Are
Workplace
Tests
Worth Taking?
Only if you do them right—which you probably don’t.

By James Krohe Jr.

What can psychological screening of job applicants do for you? According to many of the firms that peddle tests, everything from straighten out your slice off the tee to raise your kid’s SAT scores to reconcile you to your first spouse. Ask potential new hires to scratch up a few pieces of paper, and the workforce gods will grant you (as one typical come-on promises) “extraordinary increases in productivity while reducing employee relations problems, employee turnover, stress, tension, conflict and overall human resources expenses.”

No surprise, then, that 30 to 40 percent of large firms use one or more forms of psychological testing in the hiring process, from entry-level to corner-office. But the casual observer might wonder whether the results match the hype, since to judge from newspaper headlines—or a month or two inside a typical company—all this ritual filling in of circles with No. 2 pencils hasn’t appeased the workforce gods one bit. The large corporation remains, it would seem, a nest of boneheads, thieves, time cheats, and bellyachers.

So it’s worth asking: Are companies getting their money’s worth by psychological testing of workers? More to the point, is there any real evidence that years of testing have produced more honesty among employees, identified more leaders, or eliminated more crazies—or is testing merely another management hallucination brought on by breathing copier toner fumes?

Some quick answers. While testing is better than not testing, there are many more bad tests than good ones; picking the good ones requires care, and no single test will tell you all you need to know. As a result, few com-

JAMES KROHE JR. is a Chicago-area writer and editor. His most recent article was “Why Is Customer Service So Bad?”, the Jan/Feb cover story.
panies do it right, and thus few realize more than a fraction of testing's potential value.

Are you doing it right? Probably not.

Psychological Testing Works—True or False?

It depends on what you mean by works. Psychological tests come in hundreds of versions of dozens of types, including personality profiles, integrity tests, and interest tests. All purport to divine the future job performance of job applicants by measuring, variously, work styles, personal values, vocational interests, emotional stability, and what might be called regular-guy-ness. A few—very few—promise to identify the Jack Welch buried inside the unassuming body of an operations manager.

Such a test, like any pre-employment assessment instrument, can be said to be scientifically valid if it identifies what it promises to identify—if, for example, the applicant whom the test tags as warm, bold, open to change, sensitive, or self-reliant is in fact warm, bold, open to change, sensitive, and self-reliant. But HR managers care less about whether a test is deemed valid than whether it produces useful results. Utility is what matters. A valid test can be said to be genuinely useful only if using it keeps bad apples out of the corporate barrel or steers nature's go-getters into sales, its born geeks into IT, its genetic dweebs into accounting. Do tests work as promised? Some work much better than others.

What Works, More or Less: Integrity Tests

The star of the psychological-test team is the integrity test. Sometimes known as an honesty test, this instrument became the most popular form of psychological testing in the 1980s, after a crime wave engulfed businesses. Losses from internal bank fraud and embezzlement more than tripled in that decade, and losses to retailers from crime—most of it committed by employees or people in cahoots with them—reached as high as 2 percent of sales. It was in the hope of heading the varmints off at the pass that some companies resorted en masse to pre-employment polygraphs. These wonder machines, sadly, proved to be about as accurate as Ouija boards and were outlawed for private use in 1989. The now-familiar pen-and-paper integrity test was offered as a more reliable substitute.

“I have always said the first step to reducing internal theft starts at the point-of-hire,” says Mark R. Doyle, president of security consultancy Jack L. Hayes International. “That is, do not hire the bad apple. We always encourage companies to spend their money up-front to weed out undesirable candidates. It is always more effective and desirable not to hire at-risk candidates than it is to deal with them after the fact.”

Keeping out at-risk candidates in the first place is, of course, what integrity testing is all about. One company, a major North American specialty retailer, went beyond the survey-givers' claims and actually put the tests to the test. The retailer—which insists on anonymity, wanting to divulge neither the extent of its problems nor the secret of how it solved them—was suffering the woes common in that industry: too many staffers walking out the front door, and too much inventory walking out the back door. The company began using an integrity test to screen new hires in approximately 600 of its 1,900 stores. After a year, management compared the testing stores to its non-testing stores. In the testing stores, inventory loss shrank more than 35 percent, while in the non-testing stores it rose more than 10 percent. An added bonus: Testing stores saw employee turnover drop 13 percent, while in non-testing stores turnover increased 14 percent.

This specialty retailer isn't alone. Convenience-store, home-improvement, and supermarket chains have reported similarly happy results. Such findings have been widely
touted by publishers of integrity tests as product endorsements. In-house studies can be self-interested or poorly done, however, and a firm plagued by thieves usually has more than one loss-control initiative under way, which means that it’s misleading to attribute all improvements to the overall security program’s testing component. Also, an integrity test can “work” even if it has no predictive power at all. The Society for Industrial and Organizational Psychology (SIOP) lists among the advantages of honesty testing the fact that it “sends[es] the message to test takers that integrity is an important corporate value.” That message will itself scare off a few rogues.

Integrity testing has its critics, but their ire is usually reserved for poorly written versions of the tests, which can tag a too-honest applicant as dishonest. Judged solely on business utility, however, enough evidence exists confirming the utility of integrity tests to justify a cautionous endorsement. SIOP concurs, concluding that integrity tests can reduce business costs by identifying individuals who are less likely to be absent or engage in other “counterproductive”—corporate-ese for “money-losing”—behavior.

What Doesn’t Really Work: Everything Else

Do you know your MBTI type? If so, you’re in good company: Each year, more than 2 million Americans—many of whom work in Fortune 500 companies—undergo Myers-Briggs Type Indicator assessments, according to publisher Consulting Psychologists Press Inc. Corporate America takes the test awfully seriously, considering its origins: Essayist Katharine Cook Briggs came up with the idea in a moment of curiosity about the personality of her daughter’s new boyfriend, and she and that daughter, mystery novelist Isabel Briggs Myers, developed the MBTI in 1942 based on Jungian theories of personality types.

It is not familiarity with the widely used test that breeds contempt among its many critics, but the fact that it is—to borrow a non-scientific term—hooey. The sixteen distinctive personality types Briggs and Myers described were culled from biographies and similar sources and have no real scientific basis—confirmed by the fact that as many as three-quarters of test-takers have been found to have a different personality type when taking the same test more than once.

The Myers-Briggs, says personality-test-

Corporate America takes the Myers-Briggs test awfully seriously, considering its origins.

better. There are hundreds on the market, and vetting them requires skills somewhere between a computer programmer and a soothsayer—which is why most experts advise seeking a consultant’s help in picking the one for the job. Finding a scientifically sound test is difficult, and finding a scientifically sound test that is right for a particular need is doubly so.

Indeed, evidence of the utility of psych testing other than integrity testing, such as personality tests and interest inventories, is less persuasive, where it exists at all. For example, interest inventories—questionnaires that ask about respondents’ interests and spit out appropriate career tracks—have very low validity for predicting an applicant’s future job performance and thus, presumably, his value to the company.

Even if a test or tests can accurately assess the potential for creativity, collegiality, and other workplace virtues, the collective impact of these virtues on a firm’s actual performance must be inferred rather than proven. It’s easy to calculate how much inventory is leaving a building unpaid-for, but very hard to calculate in dollars the effect on output of improved creativity or collegiality. This doesn’t mean that an assessment of such traits returns no value to the firm, mind you—only that this value is hard to measure. (See “Where’s the Data?,” page 19.) And that admission can lead to questions too uncomfortable to answer with any degree of confidence.

What’s Worst of All: Not Testing

So apart from entry-level integrity questionnaires, why bother with testing at all? Because just about any type of psychological screening is better than the usual alternatives. Think about what those alternatives are—that is, the haphazard way that most companies go about hiring...
people and setting them on particular career paths.

David W. Arnold, a twenty-year veteran of the testing industry as a senior VP of R&D and now corporate counsel to test publisher Wonderlic Inc., notes that too many employers look for what he calls a gastro-thermic response—"that nice warm feeling in the gut." Wareham likens the process of hiring to falling in love with someone. "If he looks like you and says the things you want him to say, you make what I call a 'romantic rush to judgment.' You're as blind as Paul McCartney looking at his second wife."

Other employers judge the fitness of a job applicant instead by the firmness of her handshake or the steadiness of his putting stroke. Easy to sneer at, but reference checks, conventional unstructured interviews, and even college degrees aren't much better tests of fitness—each is a surprisingly poor indicator of job performance. Compared to such conventional measures, using even a psychological test with low predictive power increases the accuracy of prediction, particularly when the applicant pool is large and the number to be selected is relatively small.

So testing is, at least potentially, better than not testing. The problem is—and on this the experts are practically unanimous—that the potential value of these tools (whatever that value is) is not being realized by more than a fraction of the firms that use them and, of course, not at all by the firms that don't use them. Getting the most out of tests is no simpler than getting the most out of a new IT system—developing a useful assessment tool involves many steps and safeguards, and failure to attend to the many niceties explains why so many of the firms that do use psychological tests to screen job applicants are unlikely to get full value for their money.

The pros in the field all offer a litany of failures: Companies use tests that are poorly written or are wrong for the purpose; they administer the tests ineptly or carelessly; they use tests as a cheaper substitute for, rather than a refinement of, other selection mechanisms. Any one of these faults can reduce testing's value to a firm; a testing program flawed by more than a couple of them would be better off choosing new hires by reading the leaves left in the bottom of the tea cups in the cafeteria.

Picking the Wrong Screwdriver

A good test applied to a task for which it was not developed becomes a bad test, says Wendell Williams, managing director of Scientific-Selection.com, a Georgia purveyor of organizational hiring systems. Some of the most widely used tests have little to do with work at all; they were developed for use in other settings. Perhaps the best example is the Minnesota Multiphasic Personality Inventory, a 567-item questionnaire concocted at a Gopher State mental hospital in the 1930s. The MMPI has been validated by decades of research—as a tool to help psychologists to diagnose personality disorders and mental disturbance. The workplace is not the home, certainly, but—the rumbling in the cubicles to the contrary—it is not all that much like a mental hospital either. "HR types often do not understand the limitations of the test," Williams insists. "Even its publisher says it is not for use for hiring."

Wonderlic's Arnold is a little less dismissive of the MMPI, although he agrees that it is widely misused as a hiring assessment. "That test is not appropriate in such situations," he concedes, although he holds that the MMPI is appropriate when used in safety-sensitive jobs. "For one am happy the crew of the United Air Lines plane I'm on has been given the MMPI and sat down with a psychologist," he says with a laugh.

Of course, administering even the best available test can be a waste of money if it does a great job of accurately measuring traits that are irrelevant to the job being filled. Fitting test to traits, however, is a test in itself. "You just don't go in and say, "These are the ideal personality traits," explains John W. Jones, who was for years a senior VP of research and development for Reid London House (today's Pearson Performance Solutions) and now is president of test publisher IPAT. Jones adds that to learn which personality traits are best suited to a specific organizational culture and strategy—and thus which kind of test to use to measure them—a company must perform some personality- oriented job analysis first. Not doing so is like picking a screwdriver out of the toolbox before you know what kind of screw you are about to tighten. (Yes, analyzing each important position in terms of personality can be time-consuming and expensive, but the process can be worth it; Procter & Gamble, for instance, believes so.)

Justin Menkes, an industrial/organizational psychologist and consultant whose 2005 book Executive Intelligence: What All Great Leaders
Where's the Data?

Both test-makers and test-givers have made valiant efforts to overcome skeptics' objections to psychological testing, using a variety of methods. “Everyone knows that they need to show whether a test works,” says personality-testing expert Ann Marie Ryan, “but not everyone knows how to do it.” It should be added that not everyone has the means to do it. Ryan cites Procter & Gamble among those with both means and expertise. The company “does a tremendous amount of testing, and they invest a lot of resources finding supporting evidence for it.”

The best way to determine whether a test works for your company is to test it: Use it, then check the results against the eventual performance of the people measured by it. Measuring is not necessarily believing, however, even at larger firms. The people in the company who pick which tests to use are usually the ones who judge their efficacy, and HR departments have an interest in confirming their own sagacity. “They’ve committed a lot of the CEO’s money, and they don’t want to be wrong,” says Wendell Williams, managing director of ScientificSelection.com. HR departments may report that testing is working regardless of evidence pro or con, since to report otherwise would reveal themselves as dupes of the testing racket or worse. Inconvenient results can be ignored or unreported.

Psychologist and consultant Justin Menkes concurs. “I did a fairly large-scale evaluation of a senior-executive development program that consisted of year-long coaching,” he recalls. “The HR people took the success stories and talked them up. It was true that 25 percent did better—but 50 percent saw no change, and 25 percent did worse!”

Surprisingly few studies have been published on the links between the use of selection practices and organizational-level measures of performance such as annual profits or employee turnover. A few firms have conducted commission of studies of the return on their HR practices, but these rarely get published, lest the companies give away some competitive advantage. Academics have looked at psychological-screening instruments in depth, but their curiosity tends to fasten on questions of formal validity that do not much interest HR managers and their bosses.

J.K.

Have caused a stir, sees problems with even those tests, such as personality style inventories, that were developed in workplace settings. “They were validated in environments that are dramatically different from that of the job application process,” he explains. Participants in the test-development process are typically promised confidentiality and told that answers will be used for research purposes only. Ask the same questions of someone applying for a job, Menkes insists, and you’re unlikely to get the same level of frankness in her response—and thus not the same level of predictive validity.

Using inappropriate tests not only wastes the company’s money—it can be illegally discriminatory to sort out job applicants using tests not provably related to job performance. Clinical personality tests such as the MMPI, for example, have been found to be tantamount to medical examinations, and screening for medical conditions violates the Americans With Disabilities Act. Arnold cites one firm—Rent-A-Center—that required that applicants for promotions take the MMPI. The practice led to a class-action lawsuit and a 2006 appeals-court ruling that the company violated the ADA because the test “likely had the effect of excluding employees with [mental] disorders from promotions.”

Good Test, Bad Test

There is the good test used appropriately, and then there is the bad test. “There are tests from good companies that are validated by evidence using strict professional guidelines,” explains Ann Marie Ryan, professor of psychology at Michigan State University and a past president of the Society for Industrial and Organizational Psychology, “but there also are publishers who fudge or who don’t report all their data.”

“Lots of these tests are a lot of hot air,” Menkes agrees. “The error rate is very high.”

Unfortunately, HR departments don’t typically select and administer tests based on hard evidence. “Critical evaluation of these instruments is not HR managers’ best skill,” Menkes says. “It shouldn’t be as mysterious as it is. There is so much research, so many independent academicians looking at what works and what doesn’t. You don’t want to depend on people who are selling things.” In short, when it comes to such instruments, he insists, business in general is “uninformed even at senior levels.”

The result—in Williams’ opinion, anyway—is dire. He estimates that less than 10 percent of corporate America’s psychological tests actually measure job fit, actually predict performance, and were designed to be used for hiring purposes.

Too often, then, a test is chosen not because it is more accurate a predictor than competing products but because it costs less, or is easier to use, or because the legal department likes it, or because the head of HR saw a compelling conference presentation—or because the company has always used it. Usually in business, Williams insists, it is the “feel” or “flavor” of a test, not its formal validity, that determines whether companies elect to use it for pre-hiring screening. Or its popularity. Decisions about screening mechanisms are often made according to the “Sixty Million Frenchmen Can’t Be Wrong” principle.

Such practices may earn a department manager a pat on the head from the executive suite, but not from the ivory tower. In a 2004 journal article, Ryan and colleague Nancy T. Tippens warned corporate cost-cutters that “it never makes sense to employ a tool without sup-
The University of Iowa's Frank Schmidt and Michigan State University's John Hunter came to the same conclusion in a 1998 study: "By using selection methods with low validity," they warned, "an organization can lose millions of dollars in reduced production."

Nearly as dangerous as picking a bad test is to pick a good test and then rely on it too much. Even the best tests measure only some of what a person needs to do well on the job, and they don't do that perfectly. That is why, whatever their disagreements about specific tests, experts in publishing, consulting, and academia agree that the best pre-employment screening approach is the one that is most comprehensive. Schmidt and Hunter found that when a firm adds written pre-employment tests to a battery of non-test screening mechanisms such as references, in-depth interviews, and job simulations, the successful prediction rate rises to something like 75 percent. That's as close to a sure thing as you can get in the hiring game.

Hayes International's Mark Doyle says that a thorough pre-employment screening process should consist of not only a pre-employment "honesty test" but a weighted application form (if applicable), multiple interviews, reference checks, a criminal background check, and a drug screen.

Wonderlic's Dave Arnold points out the obvious problem with such conscientiousness: The more comprehensively employers screen new hires, he notes, "the more it costs." Excellence is expensive in hiring as in everything else, and there are plenty of quick-and-dirty paper tests out there that are cheap to buy and can be administered—if badly—by an intern or secretary. Many are from top-tier test publishers, and they may not look all that much different than their expensive, labor-intensive competitors. The off-the-rack tests save money today; tomorrow's costs—in absenteeism, theft, lost productivity, turnover—are hypothetical and won't show up in a budget review anyway.

There is, in short, a difference between cost-effective and cheap. "There are many companies that try to get by doing as little as they can, maybe doing from one to three of the six screening steps we advise instead of all of them," Doyle says. "More are now doing criminal background checks and/or drug tests than supporting validity evidence—regardless of how cheap, how low the adverse impact, or how easy to administer." The University of Iowa's Frank Schmidt and Michigan State University's John Hunter came to the same conclusion in a 1998 study: "By using selection methods with low validity," they warned, "an organization can lose millions of dollars in reduced production."
in the past, and some figure if they do these, maybe they don’t need an honesty test.”

Pencils Down, Please

The world's best test is a waste of paper if it is administered improperly, or not used in conjunction with other assessments, or if its results are ignored. Each is a common fault in testing programs as they are seen by consultants and academics.

A test not administered is hardly less useful than a test not administered properly. Bad things happen if, for example, test-givers fail to follow the instructions on the package. Many tests are supposed to be completed within a set time; if that time limit is not observed scrupulously for each group, Williams explains, “people end up taking what amounts to separate tests.” More companies are offering tests online, but such administrative efficiencies come at a price. “If tests are done on the Web,” he warns, “you don’t know who’s taking the test.”

The Association of Test Publishers stresses proper administration of psychological tests. In 1996, the group issued model guidelines for the conduct of pre-employment integrity testing, for example, but the ATP is a trade association, not a professional society, and its injunctions to its members to behave well are “shoulds” rather than “musts.” Even the firms that try to live up to them run into opposition from skinflint HR departments. “We try to train the administrators in the organizations that use the tests,” says Wonderlic’s Arnold, “but you can only do so much.”

And of course, the payrolls of even companies that do testing right—from selection to administration—are hardly free of slackers and bozos. The best a company can hope for is not a perfect workforce but one that is slightly less imperfect than it would be without testing. Tests are like any tool—useful if well-matched to the job and used with skill. But nothing about the human psyche—even that of a prospective data-entry clerk—is simple, including assessing it. Psychological screening instruments may be marketed as cheap and quick ways to simplify a chore that would give Budda ulcers, but at best they make it possible to do the complex job of staffing more precisely. Sadly for HR managers who still believe in Santa Claus, the test hasn’t been invented that can make picking new hires easy. “The first rule in hiring is, caveat emptor,” says John Wareham. “People forget that. Smart people—and they are in the minority—always take an independent look.”

ACROSS THE BOARD JULY/AUGUST 2006 23