

as seen in:



The Problem with Using Personality Tests for Hiring

Whitney Martin

August 27, 2014

A decade ago, researchers discovered something that should have opened eyes and raised red flags in the business world.

Sara Rynes, Amy Colbert, and Kenneth Brown conducted a study in 2002 to determine whether the beliefs of HR professionals were consistent with established research findings on the effectiveness of various HR practices. They surveyed 1,000 Society for Human Resource Management (SHRM) members — HR Managers, Directors, and VPs — with an average of 14 years' experience.

The results? The area of greatest disconnect was in staffing — one of the lynchpins of HR. This was particularly prevalent in the area of hiring assessments, where more than 50% of respondents were unfamiliar with prevailing research findings.

WHAT HR MANAGERS GET WRONG ABOUT HIRING RESEARCH

The most common items answered incorrectly as part of a 2002 survey of SHRM members.

PERCENTAGE ANSWERING INCORRECTLY OR INDICATED THEY WERE UNCERTAIN
(THE CORRECT ANSWER IS 'FALSE')

84%	Companies that screen for values have better performance than those that screen for intelligence
82	Conscientiousness is a better predictor of job performance than intelligence
69	Integrity tests have high degrees of adverse impact
68	Integrity tests are not very effective in practice because so many people lie on them
58	Being very intelligent is actually a disadvantage for performing well on a low-skilled job
58	Being intelligent is a disadvantage in low-skilled jobs
51	There are four basic personality dimensions, like in MBTI

SOURCE SARA RYNES, AMY COLBERT, AND KENNETH BROWN

HBR.ORG

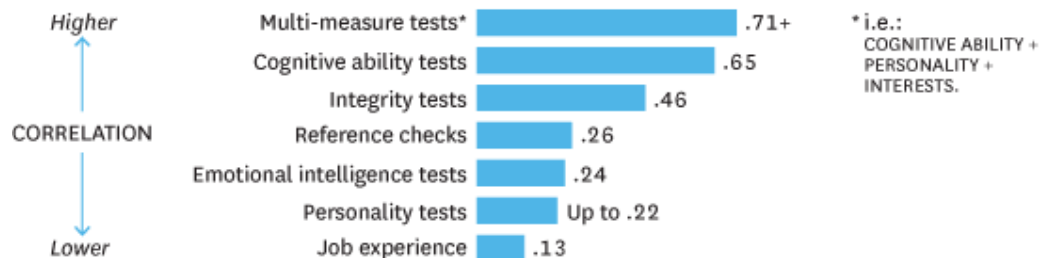
Several studies since have explored why these research findings have seemingly failed to transfer to HR practitioners. Among the causes are the fact that HR professionals often don't have time to read the latest research; the research itself is often present with technically complex language and data; and that the prospect of introducing an entirely new screening measure is daunting from multiple angles.

At the same time, anyone who has ever been responsible for hiring, much less managing, employees knows that there is a wide variation in worker performance levels across jobs. Therefore, it is critical for organizations to understand what differences among individuals systematically affect job performance so that the candidates with the greatest probability of success can be hired.

Extensive research has been done on the ability of various hiring methods and measures to actually predict job performance. A seminal work in this area is Frank Schmidt's meta-analysis of a century's worth of workplace productivity data, first published in 1998 and recently updated. The table below shows the predictive validity of some commonly used selection practices, sorted from most effective to least effective, according to his latest analysis that was shared at the Personnel Testing Counsel Metropolitan Washington chapter meeting this past November:

THE MOST EFFECTIVE HIRING SELECTION PRACTICES

And those that don't work so well, based on validity coefficients ranging from 0 to 1. The higher the number, the higher the correlation between test scores and predicted job performance.



SOURCE BASED ON DATA SHARED BY FRANK L SCHMIDT IN A NOV 6, 2013 ADDRESS TO PTCMW AS AN UPDATE TO: SCHMIDT, F. L. & HUNTER, J. E. (1998).

HBR.ORG

So if your hiring process relies primarily on interviews, reference checks, and personality tests, you are choosing to use a process that is significantly less effective than it could be if more effective measures were incorporated.

And yet that's how many companies operate. According to a 2011 NBC News article, the use of personality assessments are on the rise, growing as much as 20% annually. Especially problematic is the widespread use of Four Quadrant (4-Q) personality tests for hiring, something I see regularly in my consulting work.

A 4-Q assessment is one where the results classify you as some combination of four different options labeled as letters, numbers, colors, animals, etc. They originated around 450 BC when Empedocles noticed that he could group people's behavior into four categories which he labeled earth, water, fire, and air. Hippocrates made the same observation, but (coming from a medical background) labeled the categories blood, phlegm, black bile, and yellow bile. Since then, hundreds of iterations of these tools have been developed, all essentially based on the same premise and theory.

Generally speaking, 4-Q tools consist of a list of adjectives from which respondents select words that are most/least like them, and are designed to measure "style," or tendencies and preferences. While they can seem highly insightful — not to mention being widely available and inexpensive — they have some severe shortcomings when used in high stakes applications such as hiring.

For one, they tend to be highly transparent, enabling a test taker to manipulate the results in a way that they feel will be viewed favorably by the administrator. Also, since they are designed to measure "states" (as opposed to more stable "traits"), there is a significant chance that the results will change over time as the individual's context changes (most publishers of 4-Q tests recommend that individuals re-take them at fairly frequent intervals for this reason).

This begs the question: How can an individual's assessment results be used to predict future job performance if there is a reasonable chance that their scores will change over time?

When using any assessment, managers need to step back and ask themselves one basic question before giving it to a potential employee: Is this test predictive of future job performance? In the case of 4-Qs, probably not. They can provide tremendous value for self-discovery, team building, coaching, enhancing communication, and numerous other developmental applications. But due to limited predictive validity, low test-retest reliability, lack of norming and an internal consistency (lie detector) measure, etc., they are not ideal for use in hiring.

The strongest personality assessments to use in a hiring context are ones that possess these attributes:

- Measure stable traits that will not tend to change once the candidate has been on the job for some length of time.
- Are normative in nature, which allows you to compare one candidate's scores against another's to determine which individual possess more (or less) of a particular trait.
- Have a "candidness" (or "distortion" or "lie detector") scale so you understand how likely it is that the results accurately portray the test-taker.
- Have high reliability (including test-retest reliability) and have been shown to be valid predictors of job performance.

Even when using a tool that meets the criteria outlined above, personality constructs are not the most predictive measure available. Personality tests are most effective when combined with other measures with higher predictive validity, such as integrity or cognitive ability.

Using well-validated, highly predictive assessment tools can give business owners and managers a significant leg up when trying to select candidates who will become top producers for the organization. However, all assessment approaches are not created equal. And some will not offer a significant return on your investment. Accordingly to a 2014 Aberdeen study [registration required], only 14% of organizations have data to prove the positive business impact of their assessment strategy. Knowing which types of assessments will be most effective in accomplishing the specific objectives you have identified for your organization will enable you to select a tool with a measurable impact on the bottom line.

As a measurement strategist at ProActive Consulting, Whitney's passion and expertise lies in the field of surveys and assessments. A self-professed "data nerd," Whitney has a Master's Degree in the area of Human Resources Measurement and Evaluation and has conducted extensive research on the predictive validity of various hiring assessment strategies.