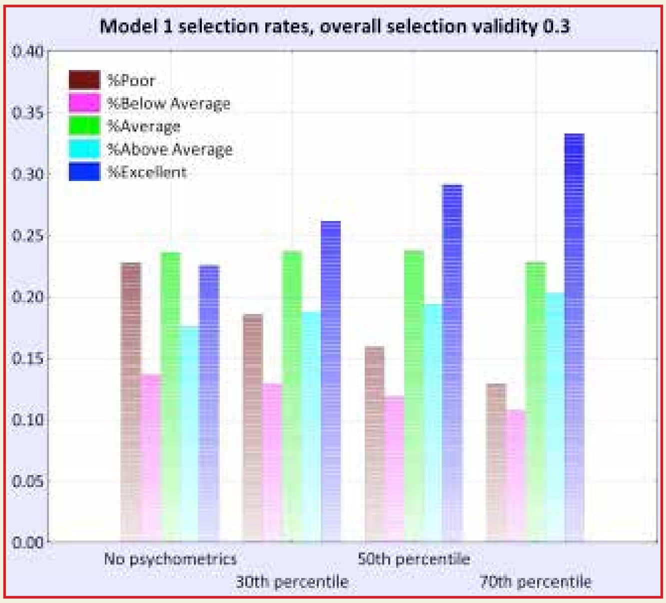
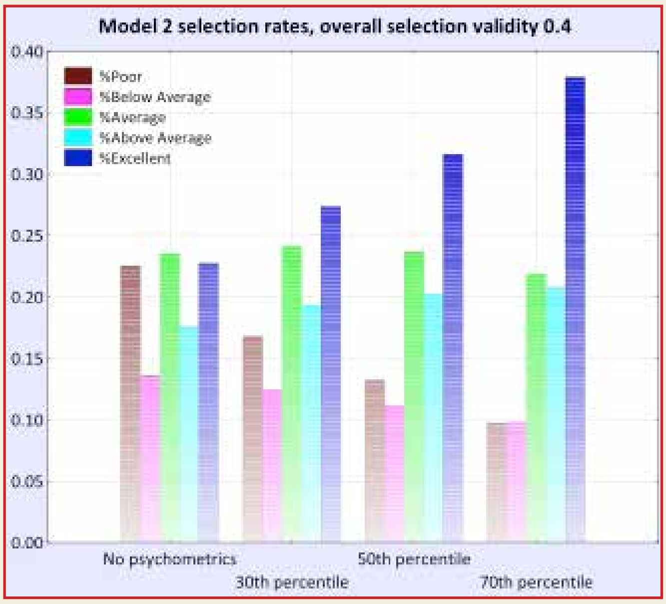


Figure 1 shows the average selection-rates across the five predicted job-performance groups, for a test score validity of 0.2, with a 0.1 lift in validity due to additional non-psychometric selection procedures (an overall selection validity of 0.3). Three test cut-scores are utilized: at or above the 30th, 50th, and 70th score percentile. The contrast group labelled 'no psychometrics' represents selection using an employer's totality of existing selection processes, assumed to possess a validity of 0.2.

Model 2: Test score validity 0.3 and projected employee job-performance group-classification; three test cut-score thresholds.

Figure 2 shows the average selection-rates across the five predicted job-performance groups, for a test score validity of 0.3, with a 0.1 lift in validity due to additional non-psychometric selection procedures (an overall selection validity of 0.4). Three test cut-scores are utilized: at or above the 30th, 50th, and 70th score percentile. The contrast group labelled "no psychometrics" represents selection using an employer's totality of existing selection processes, assumed to possess a validity of 0.2.

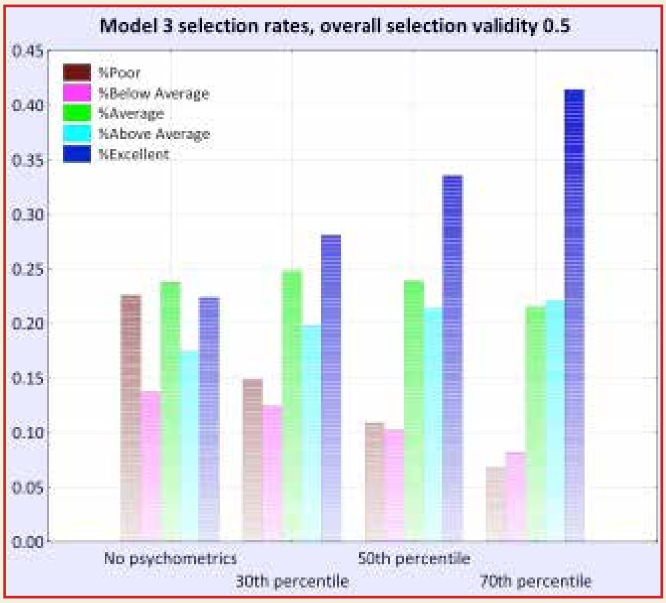
Figure 2: Selection rates for Model 2



Model 3: Test score validity 0.4 and projected employee job-performance group-classification; three test cut-score thresholds.

Figure 3 shows the average selection-rates across the five predicted job-performance groups, for a test score validity of 0.4, with a 0.1 lift in validity due to additional non-psychometric selection procedures (an overall selection validity of 0.5). Three test cut-scores are utilized: at or above the 30th, 50th, and 70th score percentile. The contrast group labelled 'no psychometrics' represents selection using an employer's totality of existing selection processes, assumed to possess a validity of 0.2.

Figure 3: Selection rates for Model 3



Although the graphs show clear trends in terms of rates of expected job-performance category employees, what's important here is seeing the actual numbers of employees selected given these rates, where only a few employees might be selected and employed over a year. Tables 1-3 show the numbers of selected new employees across two predicted performance groups: Below-Average and Average-&-Above, using the selection rates in figures 1-3, contrasted with the 'no psychometrics' selection group.

Table 1: Expected numbers of new employees across job-performance categories (5 selected)

Selection Rule	0.3 overall selection validity		0.4 overall selection validity		0.5 overall selection validity	
	Below average	Average & above	Below average	Average & above	Below average	Average & above
*No psychometrics	2	3	2	3	2	3
At or above 30th percentile	2	3	1	4	1	4
At or above 50th percentile	1	4	1	4	1	4
At or above 70th percentile	1	4	1	4	1	4

* The No psychometrics condition has a validity of 0.2 overall.

Table 2: Expected numbers of new employees across job-performance categories for Model 1 (10 selected)

Selection Rule	0.3 overall selection validity		0.4 overall selection validity		0.5 overall selection validity	
	Below average	Average & above	Below average	Average & above	Below average	Average & above
*No psychometrics	4	6	4	6	4	6
At or above 30th percentile	3	7	3	7	3	7
At or above 50th percentile	3	7	2	8	2	8
At or above 70th percentile	2	8	2	8	1	9

* The No psychometrics condition has a validity of 0.2 overall.

Table 3: Expected numbers of new employees across job-performance categories (20 selected)

Selection Rule	0.3 overall selection validity		0.4 overall selection validity		0.5 overall selection validity	
	Below average	Average & above	Below average	Average & above	Below average	Average & above
*No psychometrics	7	13	7	13	7	13
At or above 30th percentile	6	14	6	14	5	15
At or above 50th percentile	6	14	5	15	4	16
At or above 70th percentile	5	15	4	16	3	17

* The No psychometrics condition has a validity of 0.2 overall.

What might we conclude?

- Using a selection test with a validity of 0.2 looks to be of marginal impact unless you select above a substantive percentile (at least above the 50th percentile). The likelihood is that if screening candidates using a lower percentile, few employers would notice any 'stand-out' difference at all in the quality of employees they hired using a psychometric test as an initial candidate screen.
- Using a selection test with a validity of 0.3 is of lesser utility unless you select at or above the 50th percentile). But still there really doesn't look to be the kind of 'stand out effect' that would make psychometric testing an obvious addition to an already implemented selection process.
- Using a selection test with a validity of 0.4 looks more noticeable in its effects, although again it's not until you increase the cut-score percentile that the benefits really accrue. But this is beginning to look like a 'will make a substantive difference' result.
- Overall, from this very constrained analysis, I think it would be fair to conclude that for a small organization wanting to improve its employee selection using screening psychometrics, it should use tests with job-performance validity coefficients of at least 0.3, with screening cut-scores at or above the 50th percentile.

Important Caveats

- Although the figures in table 1 look better for any selection above the 30th percentile and test validity > 0.2, bear in mind these data are rounded integers. For example, the actual average value for Below-Average, at or above the 30th percentile is 1.46 (rounded to 1). But this is getting a bit too numerically picky given the host of simplifying assumptions outlined in the preamble of the modelling.
- It's easy for me to make recommendations about using more discriminating cut-scores, but the reality is that this strategy only works where sufficient candidates apply for a position who might exceed that cut-score threshold. What may happen is that by selecting a very discriminating cut-score, the employer sees no candidates at all. But the dilemma then arises that by adopting a less-discriminating cut-score, the impact of the psychometric screening will be also be diminished.
- This modelling assumed linear relations between the psychometric test-scores and job performance. Yet we know that for attributes like Conscientiousness, non-linearity is more evident than not (Le et al, 2011, Lam et al, 2014, Carter et al, 2015). So, simply using the results from this computational investigation may prove unwise when it comes to selecting candidates on those attributes shown to demonstrate curvilinear relationships with job-performance outcomes.
- What's also missing from this analysis is the cost-benefit analysis which factors in the cost of using psychometrics. But with many organizations offering instant 'have credit card – can test' options for screening-style assessments, such costs will likely be trivial. However, whether any of these 'instant' assessments possess validity coefficients of any veracity is a moot point. Without those validities, a client is forced to take a 'punt', which as this modelling shows, may result in no added advantage to how they currently select their new employees.

In conclusion

This was just a simple 'back of a matchbox' look at seeing whether a small organization might benefit at all from using screening psychometrics in its selection processes. Given the assumptions stated at the outset of the article, and the caveats at the end, I think a small organization might obtain significant 'noticeable' financial benefit if it uses psychometric tests which possess evidence of at least moderate validity (0.3 and above) and a discriminating cut-score. The modelling used here could be greatly refined, adopting the realistic usage and candidate-market-behaviour effects similar in approach to those used by Sturman's (2000) powerful investigation into the realistic vs advertised benefits of conventional utility analysis (i.e. after several real-world adjustments, the modelling showed a 96% reduction of the standard utility-formula projected 'saving'). And such 'intervention-effect' modelling need not be confined to the potential deployment of psychometric tests, but also to the deployment of expensive leadership and employee development interventions in large corporates, where costs may be in substantive and outcome expectations are based upon 'plausible reasoning' claims by providers rather than substantive empirical evidence-bases.

References

Le, H., Oh, I-S., Robbins, S.B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, 96, 1, 113–133.

Lam, C.F., Spreitzer, G., & Fritz, C. (2014). Too much of a good thing: Curvilinear effect of positive affect on proactive behaviors. *Journal of Organizational Behavior*, 35, 4, 530–546.

Carter, N.T., Guan, L., Maples, J.L., Williamson, R.L., & Miller, J.D. (2015). The downsides of extreme conscientiousness for psychological wellbeing: The role of obsessive compulsive tendencies. *Journal of Personality* (http://onlinelibrary.wiley.com/doi/10.1111/jopy.12177/abstract?campaign=wolacceptedarticle), InPress, 1–46.

Sturman, M.C. (2000). Implications of Utility Analysis adjustments for estimates of human resource intervention value. *Journal of Management*, 26, 2, 281–299.

Personality & politics: reviews of our 11th March event

Nicholas Bennett

Psychology, ideology and politics

Professor Adrian Furnham
Reviewed by Nicholas Bennett

Adrian Furnham, Professor of Psychology at University College, London (UCL) spoke during the morning session. He started by looking at what he referred to as the bottom line. In the relationship between Personality and Politics he argued that much depends on the measure of politics: such as beliefs/ideology; behaviour like voting, taking part in protests and, knowledge. He argued that all the main personality factors are involved, but that they account for relatively little variance. Other individual difference factors also play a part like 'the Dark Triad' and intelligence.

He discussed the **Authoritarian Personality**. Theodor Adorno et al (1950) described a fascist-prone individual as militaristic, conventional and anti-hedonistic, submissive to higher authority and punitive toward those below. A harsh upbringing was held to cause repressed hostility to the parents, emerging as ethnic prejudice and cruelty towards weaker persons. Typical authoritarian personality traits were listed and examples he gave included:

1. **Conventionalism**: rigid adherence to conventional middle-class values. 'Obedience and respect for authority are the most important virtues children should learn.'
2. **Authoritarian aggression**: a tendency to condemn anyone who violates conventional norms. 'A person who has bad manners, habits and breeding can hardly expect to get along with decent people.'

Coming from UCL it did not surprise me to hear Adrian refer to Hans Eysenck's original Two Factor Theory of Stable/Unstable and Introverted/Extraverted (before Neuroticism was added). Using factor analysis, Eysenck