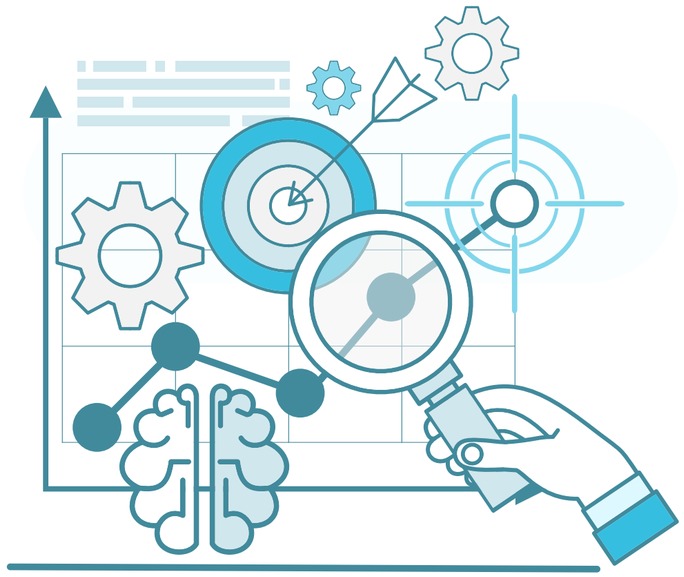
Quality Metrics of a Psychometric Test:

Why Should It Matter?

TABLE OF CONTENT

[1] Introduction



When we talk about quality, we often require reference. And if it's a conversation about psychometrics, a test within its domain ought to do the trick.

Myers-Briggs Type Indicator (MBTI) fits the criteria, owing to its extreme popularity. Reportedly, 80% of the Fortune 500 use it with their employees.[1] This includes 89 of the Fortune 100 companies also. In 2017 though, CNBC reported cited scientific concerns about the validity of the assessment.[2]

The test even has its set of consultants, which require certification. Nonetheless, and while it is difficult to believe, the test itself is built on zero scientific foundation.

It traces back to 1921 when famed psychologist - Carl Jung - unveiled a theory. He surmised that human beings could be broken down to 8 different personality types. But the theory came at a time before psychology used scientific methods such as data or controlled experiments, making the 8 types nothing more than guesses based on Jung's personal experience. [3]

These findings were later used in 1943 by mystery novelist Isabel Briggs-Myers and her mother Katharine Cook Briggs to lay the foundation of an influential test that would be used for decades to come.[4]

They doubled Jung's hypothesis of 8 personality types into 16, and published the test in 1944. Ironically, the very idea of personality types was dismissed by the original inventor himself - Carl Jung. [5]

"Every individual is an exception to the rule. This kind of classification is nothing but a childish parlor game."

Carl Jung, Swiss Psychiatrist and Founder of Analytical Psychology

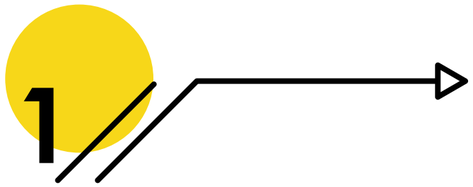
Many other tests have also spawned in the wake of MBTI, resulting in questionable qualities, rendering organizational processes dependent on the same as unscientific at best. Today, it's important to try and understand qualitative metrics that actually make a test effective. This is all the more important with psychometric tests.

[2] Deciphering & Using Test Results Accurately

Psychometric tests have found use in different stages of the employee life cycle, and also businesses close to the human resource job description - appraisals, hiring, learning & development and more. It's been known to increase chances of employee success given the correct use of both cognitive and personality tests, two of the most important components to a psychometric test.

But as explained in the introduction, too many organizations use the wrong psychometric tests in the wrong way.

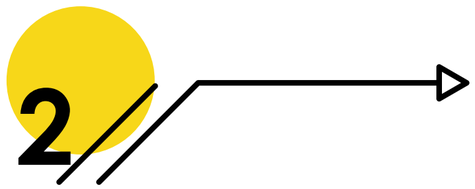
But there are measures known to minimize risk and maximize predictive accuracy for said tests.



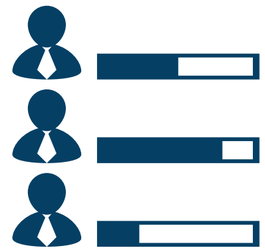
Understanding the Law:

HR generalists, specialists or organizational influencers are often advised to maintain legal compliance with the addition of psychometric tests to organizational processes. Anti-discrimination laws require - especially cognitive ability tests as referenced in the ePaper: *Knowing Your Psychometrics* - to remain job-relevant and strongly validated.

A recent example could be traced to the National Football League, an organization that changed its assessment battery due to concerns around racial discrimination and poor job-performance prediction. [6] Tests are generally required to respect privacy and not endeavour to diagnose candidates.



Understanding the Business Needs:



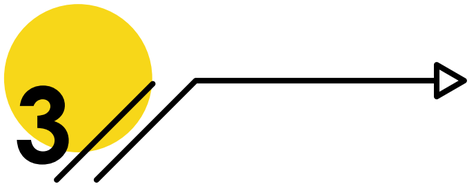
Organizations are known to focus a lot more on the “independent variables” or predictors over what’s being predicted - the “dependent variables”. If quantitative measures for employee job-performance doesn’t exist, there is little to no basis for statistical correlations for how accurately psychometric tests predict performance. Consider the following:

1. Purpose: A qualitative test is measured based on validity, and it is essential to ensure that the test being used measures what it is intended to measure. At the same time, an organization must understand the purpose for which they require an assessment before making any selections.

2. Job Roles: Psychometric tests are often a combination of different assessments; these combinations are best determined based on job roles. For example, content writers would require an assessment that measures for verbal comprehension, while hard labour would mandate a physical fitness test - both cognitive tests.

3. Industry: : Understanding industries form an important part of your assessment battery. If you look at sales, even within the same job role, skills and functionality vary depending on product and buyer sophistication. [7] A salesperson selling pens undeniably require a different set of skills from one that sells IT services.

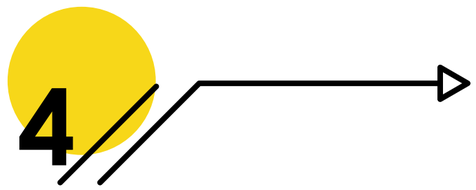
4. Geography: A test developed in India using the Indian population as a standard is remarkably more accurate than one that uses an American norm group. For example, it's more effective - in context - to use cricket analogies in India against baseball analogies, a sport most Indians are unfamiliar with. Likewise, an American audience scarcely tests well off an Indian standard.



Built to Withstand Malpractices:

Some candidates may be tempted into “gaming” results. It’s commonly referred to as “impression management”, a method used to come across as the more ideal candidate. [8] It’s recommended to compare references and ratings to test results to identify both consistency and correlation.

Some psychometric tests work with in-built measures to decipher if a candidate’s responses reflect impression management, or if they are incongruent with one another. But security measures aside, even a well-designed, legally defensible, and predictive test battery is likely to fail in adding value should a candidate find the test intrusive or time-consuming.



Assessing the Assessments

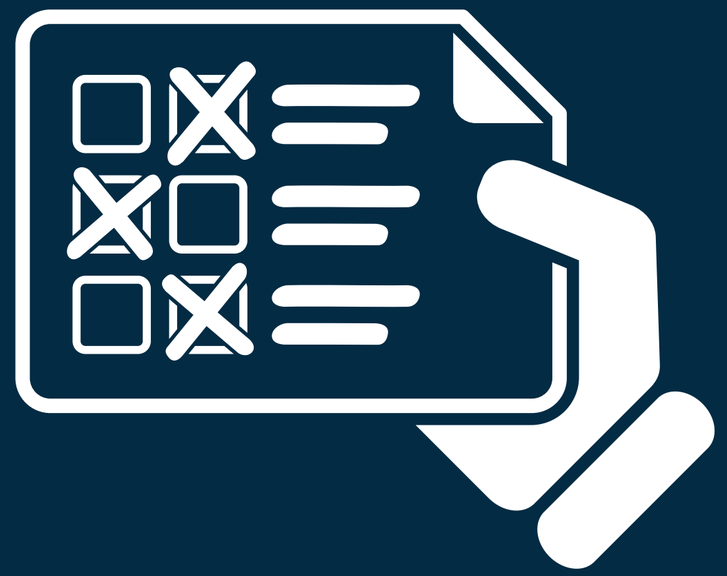
High-Performance organizations are in constant requirement of change and improvement, improving candidate evaluation systems - for example - via utilizing predictor, outcome variables, and the correlation between them. [9] Psychometric tests should also be subject to similar validation and intensive testing as the candidates they are being utilized to assess. Parameters for validity, reliability and norming weigh into this scenario.

It's assumed that when organizationally relevant professionals utilize appropriate methodologies to either retain, develop or select the right psychometric tests, they stand a chance to significantly improve the probability of selecting, developing and retaining the right talent also.[10] This holds true all the more when considering outside consultation or third-party assessment technology firms.

This raises the question of certain other parameters for quality, and these are parameters to check for pre-investment. Much of this is brought into scrutiny during the test making process itself, which is why we'll momentarily dive into the scenes behind test creation.

[3] How Do We Create or Standardize a Psychometric Test?

In definition, a standardized test is administered and scored in a consistent, or “standard”, manner. They are, in fact, designed in a way that stabilizes questions, conditions for administering, scoring procedures, and interpretations as consistent.



Standardized testing could be composed of true-false, multiple-choice, authentic assessments or essays. It’s possible to shape any form of assessment into standardized tests. When it comes to the creation of psychometric assessments, questions are measured in scales.[11] And these too are often most valid with standardization post-creation.

At the same time, in terms of creation, psychometric tests are subject to scrutiny via validity and reliability tests. On the other hand, norming frames the standardization process.

[3.1] Nature of Reliability in Psychometric Tests

A test is reliable as long as it produces similar results over time, repeated administration or under similar circumstances.

For example, it's reasonable to expect a line that measures five centimetres on one scale to measure the same on a different one. The line is essentially the same, and only a good scale can ensure it remains the same five centimetres regardless of what or who measures it. When compared to psychometric assessments, a reliable test is like that scale, with the ability to produce stable results over time.

Over the years, scholars and researchers uncovered multiple ways to check for reliability.[12] Some include testing the same participants at different points of time or presenting the participants different versions of the same test to see how consistent the results are.

Suffice it to say that an assessment has to show demonstrably good reliability in order to qualify for validity.

[3.2] Understanding Validity in Psychometric Tests

It is understandable to expect a test used in organizations to shed light on how a candidate would perform in a particular job. With this in mind, it is essential to reiterate the difference between reliability and validity, with the former being a prerequisite to the latter. [13]

Let's consider a dart player. In repeated trials, he or she continues to miss the mark consistently by about two inches. Of course, this implies reliable aim. Each shot hits the board in a region two inches from the target. It's difficult to not question his validity as a professional - considering he or she doesn't hit the bullseye as is the aim of all professional dart players - in comparison to his or her peers.

Reliability and validity go hand in hand, but reliability by no means indicates validity. As our example showed, having the first without the second hints at great consistency, but also inaccurate consistency.

There are tests for validity.

[3.3] Why Norming is Essential to Psychometric Tests

Even with a test that is both reliable and valid, there exists a question about results. An assessment fails without quantifiable results, but as often stated – human beings are far from quantifiable.

It is hard to quantify in a vacuum, competencies such as ethical integrity or teamwork; similarly, seeing a score on a personality test may be meaningless without a guide to interpret it. Experts distinguish a group of candidates from the other by comparing them to a standard – either among themselves, a relative standard, or an external criterion, an absolute standard.

The first way is to compare people against a population of interest, and this is what's more commonly referred to as *norming*. Another way is to have a solid standard against which you measure your assessment, using that standard to make decisions. Either way, it is required of a test developer to define a *cut-off* score for hiring or any other decisions dependent on the assessment.

However, even that is a delicate ball game. If you think about it, picking the relatively best apple from a batch of rotten apples would still yield a rotten apple. How then would you ensure good results from a good test? In psychometric tests, to assess overall performance, researchers have employed standardization samples, which simply refer to a large sample of test takers who represent the population for whom the test is intended.

A representative sample means using a group of children when developing a test for children, and an adult group when developing a test for adults. Also, based on the population, samples are generally made representative based on demographic factors like age, gender, education, religion, etc.[14]

When you get a 94th percentile on a trait like extraversion, you know that you are simply more extraverted than 94% of the sample group from whom the test makers derived the normal distribution. On the other hand, if you scored 94% on a math test, it simply implies that you marked about 94 in every 100 questions correctly.

Psychological constructs such as personality have no right or wrong answers associated with them, and can thereby not be marked using percentages.[15] This is why academics and researchers alike resort to norming among other methods to make sense of scores on personality assessments.

With growing concerns over costs, conveniences and other logistical challenges, technology-enabled assessments have become popular over time as well. Simply because they serve to streamline the process, reduce costs, increase efficiencies, allow employers to assess, and analyse more data points than previously deemed possible.

[4] Conclusion

Derivative stats from the Society for Industrial & Organizational Psychology (SIOP) is better used to conclude the points mentioned in previous chapters.

In it, the body highlights an active 68% of organizations to engage in various forms of job skill testing. This transcends to about 29% when it comes to one or more forms of psychological measurement and around 20% for cognitive ability tests.[16]

These tests in combination form the essentials of a psychometric test, something when used correctly enhances the chances of organizational success. Having said that, there are several applications to psychometric tests in an organizational context. We will look at those in the literature to come, but for now – we conclude here.

Do remember that despite the benefits there's only as good as its implementation. And there's another metric to look at, of course. **Quality.**

References

1 https://www.businessinsider.in/The-Best-Jobs-For-Every-Personality-Type/articleshow/41720265.cms

2 https://www.cnbc.com/2017/06/06/the-3-best-jobs-for-you-based-on-your-myers-brigg-personality-type.html

3 Stromberg & Caswell, Vox, 8 Oct 2015

4 Annie Murphy Paul, The Cult of Personality Testing, 2004

5 Carl Jung, Psychological Types, 1921

6 http://www.nytimes.com/2013/02/22/sports/football/nfl-introduces-new-way-to-test-a-players-mental-agility.html

7 http://www.law.byu.edu/news/file/Christensen_DeRosia_Lee.pdf

8 https://www.psychologytoday.com/blog/cutting-edge-leadership/201310/the-dangerous-art-impression-management

9 http://www.statisticshowto.com/criterion-variable-2/

10 https://hbr.org/2013/09/how-to-use-psychometric-testin

11 https://www.questionpro.com/blog/4-measurement-scales-every-researcher-should-remember/

12 American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: Authors.

13 Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). Educational Measurement (2nd Ed.). Washington, D. C.: American Council on Education

14 P243, Keith Coaley, An Introduction to Psychology

15 http://www.psychometric-success.com/personality-tests/personality-tests-understanding.htm

16 http://www.siop.org/workplace/employment testing/usingoftests.aspx

mettl

What is Mettl?

Mettl is a Saas based assessment platform that enables organizations to create customized assessments for use across the entire employee lifecycle, beginning with pre-hiring screening and candidate skills assessment, training and development programs for employees/students, certification exams, contests and more.

TRY FOR FREE

TRUSTED BY

amazon.com SIEMENS PUMA Optimizely Lufthansa

+91-9555114444

contact@mettl.com

Plot 85, Sector 44, Gurgaon,
Haryana, India - 122003