

Psychological Tests: Sense or Nonsense?¹

Patrick Merlevede, jobEQ's leading researcher

As long as the number of potential job applicants is greater than the number of job openings, some form of selection will be required. Moreover, using a poorly designed or improper test in the selection process may keep the organization from hiring the best candidates. The aim of this article is to ensure that the assessment instruments used during the selection process help to decrease the probability of this error and increase the overall effectiveness of the assessment process.

Most test developers and test users have good intentions. Yet, often due to inadequate training, there is considerable misuse of test data. The aims of this article are (a) to warn about the misuse of tests in general and (b) about "nonsense" claims that some commercially-oriented people sometimes make about the tests they put on the market. Most of the points in this article apply to any test; thus, anyone publishing a test should have reasonable responses to these points. Moreover, their explanations should include what they do to address these specific issues.

Test Design Problems

Test design problems can be classified into two groups. The first group has to do with the theoretical basis that underlies the test as well as with the *construction* of the test itself. A test without a solid theoretical foundation cannot deliver good results. When a test has a solid theoretical foundation, the *psychometric analysis* of the test will confirm these foundations and will point out the problem areas in the use of this test.

Let us go a bit deeper into these two issues by asking questions and establishing some criteria:

1. **Test Constructs** - What theory underlies the test? How are the test questions formulated? How do these questions relate to the theory? Which questions were rejected during test development? Are these questions context independent, or are they an exact match to the context for which the test wants to predict performance? How do language and cultural factors influence the test performance, even within one language area or one geographical area? What other environmental factors play a role in the test and how are these issues addressed?
2. **Psychometric Quality** - A task force of the American Psychological Association examined the integrity of tests and found out that they often could not get access to research-based information, not even to basic properties such as test reliability (e.g. internal consistency) and test validity (correlations with various criteria), let alone test-retest reliability, factor analysis, or other forms of internal structure validation. Test designers should be able to provide information on these critical criteria.

Test Use Problems

There are numerous ways to misuse test data. In their book, Eyde et al. (1993) indicate 86 specific elements and seven factors that represent common problems in test use. This is a summary of the problems mentioned:

1. **Comprehensive Assessment** - In part, this issue is about test follow-up. "How do you assure that the test score is an accurate description? With what other instruments is this test combined that deliver extra predictor information?" One must also consider the emotional state that the person being tested was in when the test is filled out, as well as the test's research evidence and test limitations.

¹ This article can be found on jobEQ's web site at: http://www.jobeq.com/articles/judge_a_psyc_test.php

2. Proper Test Use - What training or certification did the person or persons using the test have? How is the quality of test use controlled?
3. Psychometric Knowledge - Are correct basic statistical principles respected (concerning standard error of measurement, reliability, and validity)? Are these principles applied for interpreting the test results? How does the test limit the number of false positives and false negatives?
4. Accuracy of Scoring - Is the test filled out and tabulated by a person or by a computer? If it is scored by an individual, are there quality checks for errors? This can make a major difference in the number of errors produced in the scoring and recording processes.

Input Process	Tabulation Process	Potential Scoring Problems
Test taken via pen-and-paper	Results tabulated by a person	Possible data entry error and tabulation error
Test taken via pen-and-paper	Results tabulated by a computer	Possible data entry error
Test taken online	Results tabulated by a computer	Virtually no possible scoring problems

5. Appropriate Use of Norms - When using a standardized test, one can run into problems regarding practical applications. Is this test measuring the important traits or behaviors? How do you know objectively if “proactivity” or “teamwork” is important for success in a particular role or position? For selection, a good test will only use criteria important to job performance, and test designers must be able to prove that the information generated by the test directly relates to the role or job under consideration.
6. Feedback to test-takers - Test designers must provide correct interpretations and be staffed to do so.

Problems with Test-Takers

The seventh factor in the list of problems pertains to test-takers themselves. Here are two of the ways in which the test-taker contributes to the difficulties to the use of testing.

Lack of Self-Knowledge - Test-takers might not know enough about themselves to respond to the questions accurately. Overestimating or underestimating one’s own abilities is a common challenge.

Falsification - Psychological tests are often quite transparent. Many observers assume that job applicants would not willingly report undesirable behaviors that could ruin their chances for employment (see, for example, Goldberg et al. 1991). You should always carefully examine a test before you use it for your organization. Are there right and wrong answers? Will an educated candidate be able to know what you want to hear? For example, some tests ask:

1. At work, are you:
 - a. Always late?
 - b. Sometimes late?
 - c. As punctual as the next person?
 - d. Rarely late?

Respondents who are rarely late and those who are late but think that revealing such would reduce their chances of getting a job might both answer "d." Test designers should be able to prove to you that they can eliminate the potential for falsification.

Falsification is a major problem because it increases the number of "false positives" (when someone scores well on the test, but does not perform as expected or predicted) in test results.

Conclusion

This article identifies some critical issues regarding responsible and accurate test use. There are good tests in the marketplace and they have documented responses to the issues raised. The answers to the questions raised and the evidence related to criteria of viable tests have to come from test developers. If, as the consumer, you cannot get those answers right away, at least get a commitment that these answers will be available by the time you complete a pilot project.

We do not, under any circumstance, recommend that you proceed with an important application of a test when sufficient data about its development, reliability, and validity are not available.