Integrity Testing for Personnel Selection: The Role of Research Methods

Paul R. Sackett
University of Minnesota

► Please tell us about your current position and research interests.
I am a professor of psychology at the University of Minnesota, where my research focuses on the use of psychological testing for employment decisions and for academic admission decisions.

► What got you interested in studying the quality of tests of employee honesty and integrity?
My interest in the topic of integrity testing arose out of an inquiry from an employer wondering whether integrity tests were useful. There was little research on the topic at the time, and I began studying the topic.

► What has been the real-world impact of this work?
Integrity testing is widely used, having grown from a new and little-understood form of testing to a relatively mature field with an extensive body of supporting research.

Employers often face a complex problem when trying to choose among many applicants for a limited number of jobs. Under such circumstances, some employers turn to psychologists for help, asking whether psychological tests are available that will help them identify those applicants most likely to be effective employees. Psychologists typically respond with a wide range of possible employee-selection tests, including measures of ability, personality, interests, values, knowledge, integrity, and skill.

The goal of most testing in employment settings is the identification of applicants most likely to perform well on particular jobs. However, in some settings, employers have a different goal: They want to avoid hiring employees who might engage in various forms of counterproductive work behavior (for example, theft). In settings where employees have access to money or merchandise and work with limited monitoring or supervision, employers are highly motivated to avoid hiring anyone likely to steal. Several decades ago, employers with such concerns commonly administered polygraph examinations in an attempt to learn about applicants' possible past history of wrongdoing.

As the polygraph came under scrutiny and was eventually made illegal for most employers, paper-and-pencil tests of "honesty" or "integrity" emerged as alternatives. One test type is commonly called an overt, or clear-purpose test. This type of test focuses on respondents' beliefs about theft: how often it occurs and how widely; how it should be punished; how easy theft is; what respondents' thoughts are about theft and whether they agree or disagree with common rationalizations for theft; and how they assess their own honesty. A
second type of test is commonly called a personality-oriented, or disguised-purpose test. This type of test is closely linked to normal-range personality assessment measures, is not explicitly aimed at theft, and includes items dealing with dependability, conscientiousness, social conformity, thrill seeking, trouble with authority, and hostility.

The market for these tests is huge: Several million applicants are tested each year. Many test vendors have entered the market; at one point, they were offering over 40 tests. Mainstream psychological test publishers developed a few of these tests; many others had little to no research backing. Given the extensive interest in and use of the tests, the obvious question is whether they work: Do the tests, in fact, predict behavior on the job?

My involvement in this area began with such a question. In 1977, I received a phone call from an organization asking whether a particular test was valid. Knowing nothing about the test in question, I set out to research the topic. A colleague and I gathered up all the research studies we could find and published an article summarizing what was known about integrity tests (Sackett & Decker, 1979). Little did I know that 30 years later I’d still be interested in the topic of employee honesty and integrity. Over the years, I’ve published four more reviews of the state of our knowledge about integrity tests (Sackett & Harris, 1984; Sackett, Burris, & Callahan, 1989; Sackett & Waneck, 1996; Berry, Sackett, & Wiemann, 2007). In 1979, I could locate only a handful of research studies; by the time I wrote the later reviews, I was able to examine and critique hundreds of studies.

Research Strategies for Evaluating the Effectiveness of Integrity Tests

To gain an understanding of integrity tests and how well they work, researchers have employed a wide variety of research methods. Among the most widely used are criterion-related validity studies in which tests are administered to a sample of job applicants or current employees and measures of on-the-job behavior (or whatever criterion you are attempting to predict) are obtained. Various statistical measures are used to index the degree to which test scores are related to job behavior. In other words, data are examined to determine whether individuals with high test scores do better on the job than those with low test scores.

Early studies of integrity testing used scores on a polygraph examination as the criterion. The question in this research was whether individuals with low integrity test scores were more likely than those with high scores to fail a polygraph exam, and research found this to be the case. However, critics were skeptical of these findings, questioning the accuracy of the polygraph method. If the polygraph is not a credible measure, showing that integrity test scores correlate with polygraph scores is not persuasive.

In response to this criticism, researchers conducted studies that did not use the polygraph as the criterion. Instead, test-takers underwent questioning directed at getting them to admit to prior theft from employers. Researchers using this strategy found that test scores are correlated with admissions of theft (that is, those with high integrity test scores admit less prior theft). Again, this evidence was met with skepticism. Among the complaints was the suggestion that test scores don’t really predict theft, but merely predict willingness to admit to theft. In other words, high and low test scorers may be just as likely to steal; they may merely differ in their willingness to admit to wrongdoing.

So, researchers tried yet another approach. Because self-reported theft was not seen as a credible measure, the focus shifted to measures of theft detected on the job. In these studies, firms would keep track of all employees caught stealing over the course of a year and then the test scores of employees who were caught stealing would be compared with those of employees who were not caught stealing. These studies showed that those caught stealing were very likely to have low test scores. At the same time, though, only a very small number of employees were caught in a given year, thus raising questions about how to interpret the test scores of those not caught. Skeptics could point out that many people with low integrity test scores were not caught stealing and, thus, assert that the test mislabeled innocent people. Most importantly from the employer’s point of view, those caught stealing may represent only a fraction of those actually stealing. And it’s even possible that those caught stealing are not a random sampling of those who steal, but instead represent those who simply aren’t very good at it. So, if we use theft detection as the criterion measure, we’re left with the interesting finding that those caught stealing do tend to have low test scores, a finding consistent with what we’d expect if the tests worked as intended, but also with uncertainty about the issue of undetected theft.

Researchers continued exploring alternative research strategies. The idea of an objective measure of wrongdoing was attractive, but difficulties in detecting theft led to the use of indirect measures. In one creative study, researchers looked at the behavior of a sample of Salvation Army bell-ringers at Christmas time (Jones & Terris, 1981). The researchers gave an integrity test to all bell-ringers at the start of the holiday season, but they did not score the test or use it in the hiring decisions. They then measured the average amount of money collected each day for each bell-ringer. They had historical data for each location, and so each bell-ringer’s dollar intake could be compared with the historical average for that location.

At the end of the holiday season, those with low integrity test scores turned in less money than those with high integrity test scores, a finding that they interpreted as evidence that dishonest bell-ringers were pocketing some of the money collected. However, low-integrity bell-ringers could have been taking longer breaks, working fewer hours, and thus soliciting less money. Of course, taking extended breaks can be seen as a form of counterproductive behavior, and thus the integrity test is effective in identifying those prone to taking in less money, regardless of whether it is due to extended breaks or to theft. Nonetheless, we are left unsure of exactly why those with low integrity test scores turn in less money.

These methodological difficulties led to yet another approach, namely to move from on-the-job field settings to research carried out in the psychologist’s laboratory. The appeal of laboratory research is the degree of control available to the researcher. Here researchers can design settings where
opportunities for forms of wrongdoing or misbehavior are arranged and where settings are designed in such a way that the researcher knows with certainty how each participant behaves. In one study, participants took an integrity test and then engaged in a number of experimental tasks for which they were to be paid a certain amount (for example, $5). When they were finished, they were told, "Here's your $5," and handed an envelope containing two $5 bills. Not unexpectedly, those with high integrity test scores were more likely than those with low scores to report the extra money. Findings from these kinds of studies are consistent—for example, those with low integrity test scores are more likely to cheat on an experimental task or to give themselves more points than deserved when self-scoring a test they've taken (cf. Berry, Sackett, & Wiemann, 2007).

I find this approach an interesting inversion of the typical research strategy as advocated in textbooks. The common sequence is for research to begin in a laboratory and later to move into field settings to examine whether lab findings generalize to real-world settings. In the area of integrity testing, the question of interest is inherently an applied one (Do these tests predict wrongdoing on the job?), and thus most research has been done in applied settings. The use of a lab setting, with its accompanying degree of control over the situation, proved a useful supplement in response to questions about possible flaws in the measures obtained in field settings.

Lessons Learned

This story illustrates several important issues regarding psychological research methods. One is that in settings where research outcomes have significant real-life consequences (for example, the investigation of a type of test affecting the job opportunities of millions of people), research will be scrutinized very closely. Critics will examine the work carefully to determine whether possible alternative explanations for the results can be identified. This kind of analysis can be frustrating to the individual researcher whose work is being criticized, but it is clearly good for the advancement of knowledge: Criticism prompts additional research to address the concerns raised by critics.

Second, evidence from a single study is rarely seen as fully persuasive. As developments in the investigation of integrity testing illustrate, triangulation across multiple research methods is crucial. Each method has strengths and weaknesses, and convergence of findings across methods plays a large role in the eventual acceptance of research findings. Note that the research on integrity testing has found that integrity tests are predictive of (a) polygraph exam performance, (b) self-reports of theft, (c) detected theft in organizational settings, (d) objective measures of various forms of wrongdoing in organizational settings, and (e) objective measures of various forms of wrongdoing in carefully controlled laboratory settings. This consistency in findings is quite convincing. It would have been quite a different matter had the results been different (for example, had the tests proven to predict polygraph exams but nothing else). A body of work showing that a research finding is not specific to one research setting or to one way of measuring the variables of interest is far more persuasive.

Other Research Issues in Integrity Testing

When psychologists assess the quality integrity tests, a wide range of other important research questions arise: Do integrity scores vary by features such as race or gender? Can applicants raise their scores by giving distorted responses (that is, by faking)? Are the test scores reliable (that is, would a test-taker get the same or nearly the same score if he or she retook the test after a short interval)? For what range of jobs are the tests effective (that is, is predictive power limited to entry-level jobs, or are these tests also useful for higher-level jobs)? How do integrity test scores relate to other psychological measures (that is, are they correlated with measures of central personality traits)? The answer to the last question is yes; they correlate with three measures: conscientiousness, agreeableness, and emotional stability.

Important practical issues also influence employers' decisions as to whether or not to use an integrity test. One key notion is that the opportunity to be selective is a scarce resource. Consider an employer with, say, two applicants for each opening. The employer could choose to use an integrity test and screen out the lowest 50 percent of the applicant pool. But the employer may also want employees who learn quickly and solve problems effectively, and so another option would be to use a cognitive ability test and screen out the lowest 50 percent of the applicant pool on that measure. "Spending" your opportunity to be selective on screening for integrity prevents screening on cognitive ability, and vice versa. Alternately, the employer could use both tests but be less selective on each (that is, screen out the bottom 25 percent on both cognitive ability and integrity). A further research question is whether it is more useful to screen for one characteristic than for another.

Conclusion

In sum, psychologists have made great progress in understanding how well integrity tests perform as predictors of counterproductive work behavior. Over several decades, our knowledge base has moved from a handful of studies to hundreds. And, as is typical with research evidence, no single grand study answers all questions; rather, each smaller study serves as a piece in a larger puzzle. Researchers identify potential weaknesses in existing studies and work to design additional studies to remedy these weaknesses. For this to occur, they must not only be highly knowledgeable about their specialty area but also well versed in research methodology. Thus informed, they can effectively evaluate the strengths and weaknesses of existing research studies and design and conduct studies that further understanding of issues that affect the lives of all of us.

Suggested Further Reading
