DBS Business Review
ARTICLE | dbsbusinessreview.ie
03 | 2019

Psychometrics at Work: How to Ensure Test Results You Can Trust

Dr. Barbara Caska
Selection by Design
Dublin, Ireland

Abstract

Psychometric testing is considered as the intersection of the fields of psychology and business, with potential benefits to both employees and employers. Recommendations for maximising contributions of occupational testing are presented. These are evaluated across five phases of the assessment process. The first is choosing a test, bearing in mind both psychometric and practical qualities of measures. Next, ways to increase positive responses to testing are discussed. The process of administration is examined, with suggestions for improved accuracy. Recommendations for score interpretation are provided, taking measurement error into account. Finally, implications for communicating test results are drawn. It is explained that these facets of psychometric testing are key in ensuring accurate, meaningful and trustworthy workplace measurement.

Keywords: Psychometric testing; Selection; Employee assessment; Test user.

Introduction

Psychometric testing is now a well-engrained part of employee decision making within organisations. Assessment results are used to help evaluate suitability for hire, determine match with organisational values and culture, examine best-fitting career options, build effective work teams, and ascertain potential for leadership (Bailey, 2017; Kantrowitz, Tuzinski and Raines, 2018). Standardised testing offers the potential for accuracy, validity and fairness. Through tests, detailed information on candidates can be easily compiled and compared. As a result, decisions may be made in an efficient and cost-effective way.

Psychometrics, Psychology and Business

Psychology and business intersect at the point of psychometric testing. Psychology contributes the mechanisms of psychometric measurement, including test construction and validation (e.g., Kline, 2005). It also provides a pathway for understanding the impact of testing on individual employees. Businesses provide an avenue to widely apply test results, and to realise associated gains.

In applications such as selection, psychometric testing offers businesses the chance for improvements in the efficiency, validity and utility of employment assessment

(Schmidt *et al.*, 1979). Testing makes it possible to provide transparent and fair comparison of employee attributes. Positive impressions of employers may result. Those participating in work-related assessments may gain self-understanding, personal development, and defined opportunities for change and growth.

Key considerations are discussed below, with recommendations on how to ensure test results that are meaningful, useful and trustworthy.

1. Choose the Right Test

Before selecting a test for workplace applications, what needs to be measured must be clear. A systematic job or work analysis should determine the knowledge, skills, abilities and key characteristics required for a position (Cook, 2016). Based on this analysis, a job specification can be prepared, identifying the personal attributes needed to perform the work (Riggio, 2013). Informed decisions on how to best assess those qualities can then be made. Testing may be a useful option for measuring attributes needed for the job.

A multitude of work-related tests are available today. These include measures of ability and aptitude, ranging from job-focused administrative or mechanical skills to advanced reasoning. Tests of work values, motives and interests may help to ensure the right match between employees and their organisations. Measuring capacity for emotional intelligence, coping or resilience may be useful for career and leadership development. Personality type assessments such as the Myers-Briggs Type Indicator® can promote self-understanding and team building, while more in-depth trait measures like the NEO personality inventory and the 16pf® are invaluable in selection and talent management. Regardless of the kind of test, decisions as to which to use should take into account evidence of the psychometric qualities of the instrument. These include:

Reliability and Validity

Evidence for reliability, or consistency of the test, and the validity, whether it measures what it claims to, may be available from the test publisher. It is up to the test user to determine whether or not there is sufficient evidence of these qualities to meet the purpose at hand. For example, is there evidence that similar scores can be expected over brief periods of time? In other words, is there sufficient test-retest reliability? This question becomes important if a test will be used both for remotely administered, unproctored screening of job candidates and securely managed re-assessment to verify applicant scores before hiring.

Standardised scoring

With normative tests, performance is determined through comparison with scores from a group of persons who previously completed the assessment. In this way, scores are translated from raw data into standardised formats such as percentiles, t-scores or stens.

A question for consideration is whether the comparison or norm group used for a specific measure appropriately matches the present testing participants (e.g., Miller and Lovler, 2019). For example, when assessing verbal reasoning ability for a group of office administrator applicants, could a comparison group of professionals and managers be used? If so, the administrator candidates might be disadvantaged. The professional/managerial group may have high verbal ability scores due to advanced education or beneficial experiences. This would result in the administrator scores comparing poorly. A more suitable comparison group might be drawn from a general working population, representing persons from backgrounds that include office administrators.

Practical considerations should also play a role in determining which test best matches present assessment requirements. For example:

Accessibility

Is the test accessible, based on the purchaser's qualifications and background?
Specific training or qualifications are often necessary to purchase and work with psychometric measures. Requirements vary by the type of test and proposed use, so it is important to check with a test's publisher on qualifications needed to purchase or use the test. If regular work with psychometric measures is anticipated, it is helpful to complete a course such as the British Psychological Society/European Federation of Psychologists Associations (BPS/EFPA) Occupational Test User training (BPS, 2019). With in-depth measures, additional specific instruction may also be required in order to purchase and work with the test.

Pricing

Does use of a test fit within allowed budget?
Pricing varies widely across different kinds of assessments. Factors affecting costs include the format for administration, method of scoring, and type of report or results required.

Generally, ability and aptitude tests are less costly than in-depth measures of traits or complex attributes. For example, Selection by Design offers on-line tests of verbal, numerical and abstract reasoning for under €15. Complex, in-depth measures are generally priced higher. Selection by Design's 16pf competency reports range in price from €20 to €35, while personality reports range from about €25 for a profile report to around €85 for audience-tailored, comprehensive personality evaluations (https://www.selectionxdesign.com/16pf-report-options/). Similar patterns of pricing can be found across UK based test publishers, such as PSIonline, Criterion, and Hogrefe.

Paper and pencil formats and hand-scoring may offer savings, but are time consuming to administer and score. Many test companies are presently moving towards only offering on-line assessment and computer-generated results reports.

Security

Choice of format will also depend on needs for security. This relates to the question of how the test will be administered to test-takers.

As outlined by Bartram (2006), four different modes of test administration may be considered. These range from the open, uncontrolled mode to the very secure, managed mode. Many tests used in employment settings require either a single use login link, or supervised administration in a controlled setting. If remote testing is initially used, it may be followed with secured, in-person re-assessment to verify scores. Options available for administration mode vary by test, so it is best to check with the publisher on whether present needs can be accommodated.

For computerised administration, individually-tailored test versions created through adaptive modelling or randomised item ordering may be available. For example, with the Smart-aptitude series by PSIonline, items interactively adapt to the individual's response pattern as they progress through the test. Correct responses lead to more difficult items being presented for attempt, as increased ability is indicated. Each test taker may receive a unique set of questions. Adaptive tests, therefore, offer the potential to decrease the risk of cheating and improve security (Kantrowitz, Dawson and Fetzer, 2011; Sanderson, Chockalingam and Pace, 2011).

2. Maximising Engagement and Encouraging Honest Responses

Accurate measurement through testing depends on the process of administration. Steps taken prior to testing will influence the candidate's engagement with the assessment.

Maximising engagement during assessment means motivating candidates to work earnestly and honestly as they complete their tests. People need to see the value of putting effort into testing, and to consider potential outcomes of psychometric assessment worthwhile. To maximise engagement, make sure the process and purpose of testing is clearly explained to each test-taker. Consistent with this, a meta-analysis by Truxillo *et al*. (2009) indicated that job applicant motivation and performance were positively associated with explanations as to the job relevance of their selection procedures, such as testing.

Many work-related assessments involve high-stakes testing. Decisions based on test results may have important consequences for an individual's career, salary, and professional opportunities. Failing to land the "perfect" job or being bypassed for a promotion may have negative emotional, attitudinal, cognitive and financial effects on a person. Negative responses to assessment can also have long-term implications for organisations. For example, McCarthy, Hrabluik and Jelley (2009) suggested that discontentment with testing may deter candidates themselves from entering future competitions for promotion, and even lead to discouraging their colleagues from taking part. A reduced pool of applicants may make it difficult for a company to best fill job openings.

It is important that candidates trust the assessment process to be accurate, fair, and useful in identifying the best applicants. This is another way of saying that testing should be standardised, reliable and valid. It is up to those interacting with persons throughout the testing process to communicate these qualities in easily understood, non-technical terms. These practices are consistent with general recommendations for ensuring an effective employee selection process (Bauer *et al.*, 2012).

Reactions to testing will be improved if a test has both face and faith validity. Face validity relates to whether or not the test appears to be a good measure for its purpose (Bornstein, 1996). Does it look the part? Faith validity is the users' belief that the test really will deliver the benefits it claims (Bailey, 2017). Again, this is about the test-taker's impression of the measure. Face and faith validity are both subjective reactions that can make a significant difference in how a test is undertaken and responses to being assessed. In a meta-analysis examining job applicant reactions to the process of selection, Hausknecht, Day and Thomas (2004) found a sizable overall correlation of .60 between perceptions of face validity and procedural justice. This implies that applicants may consider the selection process fair if they see the job relevance of a test. Impression is important!

It is also helpful to explain clearly why candidates should respond accurately and truthfully when completing a test. This can include emphasizing why and how the test results will be used. Benefits to the person and the business can be clarified. For example, scores can help applicants decide whether they are a good match for a particular job.

It may be useful to ask applicants to sign an honesty contract (e.g., Bartram and Tippins), particularly if tests will be completed remotely and without proctoring. An honesty contract requires signed agreement from the test-taker on conditions such as completing the assessment independently and responding truthfully. The consequences of failing to do so should be explained. For example, will the person be eliminated from further consideration in the present campaign if they respond dishonestly? Means used to verify test results should be communicated. Will respondents be asked to re-take a test at a later point, particularly to check scores from unproctored assessments?

Drawing from Fahey (2018), highlighting the importance of honesty as a moral standard can reduce moral hypocrisy. Creating honesty is particularly powerful when combined with raising objective self-awareness. Moral hypocrisy is the tendency to act in ways that maximise self-benefit while maintaining an impression of adhering to ethical standards (e.g., Batson *et al.,* 1997).

Fahey (2018) further suggests the importance of using impression management scales to detect persons who endeavour to ‘fake good’ on personality tests. Such scales may be included within complex personality assessments. For example, both the 16pf and the Eysenck Personality Scales include social desirability subscales (British Psychological Society, 2020). Stand-alone measures may alternatively be used, such as the well-established Crowne-Marlowe social desirability scale (Crowne and Marlowe, 1960). Scores may then be examined to evaluate, or control for, potential misrepresentation when answering test questions.

Unlike other employee selection methods, tests often directly incorporate techniques to detect cheating. If tests are administered on-line, potential analyses include overall time taken to complete the full test as well as individual items, and patterns of incorrect vs. correct answers to questions which vary in difficulty (Sanderson, Viswesvaren and Pace, 2011).

3. Administering the Test

The way a test is administered affects the accuracy of results. Measurement through psychometric testing is not perfect, even under the most careful conditions. Classical test theory (e.g., Lord and Novick, 1968) posits that a test score is likely an approximate indication of one’s actual ability or trait level. Results may also reflect random environmental influences. Examples include aspects of the physical testing environment, such as lighting or noise. The administrator may be another influence on scores. An untrained or inexperienced administrator may provide test instructions that give advantages to some test-takers, but hinder performance of others. In addition, the mood, physical and emotional state of the test-taker will affect concentration and performance, shifting scores in one direction or the other.

Consistency and control over the assessment processes help to keep unwanted influences from affecting test scores (Coaley, 2014). Psychometric assessments typically have specific and standardised instructions for administration. These may include time allowed, a script for introducing the test, appropriate methods for administration, and ways to maintain security and maximise honest responding. Strictly adhering to these guidelines helps to ensure recommended procedures for testing are in place, and maximises the opportunity for accurate measurement.

4. Interpreting Results

Verify test scores

Test results should be treated as hypotheses rather than as flawless indicators. As stated by Cripps (2017, p. 18), “Scale scores on all instruments should be regarded as expectations or hypotheses and subject to movement.” Accordingly, verifying results is an important part of interpreting test scores. In part, this is accounted for by considering measurement error. Additional data from alternative sources may also be taken into account. For selection, this could include examining competencies through structured interviews. Ratings from others who know or work with the test-taker might be used as part of 360-degree feedback for development.

Candidates themselves can provide useful perspectives, including the way that they approached assessment and how they feel they performed. This may take place through an interactive, two-way discussion following assessment (Duggan, 2017). Information shared by the individual can indicate whether or not they were focused, motivated and optimally engaged in test-taking. If so, their scores are more likely to indicate their true attribute levels. If not, caution is in order.

Use Confidence Intervals

As discussed under ‘Administering the Test’, errors in measurement occur with testing. The use of Confidence Intervals considers error when interpreting scores.

Confidence intervals may be constructed using a statistic called the Standard Error of Measurement (Coaley, 2014). The interval is a range of scores that is likely to include an individual’s actual ability or attribute level. This range should be interpreted as an indication of performance, rather than assuming that a single score point (e.g., a ‘raw score’ of 57, or a percentile of 68) is completely accurate. Using confidence intervals increases the chance that you have really 'captured' or identified the person’s true performance. For example: You may not be completely certain that Mary’s score of 70 precisely indicates her computational skills, but you can be 68% certain that her numerical skills fall within the range of 62 to 78.

By taking standard errors and confidence intervals surrounding test scores into account, you can more fairly and justifiably make comparative decisions among candidates. This will require considering error surrounding both test-takers' scores.

For example: two candidates are being considered for progression to the interview stage of selection decisions. Steve's verbal ability test score is 53. Frank’s score on the same test is 57. Are these scores different enough to justify choosing Frank for the job? If you are not sure, imagine how Steve would feel, or what his reaction would be!

Carefully choose cut-off points

Give careful consideration to where to place a passing cut-off point for test scores. As illustrated by the Taylor-Russel model (1939), a choice of cut-off score is a key determinant of correct decisions being made as to who should be selected or progressed to the next stage of assessment. High requirements, such as the upper levels of a percentile range of ability tests, increase the chances that all persons passing will have very strong skills. For example, a cut-off point of the 75th percentile on a verbal reasoning test will result in the top 25% of persons considered as passing. Whether this is justifiable depends on the level and importance of verbal reasoning for the job in question. An important question becomes whether succeeding on that particular job really requires such high levels of this ability. Overqualified candidates may be a poor fit for work.

In addition, very high cut-off scores may result in adverse impact by eliminating persons from minority or protected groups. Bailey (2017) suggested that a conservative point of the 30th percentile may be used towards removing the lowest third of scores, while controlling for the likelihood of adverse impact.

5. Communicating and Storing Test Results

Communicating results

It is essential that the results of assessment are communicated to the intended audience in a way that is sensitive, confidential and understandable.

With work-related testing, test results or feedback from a single campaign may be needed for multiple audiences. Human resource management may require psychometric details on score results across numerous candidates. Non-technical reports are appropriate for audiences with limited backgrounds in testing, potentially including managers and the test-takers. Test publishers may offer reports tailored to user requirements and technical background (e.g., PSIonline, Hogan). An alternative option is for the test user to prepare their own reports tailored to audience and purpose.

Test data storage

Policies on test data storage, access and maintaining security vary from company to company. The BPS recommends that practitioners or businesses develop a test user policy, specifying practice in each of these areas (BPS, 2018). It is essential that test results are used, stored and shared in a way that is consistent with local requirements and legislation such as the General Data Protection Regulation (GDPR) and the Data Protection, Act, 2018.

Conclusion

Based in psychology and applied to businesses, psychometric testing has a great deal to offer both employees and employers. As discussed, wide-ranging benefits can be realised through including psychometric testing as part of workplace decisions.

However, ensuring accurate measurement requires planning and preparation across the stages of test selection, administration, result interpretation and communication.

Businesses may capitalise on the contributions of testing by incorporating suggestions across five areas: Begin with a careful choice of psychometric test. Take steps to maximise positive responses from test-takers. Carefully plan administration. Interpret results cautiously, seeking verification of apparent scores. Communicate test results sensitively, considering legal requirements for data protection.

Incorporating the recommendations presented throughout this paper will maximise the chance of achieving assessment results that can be trusted – the key to effective workplace testing.

REFERENCES

Bailey, R. (2017) 'HR applications of psychometrics', in Cripps, B. (ed.) *Psychometric testing: critical perspectives*. Chichester: Wiley Blackwell, pp. 87-112.

Bartram, D. (2006) 'The internationalization of testing and new models of test delivery on the internet', *International Journal of Testing*, 6(2), pp.121-131. Academic Search Complete, EBSCO*host* [Online]. doi: 10.1207/s15327574ijt0602_2 (Accessed: 5 November 2019).

Bartram, D., and Tippins, N. (2017) 'The potential of online selection', in Goldstein, H.W., Pulakos, E.D., Passmore, J., and Semedo, C. (eds.), *The Wiley Blackwell Handbook of the psychology of recruitment, selection and employee retention*, Chichester: Wiley Blackwell, pp. 271-292.

Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., and Wilson, A. D. (1997) In a very different voice: unmasking moral hypocrisy', *Journal of Personality and Social Psychology*, 72(6), pp.1335-1348. PsycArticles, EBSCO*host* [Online]. (Accessed: 5 November 2019).

Bauer, N. T., McCarthy, J., Anderson, N., Truxillo, D.M., and Salgado, J. F. (2012) *What we know about applicant reactions to selection: research summary and best practices*. SIOP White Paper Series. Bowling Green, OH: Society for Industrial and Organizational Psychology, Inc.

Bornstein, R. F. (1996) 'Face validity in psychological assessment: implications for a unified model of validity', *American Psychologist*, 51(9), pp. 983-984. PsycArticles, EBSCO*host* [Online]. (Accessed: 5 November 2019).

British Psychological Society (2020) 'Search test reviews: search over 160 comprehensive test reviews to assist you with test selection', Available at: https://ptc.bps.org.uk/test-registration-test-reviews/search-test-reviews (Accessed: 27 January 2020).

British Psychological Society (2019) 'The BPS qualifications in test use', Available at: https://ptc.bps.org.uk/bps-qualifications-test-use (Accessed: 5 November 2019).

Coaley, K. (2014) *An introduction to psychological assessment and psychometrics*. 2nd edn. London: SAGE.

Cook, M. (2016) Personnel selection: adding value through people – a changing picture. 6th edn. Chichester: Wiley-Blackwell.

Cripps, B. (2017) 'Ride the horse around the course: triangulating nomothetic and idiographic approaches to personality assessment', in Cripps, B. (ed.) *Psychometric Testing: Critical Perspectives*. Chichester: Wiley Blackwell, pp. 3-14.

Crowne, D. P. and Marlowe, D. (1960) ‘A new scale of social desirability independent of psychopathology’, *Journal of Consulting and Clinical Psychology*, 24, pp 349–354. PsycArticles, EBSCO*host* [Online]. (Accessed: 5 November 2019).

Duggan, G. (2017). ‘The practical application of test user knowledge and skills’, in Cripps, B. (ed.) *Psychometric testing: critical perspectives*, Chichester: Wiley Blackwell, pp. 65 – 76.

Fahey, G. (2018). ‘Faking good and personality assessments of job applicants: a review of the literature’, *Dublin Business Review*, 2, pp. 45-68. Available at: https://dbsbusinessreview.ie/index.php/journal/article/view/25. (Accessed: 5 November 2019).

Hausknecht, J. P., Day, D. V., and Thomas, S. C. (2004). ‘Applicant reactions to selection procedures: an updated model and meta-analysis’, *Personnel Psychology*, 57, pp. 639-683. Business Source Complete, EBSCO*host* [Online]. (Accessed: 5 November 2019).

Kantrowitz, T.M., Tuzinski, K.A., and Raines, J.M. (2018) SHL 2018 Global Assessment Trends Report https://www.shl.com/en/assessments/trends/global-assessment-trends-report/. (Accessed: 5 Nov 2019).

Kantrowitz, T.M., Dawson, C.R. and Fetzer, M.S. (2011) ‘Computer Adaptive Testing (CAT): A faster, smarter, and more secure approach to pre-employment testing’, Journal of Business Psychology, 26(2), pp. 227-232. Available at: https://www.jstor.org/stable/41474872. (Accessed: 5 November 2019).

Kline, T.J.B. (2005) *Psychological testing: a practical approach to design and evaluation*. Thousand Oaks, CA: Sage Publications.

Lord, F.M. and Novick, M.R. (1968) *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.

McCarthy, J., Hrabluik, C. and Jelley, B (2009) ‘Progression through the ranks: assessing employee reactions to high-stakes employment testing’, *Personnel Psychology*, 62, pp. 793-832. Business Source Complete, EBSCO*host* [Online]. (Accessed: 5 November 2019).

Miller, L.A, and Lovler, R.L. (2019) *Foundations of psychological testing: a practical approach*. 6th edn. Thousand Oaks, CA: Sage Publications.

Riggio, R.E. (2013) *Introduction to industrial/organizational psychology*. 6th edn. London: Pearson.

Sanderson, K.R., Viswesvaran, C. and Pace, V.L. (2011) ‘UIT practices: fair and effective?’, *The Industrial-Organizational Psychologist*, 48(3), pp. 29-37.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C. and Muldrow, T. W. (1979) ‘Impact of valid selection procedures on work-force productivity’, *Journal of Applied Psychology*, 64(6), pp. 609–626. Business Source Complete, EBSCO*host* [Online]. (Accessed: 5 November 2019).

Truxillo, D. M., Bodner, T. E., Bertolino, M., Bauer, T. N., and Yonce, C. A. (2009) ‘Effects of explanations on applicant reactions: a meta-analytic review’, *International Journal of Selection and Assessment*, 17(4), pp. 346-361. Academic Search Complete, EBSCO*host* [Online]. (Accessed: 5 November 2019).