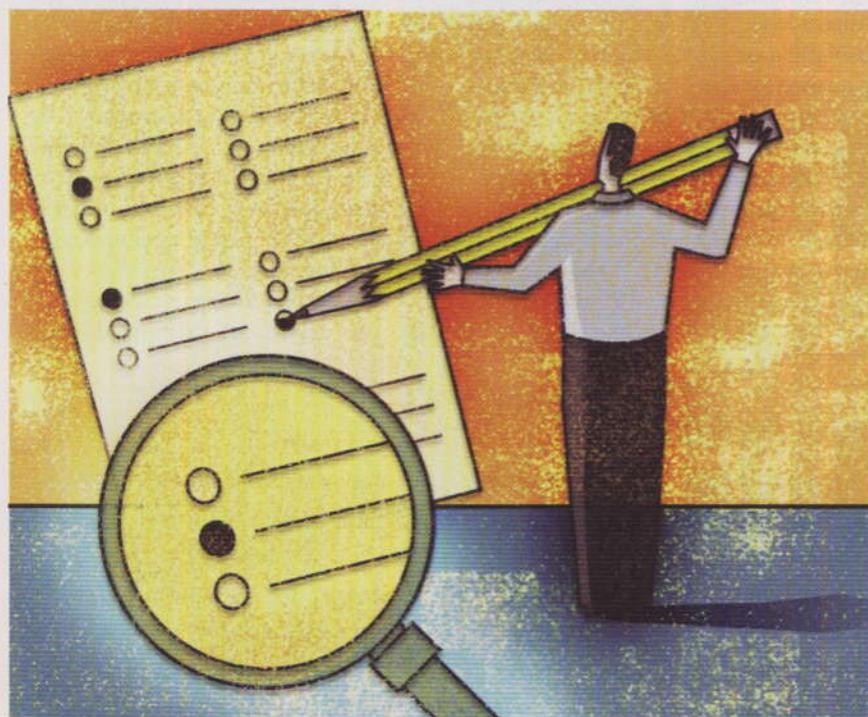


# Avoiding Bias in Pre-Employment Testing

Research into testing bias raises questions about the effectiveness of tools designed to measure such bias.



By Theresa Minton-Eversole

**D**ebate about whether to use psychological assessments during the hiring process has gone on for years. Advocates say that as long as such assessments have been validated and proven to be without bias, such testing is acceptable. Others disagree, saying that this form of testing should not be used, as it could lead to discrimination, albeit inadvertent, based on race or national origin, thereby putting a company at risk for claims of adverse impact.

New research from Indiana University's Kelley School of Business has sparked more heated debate by suggesting that the tools used to check for bias in tests of "general mental ability" could themselves be flawed, thus raising further questions about whether employers should rely on these cognitive exams to make objective hiring decisions. Such tests measure verbal, numerical or reasoning abilities or reading comprehension, for example.

"I'm a believer in tests," says study co-author Herman Aguinis, professor of organizational behavior and human resources at Kelley and director of the university's Institute for Global Organizational Effectiveness in Bloomington, Ind. "Pre-employment tests add tremendous value to the hiring process. The irony is that for 40 years we have been trying to assess potential test bias with a biased procedure."

## The Validity of Validity Tests

The study, published in the July *Journal of Applied Psychology*, investigated an amalgam of scores representing a vast sample of commonly used exams—such as university entrance exams and civil service exams—and pre-employment tests. To look for bias, scientists typically examine samples of real test scores organized by demographic groups and compare the scores to some measure of job performance, explains Aguinis. A test is assumed to be bias-free when the prediction of performance based on test scores remains similar across demographic groups.

Aguinis and co-authors Steven A. Culpepper at the University of Colorado-Denver and Charles A. Pierce at the University of Memphis in Tennessee challenged that assumption by creating a large-scale computer simulation to assess the accuracy of the tools used today to identify test bias. The researchers created bogus test and job-performance scores that indicated clear differences in responses among different demographic groups. However, the data analyses failed to reflect the bias that researchers had built into the tests.

"We created billions of biased test scores and we tested random samples" of the scores, says Aguinis. "But the results showed there was still no test bias, though we were certain there was. As a result, we determined that the tools used to identify test biases are themselves flawed, and we proved that bias can be present but not be detected, which could result in inaccurate prediction of outcomes such as job and academic performance." >

*The author is an online editor/manager for SHRM.*

### Testing: One Piece of Puzzle

Some industrial psychologists are underwhelmed by Aguinis' research findings, however. They note that test bias is a fact of life that doesn't have to be proven, just dealt with effectively. Some say such hypersensitivity around the use of psychological assessments during the hiring phase distracts HR professionals from the bigger issue of how best to create a sound selection system that encompasses multiple data points on which to base a hiring decision.

The federal government's Uniform Guidelines on Employee Selection Procedures notes that cognitive tests are most likely to result in adverse impact, says Larry O'Leary, an industrial psychologist and testing consultant based in St. Louis. These guidelines provide a framework for determining the proper use of tests and other selection procedures.

Aguinis' "research shows that the tools we have to detect bias in hiring assessments

may not be sufficiently sensitive," says Elaine D. Pulakos, chief operating officer for the industrial/organizational psychol-

ogy consulting firm PDRI, a subsidiary of Atlanta-based PreVisor, a pre-employment testing and assessment vendor. "But there is no need for alarm based on this research, and it certainly does not mean that the assessments organizations are using today for hiring are biased. All the research means is that we would benefit from additional research into the development of new tools that may better detect bias."

**'We would benefit from additional research into the development of new tools that may better detect bias.'**

ogy consulting firm PDRI, a subsidiary of Atlanta-based PreVisor, a pre-employment testing and assessment vendor. "But there is no need for alarm based on this research, and it certainly does not mean that the assessments organizations are using today for hiring are biased. All the research means is that we would benefit from additional research into the development of new tools that may better detect bias."

"Testing should be used within a system of multiple assessment methods,"

are very good at initially screening people in or out of jobs on basic job requirements. Once you have a pool of candidates that meet some minimum competency requirements, further assessments can be very beneficial in selecting the best talent [to fit the job] and the organization. Testing is very economic, especially in cases where a large number of applicants need to be evaluated."

But, as a rule, test results should be the last piece of data HR and hiring managers

### Factors to Consider in Test Creation

Employers develop their own tests for cost effectiveness, test-security concerns, company culture, position uniqueness and other reasons. Test development is a complex and time-consuming process, so experts either inside or outside the organization should be involved to ensure the company has developed a legally defensible test and proper supporting documentation, regardless of what it plans to measure with the test, the type of test or mode of administration.

Whether tests are created in-house or procured, the Society for Industrial and Organizational Psychology Inc. of Bowling Green, Ohio, recommends that test documentation includes:

**Test development information.** The research and development that went into creating the instrument.

**Validity.** Information about products' validity, or the accuracy of the inferences made based on test results. There are many forms of validity evidence.

**Test bias.** Evidence that the test does not

contain bias on the basis of race or sex; that is, the test is related to outcomes in a similar manner for all individuals.

This does not mean that the test will have similar results for different groups of people, but that it is not a biased indicator of an outcome of interest.

**Reliability data.** Information on the reliability, or consistency of results, of testing products.

**Information on administration.** A description of all materials required for administration and administration instructions, and discussion of issues such as standardization of testing conditions.

**Data for test interpretation.** Information about the different norm groups that are available for the test. Ideally, one uses a norm group similar to the group of people in the position being recruited for. Information should be made available on appropriate test score interpretation.

**Scoring options.** Explicitly stated qualifications for scoring. Some tests can be scored

on-site, either by hand or by machine. Others require that an employer call, mail or fax test results to a test publisher for scoring.

**Ongoing test research and refinement.** Information about when the test was developed and when the test was last updated to comply with new legal requirements or to reflect changes in vocabulary.

**Time requirements.** Information on how time limits were determined and why they are necessary.

**Credentials and experience.** The educational background and work experience of the people who developed the test. Some tests also require administrators or individuals interpreting test scores to have certain credentials.

**Cost.** Direct costs of the test, which usually includes test booklets, answer sheets and a test administrator's manual. Computer-based testing generally includes software valid for a prepaid number of uses.

—Theresa Minton-Eversole

review in drawing conclusions about a person's abilities, Kozloff adds.

O'Leary agrees: "Since cognitive tests do predict success, but do have a substantial degree of adverse impact, I advise HR practitioners to avoid using them by themselves as a predictor in a selection process."

"Good assessment instruments are the most effective and efficient tools human resource professionals have available to make hiring decisions," says Pulakos, who wrote a guide, *Selection Assessment Methods*, one of a series of Effective Practice Guidelines published by the Society for Human Resource Management Foundation. "And good assessment instruments have been repeatedly shown to yield important bottom-line outcomes for organizations, such as increased productivity, reduced turnover, and enhanced employee engagement and manager satisfaction, among others."

Kozloff notes that in terms of selection, all U.S. organizations—government or private—must address two issues that are spelled out in the federal guidelines: adverse impact and job-relatedness.

Adverse impact exists if the proportion of protected group members selected for a job is less than 80 percent of the proportion of majority group members selected. If an assessment method is shown to produce adverse impact and the organization chooses to continue using that assessment, then an employer must show job-relatedness through a professionally sound validation study. The strongest validation evidence is obtained by demonstrating that people who score higher on the test actually perform better on the job, says Pulakos.

Kozloff says hiring decisions also can be improved with triangulation, or the process of using all of the information available to hiring professionals from all the methods of data collection, and not just test scores or interview data. Specifically, he advises employers to:

- Look for overlapping data from two to three informational sources.
- Wait for all available data before making a decision.
- Use recruiter interviews as threshold screening.

Kozloff also cautions employers not to:

- Eliminate candidates after a single interview.
- Influence interviews by discussing results before all data are in.
- Use a single numeric score to make a conclusion about a person's capability or lack of capability.

Likewise, in the case of multiple interviews, Interviewer A does not talk

"If I were an HR professional responsible for buying testing, I would require vendors to show me how the means by which they assess test bias in their instruments can actually detect bias," he says.

To do this, he adds, HR professionals need to become shrewd, educated consumers of testing and assessment instruments. "The huge cost and performance ramifications associated with poor em-

## Managers need to be trained to use tests so data support decisions rather than define them.

to Interviewer B until Interviewer B has conducted an interview, to avoid biasing Interviewer B.

Finally, he says, managers need to be trained to use tests in decision-making so they can ensure test data support decisions rather than define them.

### Testing Demands

Employers have two choices when implementing an employment test. They can either purchase a test or create their own. Most test publishers provide a technical summary or manual that describes the qualities and characteristics for any given test. The technical manual should provide information on most, if not all, of the factors to consider before purchasing a test. This can lead to information overload for HR managers, so obtaining professional help from an industrial psychologist who is expert at interpreting tests is often necessary, Kozloff says.

Aguinis says the testing industry needs to develop more sensitive tools to better assess test bias, but only demanding market conditions will speed this development along.

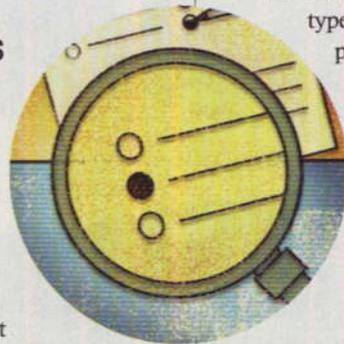
ployee selection make this a critical competency [for] all HR professionals."

This may be an unrealistic expectation, sources say.

"Unfortunately, many HR professionals have misconceptions about both the value of formal assessments and the types of assessments that have proven to be most effective," says Pulakos. "This, coupled with the fact that the area of selection testing is inherently technical and difficult to understand, has led to an underutilization of formal assessments in organizations.

"With everything that HR generalists have on their plates, it's unlikely that they're going to be able to devote the time necessary to becoming experts in psychological assessments," she says. However, she adds, they may find that they want to consult with an independent industrial/organizational psychologist who possesses the necessary expertise.

Regarding the ramifications of Aguinis' research on the testing industry, "There is no reason to overreact and set expectations that can't be met," Pulakos says. "More research is needed to determine if better tools can be developed for detecting bias, and, until that happens, companies should continue using their hiring assessments and not abandon or even reconsider use of the most effective hiring tools we have available today." ■



### Online Resources

For links to the *Selection Assessment Methods* guide and more resources and information on pre-employment testing, see the online version of this article at [www.shrm.org/hrmagazine/1210EmploymentStaffingAgenda](http://www.shrm.org/hrmagazine/1210EmploymentStaffingAgenda).