5

CRITICISMS OF EMPLOYMENT TESTING: A COMMENTARY

ERNEST H. O'BOYLE JR. AND MICHAEL A. McDANIEL

When people think of employment testing, they typically consider only written tests of cognitive ability and personality. However, applicants are screened for jobs with a variety of measures, including reviews of resumes and applications, cognitive ability tests, personality assessments, integrity testing, employment interviews, drug testing, and reference checks. The U.S. federal guidelines on employee selection (Equal Employment Opportunity Commission [EEOC], Civil Service Commission, Department of Labor, & Department of Justice, 1978) consider any method used to screen job applicants to be an employment test. Our use of the word *test* in this chapter is consistent with these federal guidelines. Employment testing is one example of high-stakes testing (Sackett, Schmitt, Ellingson, & Kabin, 2001), in which those who perform well on the test are considered for employment and those who perform less well are not considered for employment.

The purpose of this chapter is to review commonly offered criticisms of employment testing. We then evaluate these critiques in light of research findings. In selecting the criticisms to address, we relied on our experience with employment testing and solicited common criticisms from others who have substantial experience with employment testing. We identified 11 such

criticisms. Although one can classify criticisms of employment testing in several ways, for this chapter we organized the criticisms into three categories: fairness, accuracy, and administrative efficiency.

Concerning fairness, employment tests are said to be

- unfair to ethnic and racial minorities,
- illegal when ethnic and minority groups obtain lower scores on average,
- unfair because they have not been validated for every type of job and across every type of context, and
- an invasion of privacy.

Concerning accuracy, employment tests are said to

- result in some bad hiring decisions,
- tell one nothing one would not learn by talking with the applicant,
- be useful but often assess the wrong content, and
- be easily faked.

Concerning administrative efficiency, the employment tests are said to

- be unnecessary because one can train anyone to do any job,
- take too long, and
- be too expensive.

We address each of these criticisms in turn.

CRITICISMS RELATED TO FAIRNESS

Employment Testing Is Unfair to Ethnic and Racial Minorities

Many criticisms of employment testing are based on concerns regarding the lower performance, on average, of Blacks and Hispanics on employment tests relative to the performance of Whites and Asians. Some criticisms are based on the assumption that there are no job-related ability differences among the various racial and ethnic groups, and so if employment tests show such differences, something must be wrong with the tests. The critics are correct in their observation that Blacks and Hispanics on average obtain substantially lower scores than Whites and Asians on most employment tests, particularly on tests of cognitive ability (Murray, 2005; Sackett et al., 2001). Such differences are also common when educational credentials (degrees, grade point average) are used as a test or when job-related knowledge or skills are assessed. Roth, BeVier, Bobko, Switzer, and Tyler (2001) documented that after correcting for range restriction, there was a difference of roughly 1 standard deviation between the means of Blacks

and Whites in general cognitive ability and a substantial but somewhat smaller difference in mean scores between Whites and Hispanics. Because many employment tests are designed to assess cognitive ability, or are at least moderately correlated with cognitive ability, it is more common than not for employment tests to show substantial mean differences between Blacks and Whites and between Hispanics and Whites. These differences are not restricted to employment tests. Large mean ethnic–racial differences are found in educational assessments as well.

Research on changes in the magnitude of mean ethnic–racial differences in cognitive ability is controversial, with some arguing the differences are narrowing in recent decades (Dickens & Flynn, 2006), whereas the bulk of the research indicates that these mean differences have been relatively intractable (Murray, 2005; Rushton & Jensen, 2006). Even if one accepts the position that mean racial differences are shrinking over time, they are still large enough to cause substantial disparities in hiring decisions.

The large mean ethnic–racial differences in employment test scores prompted much research, occurring primarily in the 1970s, examining whether employment tests are biased against Blacks and Hispanics. It was argued that employment tests might be biased through either differential validity or differential prediction (Boehm, 1972; Bray & Moses, 1972; Kirkpatrick, Ewen, Barrett, & Katzell, 1968).

Differential validity studies addressed the hypothesis that a test may have validity for Whites but not for Blacks or that the test may have a larger validity for one group than another. By *validity*, we refer to the extent to which an employment test correlates with a measure of job performance (e.g., a supervisor rating of one performance, the amount of product sold, the number of customers served). Thus, a hypothesis about differential validity might hold that a test predicts performance of Whites but not Blacks. In the 1980s, after substantial research examined this issue, it was shown that differential validity is uncommon (Schmidt, 1988; Schmidt & Hunter, 1981; Wigdor & Garner, 1982).

It became apparent that it would be more appropriate to look for differential prediction than differential validity because even if a test has the same validity for all ethnic–racial subgroups, the best regression lines to predict job performance might be different.[1] Differential prediction can occur if the regression lines for ethnic–racial subgroups yield either different slopes or

[1]A correlation is useful for estimating the relationship between an employment test and a measure of job performance. However, it does not give the one the information needed to predict a job performance score from an employment test score. For this, one needs to run a regression analysis. A regression equation for the prediction of job performance is of this form: job performance = $a + b$ (employment test score). This regression formula describes a straight line that summarizes the relationship between the employment test and the job performance measure. The a value is called the *intercept*. When the variables have been standardized, the b value (the *regression weight*) equals the correlation coefficient and describes the slope of the regression line.

different intercepts. The research indicated that different slopes for different racial–ethnic subgroups occur at no higher than chance levels (Bartlett, Bobko, Mosier, & Hannan, 1978).

Different intercepts are more common, but the error in prediction tends to favor minority groups. Specifically, when the prediction of job performance for Black or Hispanic groups and Whites is based on a common regression line, performance of Blacks and Hispanics is overpredicted on average (Hartigan & Wigdor, 1989; Schmidt, Pearlman, & Hunter, 1980). Thus, to the extent that differential prediction occurs in employment tests, it is to the disadvantage of White applicants but not Black or Hispanic applicants.

Employment Testing Is Illegal When Ethnic and Minority Groups Score More Poorly on Average Than Whites

Some critics of employment testing assert that under U.S. federal law, it is illegal to use employment tests when ethnic–racial groups score lower than Whites, on average. This criticism is not factual. The *Uniform Guidelines on Employee Selection Procedures* (EEOC et al., 1978) are the executive agency guidelines that are given deference by U.S. courts. The *Uniform Guidelines* do not address ethnic–racial mean score differences but do address adverse impact. *Adverse impact* as defined by the EEOC is a "selection rate for any race, sex, or ethnic group which is less than four-fifths (or eighty percent) of the rate for the group with the highest rate" (EEOC et al., 1978, p. 60-3.4 D). If an organization's screening instruments show adverse impact and someone challenges the employment test, the employer must demonstrate its job relatedness.

Unfortunately, the *Uniform Guidelines* are more than 30 years old and inconsistent with current scientific findings and professional guidelines (McDaniel, 2007). Specifically, they are biased in favor of local validation studies, differential validity and differential prediction studies, and a preference for costly and detailed job analysis data. As such, the procedures advocated by the *Uniform Guidelines* are often inappropriate and needlessly cumbersome. Thus, critics who argue that employment testing is illegal when mean ethnic–racial differences exist are incorrect. However, federal guidelines place an expensive burden on employers when adverse impact is present.

McDaniel (2007) raised the question, "Why have the *Uniform Guidelines* not been revised to be consistent with professional standards?" (p. 168). He speculated that the

> primary use of the *Uniform Guidelines* is to pressure employers into using suboptimal selection methods in order to hire minorities and Whites at approximately the same rates. If employers do not hire minorities at about the same rates as Whites, the *Uniform Guidelines* are invoked by enforce-

ment agencies and plaintiffs to require the employer to prepare substantial validity documentation. (pp. 168–169)

It is unfortunate that the *Uniform Guidelines*, which are inconsistent with professional principles (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Society of Industrial and Organizational Psychology, 2003) and scientific findings, are used to promote race-conscious hiring in violation of the U.S. Civil Rights Act of 1991, the very legislation that they are intended to enforce.

Employment Tests Are Unfair Because They Have Not Been Validated for Every Type of Job and Across Every Type of Context

This criticism is based on the belief that each job in a field is unique in terms of duties, environment, and personnel needed. Therefore, any attempt to generalize the validity findings of an employment test from one job to another or one location to another is flawed. This argument is known as the *situational specificity hypothesis* (Schmidt & Hunter, 1977, 1998). This hypothesis was developed in response to the frequent observation that an employment test used in one setting did not yield the same or nearly the same validity coefficient when used in another similar setting for the same or similar job. This caused some to speculate that there were yet-to-be-discovered characteristics of situations or jobs that caused the differing validity coefficients. It was the prevalence of the situational judgment theory in the mid-1970s that influenced the authors of the *Uniform Guidelines* to emphasize the need for detailed job analysis information and local validation studies. The seminal work of Schmidt and Hunter in the 1970s and 1980s showed that the situational specificity hypothesis was incorrect (Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980; Schmidt & Hunter, 1977; Schmidt, Hunter, & Pearlman, 1981). Specifically, they found that the validities varied from study to study primarily because of simple random sampling error. Thus there is no scientific basis to support the position that employment tests need to be validated for each situation in which they are used. There is, however, a pragmatic reason for employers to continue to validate employment tests. This reason is legal defensibility under EEOC guidelines. As discussed in this chapter, job-specific test validation is the federal government's preferred (although misguided) method of demonstrating test validity, and many employers would rather waste thousands of dollars on test validation than millions of dollars in litigation. Employers must weigh heavily the need for local validation and ensure that their decision is based on several contingencies that go beyond the thinking that "this test worked for someone else; therefore, it should work for me."

Employment Tests Are Invasions of Privacy

The content of some employment tests causes critics to argue that the tests are an invasion of privacy. Typically this criticism is directed toward clinically oriented personality tests designed to identify severe psychological disorders. Such clinically oriented personality tests are often used for police applicant screening and may contain questions that applicants may find intrusive (e.g., questions concerning one's private bodily functions). Other types of tests, known as *biodata* tests, ask questions about one's background and interests in areas that are not viewed as relevant to the job (e.g., recreational interests, experiences in high school). Criticisms related to invasions of privacy are primarily limited to items that have little explicit relation to the job. Employers should evaluate whether the information obtained through clinically oriented personality tests and biodata tests is sufficiently valuable as a predictor to justify their inclusion in an employment test.

CRITICISMS CONCERNING THE ACCURACY OF EMPLOYMENT TESTS

Employment Tests Result in Some Bad Hiring Decisions

Some critics argue that using employment tests sometimes results in bad hiring decisions. For example, an applicant who scored well on an employment test was an unproductive employee. Alternatively, a person who would have been an excellent employee was not hired because of a low test score. This criticism has some merit. Predictions from employment test scores are not perfect. Figure 5.1 shows an illustrative scattergram for a sample of data in which an employment test score correlates .50 with a measure of job performance. In the upper left section of the graph, an arrow identifies a person who scored relatively low (about 3) on the employment test but scored relatively high (about 7) on the job performance measure. Had the employer not hired anyone with scores as low 3, this productive employee would not have been hired. The arrow in the right section shows another employee who scored very high (almost 9) on the employment test but performed poorly on the job (about 3.5). Thus, even though an employer uses an employment test, some hires prove to be less than productive employees.

Employers can take several actions to minimize hiring errors. Abandoning the employment test and hiring people randomly will result in substantially greater hiring errors. Replacing one employment test with another is a useful solution if the new employment test has better validity (fewer hiring errors) than the original test. Another solution is to supplement the original test with additional information to reduce selection errors. For example, one could supplement a cognitive ability test with an interview in hopes of reducing hiring errors. These decisions are best made by evaluating

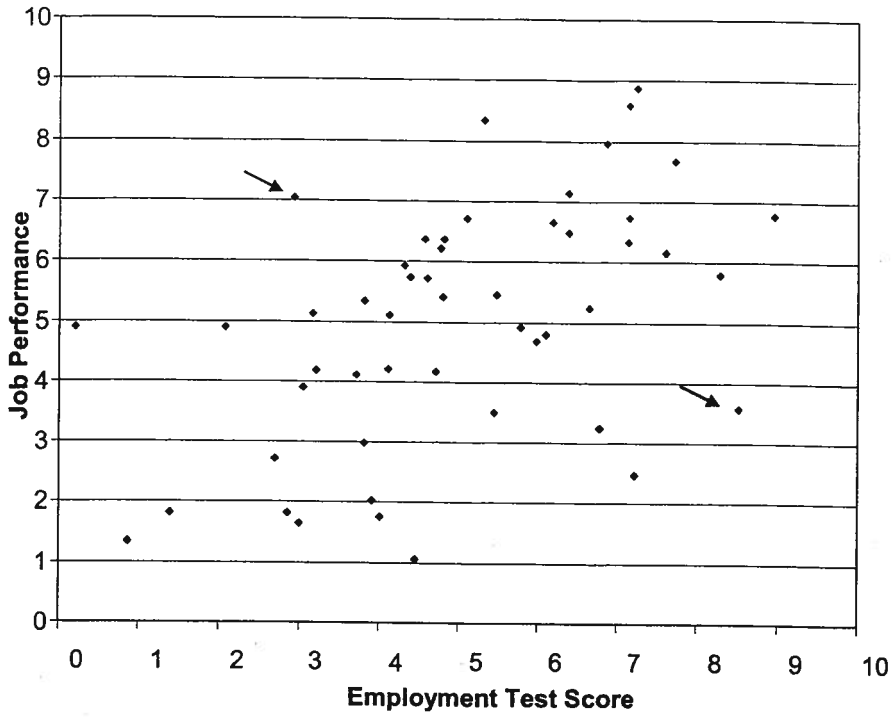

Figure 5.1. A scattergram of the relationship between an employment test and job performance (simulated data). In the upper left section, an arrow identifies a person who scored relatively low (about 3) on the employment test but scored relatively high (about 7) on the job performance measure. The arrow in the right section shows another employee who scored very high (almost 9) on the employment test but performed poorly on the job (about 3.5).

cumulative evidence concerning the effectiveness of various tests. Schmidt and Hunter (1998) summarized 80 years of employment test validity information to identify which employment tests work best and which should not be used. These cumulative research results should guide employers regarding which tests to use.

Some employers evaluate the validity of a test by trying it out on a few incumbents, for example, five employees. If the test scores match their perception of the employees, the test is judged a good measure, otherwise the test is judged a bad measure. Although intuitively appealing, this approach yields near random results. The five employees are a small sample of a population of employees, and small sample estimates of correlations have large amounts of sampling error.

By way of example, consider an employment test that correlates about .20 with a measure of job performance. If you collect data on 5 individuals, the correlation that you estimate based on the five individuals can range from –.61 to .80.[2] Even evaluating the test with 50 incumbents, the correla-

[2]The range of values in this example are the 80% confidence interval (10th percentile to 90th percentile) for a population correlation for the varying sample sizes listed in the example.

tion could range from .02 to .37. Only with large numbers of incumbents—say, 500—will the correlation likely be sufficiently free from sampling error that one can be confident of the precision of the validity coefficient (for sample size of 500, the correlations could range from .14 to .25). Therefore, relying on a limited sample of incumbent data does not tell one much about the validity of the test. To evaluate the validity of a test, one would be best served by relying on cumulative research evidence for the test.

Employment Tests Do Not Tell One Anything That Would Not Be Learned by Talking With the Applicant

We have often heard hiring managers argue that employment tests do not provide any information additional to what one would get by talking with applicants or by examining their resumes or educational transcripts. This belief is not specific to employment testing and is often referred to as the *actuarial versus clinical debate*. The belief that "gut feelings" or past experiences are more accurate predictors than diagnostic tests is not a new argument. This debate was most vehemently argued in the field of clinical psychology, and it took nearly a century and dozens of empirical studies (for a review, see Marchese, 1992) to overcome the belief that a "good" clinician does not need tests to make a diagnosis (Dawes, Faust, & Meehl, 1989).

Talking with applicants (e.g., an interview) is a method of screening applicants. These types of employment screenings are known in the literature as *unstructured employment interviews*—"unstructured" in the sense that the manager is not bound to a specific set of questions. To evaluate the suggestion that the unstructured interview outperforms other employment screening tools and methods, one needs to compare the validities of the various employment tests. The argument is now a proposition that can be empirically evaluated. Schmidt and Hunter (1998) tested this argument by means of a meta-analysis using more than 85 years of employment data. They examined how much incremental validity (added predictive capability) a variety of employment tests have after taking into account an intelligence measure. The incremental validity for unstructured employment interviews was .04—one of the lowest incremental predictors and only marginally better than a handwriting analysis (Schmidt & Hunter, 1998).

Those who offer this criticism likely lack evidence to support the assertion that their employment test is better than the employment test they criticize. In part, this criticism reflects a tendency among many to discount scientific findings when the findings are counter to their own beliefs.

Employment Tests Can Be Useful, but Many Assess the Wrong Content

Some critics argue that employment tests often do not measure the most useful content for employment decisions. This criticism contains two

arguments. The first is that employment tests do not measure content that is relevant for predicting job performance. The second is that there is other content that could be assessed in employment tests, but it is generally not assessed. These two arguments are addressed in turn.

The argument that employment tests do not measure job-relevant content is related to the face validity of the test. An employment test that is face valid contains content that is clearly relevant to the job and might even have an applicant perform a sample job activity. For example, a test of typing or word processing has clear face validity for jobs that demand fast and accurate typing. Other tests, such as measures of general cognitive ability, may have substantially less face validity. The value of an employment test is best evaluated on the cumulative evidence of its ability to predict job performance, and tests of general cognitive ability have been proved to be good predictors of job performance (Schmidt & Hunter, 1998). Thus, employment tests do not need to be face valid to be useful for evaluating job applicants. However, face-valid tests are more acceptable to applicants than tests that have no apparent relevance to the job.

The second argument holds that there is other content that could be assessed in employment tests but generally is not. The search for different content in employment testing is an active area of research in personnel selection (Frei & McDaniel, 1998; Hunter & Hunter, 1984; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Reilly, 1996, Rynes & Connerley, 1993; Taylor, Keelty, & McDonnell, 2002; Verive & McDaniel, 1996). Much of the research on employment tests focuses on efforts to reduce mean racial differences in test scores (Outtz, 2002; Potosky, Bobko, & Roth, 2005; Reilly, 1996; Sackett et al., 2001). Recently, Sackett et al. (2001) reviewed some alternative employment tests and found that the tests provided no easy method for improving validity. Likewise, Potosky et al. (2005) found little evidence to support the contention that alternative selection methods automatically result in increased predictive accuracy coupled with smaller race differences.

Some researchers have postulated new constructs to be assessed in employment tests. These concepts include practical intelligence (Sternberg et al., 2000), successful intelligence (Sternberg et al., 2000), and emotional intelligence (Bar-On, 2000, 2002; Goleman, 1995; Mayer, Salovey, & Caruso, 2000). Although research on new constructs is encouraged, the value of such constructs needs to be evaluated. For example, assertions about the structure and correlates of practical intelligence as espoused by Sternberg et al. (2000) are not supported on the basis of reviews of research (Gottfredson, 2002; McDaniel & Whetzel, 2005). Likewise, a mixed-model measure of emotional intelligence has been demonstrated to be primarily an assessment of long-standing personality constructs (the Big Five: Agreeableness, Conscientiousness, Extraversion, Emotional Stability, and Openness) and is not recommended for use as an employment test because of the ease with which

respondents can raise their score by faking (Grubb & McDaniel, 2007). Ability-based emotional intelligence models have been offered as a better alternative to the more personality-loaded mixed-model emotional intelligence tests. However, even with an ability-based emotional judgment test, Schulte, Ree, and Carretta (2004) concluded that the test was primarily saturated with cognitive ability and personality constructs. The substantial overlap of newly offered constructs with existing traditional constructs limits their ability to improve substantially the predictive accuracy of employment tests. Although research into new constructs is an important research goal for employment testing, research to date on alternative predictors is not encouraging.

Employment Tests Can Be Easily Faked

Some critics of employment tests argue that employment tests can be easily faked. This criticism is primarily directed at tests of personality, mixed-model emotional intelligence, or integrity. It is also sometimes directed at employment interviews. The criticism is based on the idea that the answers judged desirable by the organization are apparent to the applicant. Consider these items illustrative of those on personality tests purporting to measure conscientiousness:

- I am always prepared.
- I pay close attention to details of assignments.
- I seldom make errors at work.
- I complete my work successfully.

It would be clear to most applicants that the employer is seeking to hire someone who is dependable, and it would not be difficult for undependable applicants to assert that they are dependable.

The research literature supports the contention that such self-report measures are easily faked such that less suitable applicants can appear suitable. Viswesvaran and Ones (1999) conducted a meta-analysis to evaluate the extent to which one could improve one's scores on measures of the Big Five personality factors. They found that individuals could raise their scores on average by about 0.5 standard deviation. If an applicant has an average level (50th percentile) of conscientiousness, the applicant can respond so as to score at the 70th percentile. Thus applicants who choose to misrepresent themselves can respond so that they appear more suitable than those who truly are (Douglas, McDaniel, & Snell, 1996). Efforts to reduce faking on these measures is a topic of substantial research (Griffith, 2006). To date, however, no highly effective strategies to identify fakers or prevent faking on self-report measures have been developed.

Employers should recognize the limitations of such measures. When respondents are instructed to fake, the validity of employment tests sharply

decline (Douglas et al., 1996). Nonetheless, personality tests have nonzero validities in applicant samples in which one would expect at least some of the respondents to be faking (Hough, 1998). Some employers use these instruments as pass–fail screens with low cutoffs to screen out those applicants who are both ill suited for the position and willing to admit it. These employers then use other employment tests to differentiate among those who passed the highly "fakable" test. Other employers shun personality tests and instead rely on nonfakable measures such as cognitive ability, job knowledge, or situational judgment with instructions to identify the best response (McDaniel, Hartman, Whetzel, & Grubb, 2007). These are maximal performance measures (Cronbach, 1984) in that applicants know they should answer to the best of their ability. In such situations, faking is not possible because the instructions are to provide the best answers and not descriptions of typical performance (Nguyen, Biderman, & McDaniel, 2005).

CRITICISMS CONCERNING ADMINISTRATIVE EFFICIENCY

Employment Tests Are Not Needed Because One Can Train Anyone to Do Any Job

The authors have sometimes heard from practitioners that testing is not needed because one can train anyone to do any job. Ultimately, this is an argument that all people are roughly equal in most regards in terms of traits and abilities. Although this is a socially desirable position to take, it is not grounded in reality. This criticism ignores a large literature indicating that cognitive ability is a strong predictor of training performance (Brown, Le, & Schmidt, 2006; Nijenhuis & van der Flier, 2000; Pearlman et al., 1980). Cognitive ability places limits on the amount and complexity of material that one can learn. The U.S. military established minimum cutoff scores for cognitive skills of individuals whom it is willing to accept because its research showed that individuals below these scores could not readily be trained.

This criticism of testing may sometimes reflect the idea that certain jobs are so simplistic or remedial, and such large swaths of the population could be trained to do them with ease, that employment testing is unnecessary. A counterpoint to this criticism would be that in any job in which differences in performance are possible, it is in the interest of the employer to select the applicant who can achieve peak performance in the shortest period of time. More goes into making a good worker than the ability to perform specific job duties. Even in the case of the most simplistic work, the employer would still likely desire employees to possess certain personality traits such as conscientiousness (showing up to work on time), agreeableness (getting along with managers and coworkers), and integrity (not engaging in counterproductive work behaviors). Employers would be better served by

having at least some idea of where an applicant ranks on these traits regardless of how fast he or she can be trained to perform the job.

Employment Testing Takes Too Long

Some criticisms of employment testing do not concern the testing itself but the time it takes to develop and administer employment tests. Employment testing does take some time. Often job applicants apply for jobs in many organizations at the same time, and any delays associated with screening applicants may result in the best applicants being hired by the employer with the fastest screening procedure. Time needed to screen applicants increases when customized employment tests are developed. Likewise, at least for U.S. employers, the *Uniform Guidelines* encourages the use of time-consuming content validation procedures that typically entail the collection of detailed job analysis information. To minimize their legal liability in the face of adverse impact, many employers will not administer an employment test until a content validity study has been completed. Employment screening of public-sector employers is often particularly slow because of civil service regulations and other bureaucratic obstacles (Thompson, 2006).

Some of the time lag that critics attribute to employment testing is actually the result of other requirements of the larger recruitment and hiring effort. Vacant positions often need to be advertised. Applicants need time to apply. They also need time to consider offers and give notice to their current employers. The time associated with these activities would be spent regardless of whether employment testing was used.

Some technology has been applied to reduce employment testing time. Online applications have become common (Blyth, 2004), and employers are increasingly incorporating online employment testing into their application process.

Employment Testing Costs Too Much

Some employers shun employment testing because of its costs. The professional development of an employment test does require resources. However, many commercially available tests are inexpensive (less than $10 per administration). When compared with other costs associated with the hiring process (e.g., vacancy advertising, recruitment expenses), employment test costs are relatively insignificant.

More important, this criticism ignores the substantial productivity gains associated with employment testing. Good employees are much more productive than poor employees. Excellent employees are much more productive than good employees. Employment testing results in increased work output and improved learning of job skills (Hunter, Schmidt, & Judiesch, 1990). The dollar value of the increased productivity associated with using employ-

ment tests is substantial (Schmidt & Hunter, 1998) and makes the cost of testing trivial in comparison.

SUMMARY AND CONCLUSION

This chapter has reviewed 11 criticisms of employment testing. The criticisms fall into three categories: fairness, accuracy, and administrative efficiency. Two fairness criticisms arise from concerns about ethnic–racial differences in employment test results and the subsequent disparities in hiring rates. One concern over fairness arises from a belief in the discredited situational specificity hypothesis. A final fairness concern involves perceptions of invasion of privacy. Criticisms concerning the accuracy of employment testing center on the inability of employment tests to yield perfect predictions. Two of these criticisms concern some measures or test content being better than other content. A final criticism concerns the ease with which some employment tests, particularly personality and integrity tests, can be faked. Three criticisms are associated with the administrative efficiency of employment tests. One criticism falsely asserts that anyone can be trained for any job. Another concerns the time it takes to conduct employment testing, and the final administrative efficiency criticism concerns the cost of testing.

Human resources personnel must be aware of the importance testing plays in selection and the common pitfalls to avoid when improper selection procedures are used. Employers must invest some time and effort into deciding what characteristics they seek and should select measures that will help them best identify the applicants possessing those traits. There is no perfect measure, and employers should be wary of anyone (academic, consultant, or test publisher) purporting to possess one. In all likelihood, whichever testing method an employer uses will result in the rejection of qualified applicants on occasion and the acceptance of unqualified applicants on other occasions, but employers must not lose sight of the fact that formalized testing with valid measures will provide the best odds for avoiding these costly mistakes.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bar-On, R. (2000). Emotional and social intelligence: Insights from the Emotional Quotient Inventory. In J. D. A. Parker (Ed.), *The handbook of emotional intelligence* (pp. 363–390). San Francisco: Jossey-Bass.

Bar-On, R. (2002). *Bar-On Emotional Quotient Inventory: Short technical manual.* Toronto, Ontario, Canada: Multi-Health Systems.

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology, 31,* 233–241.

Blyth, A. (2004, November 16). Winning recruitment race. *Personnel Today,* p. 26.

Boehm, V. R. (1972). Differential prediction: A methodological artifact? *Journal of Applied Psychology, 62,* 146–154.

Bray, D. W., & Moses, J. L. (1972). Personnel selection. *Annual Review of Psychology, 23,* 545–576.

Brown, K. G., Le, H., & Schmidt, F. L. (2006). Specific aptitude theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment, 14,* 87–100.

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989, March 31). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17,* 913–920.

Douglas, E. F., McDaniel, M. A., & Snell, E. F. (1996, August). *The validity of noncognitive measures decays when applicants fake.* Paper presented at the Proceedings of the Academy of Management, Cincinnati, OH.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). *Uniform guidelines on employee selection procedures. Federal Register, 43*(166), 38290–39315.

Frei, R., & McDaniel, M. A. (1998). The validity of customer service orientation measures in employee selection: A comprehensive review and meta-analysis. *Human Performance, 11,* 1–27.

Goleman, D. (1995). *Emotional intelligence.* New York: Bantam.

Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence, 31,* 343–397.

Griffith, R. L. (2006). *A closer examination of applicant faking behavior.* Greenwich, CT: Information Age.

Grubb, W. L., III, & McDaniel, M. A. (2007). The fakability of Bar-On's Emotional Quotient Inventory Short Form: Catch me if you can. *Human Performance, 20,* 43–59.

Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11,* 209–244.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75, 28–42.

Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., & Katzell, R. A. (1968). *Testing and fair employment*. New York: New York University Press.

Marchese, M. C. (1992). Clinical versus actuarial prediction: A review of the literature. *Perceptual and Motor Skills*, 75, 583–594.

Mayer, J. D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 396–420). New York: Cambridge University Press.

McDaniel, M. A. (2007). Validity generalization as a test validation approach. In S. M. McPhail (Ed.), *Alternative validation strategies* (pp. 159–180). San Francisco: Jossey Bass.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L., III. (2007). Situational judgment tests, response instructions and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.

McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, 33, 515–525.

Murray, C. (2005, September). The inequality taboo. *Commentary*, 120, 13–22.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instruction on faking a situational judgment test. *International Journal of Selection and Assessment*, 13, 250–260.

Nijenhuis, J. T., & van der Flier, H. (2000). Differential prediction of immigrant versus majority group training performance using cognitive ability and personality measures. *International Journal of Selection and Assessment*, 8, 54–60.

Outtz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance*, 15, 161–172.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373–406.

Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment*, 13, 304–315.

Reilly, R. R. (1996). Alternative selection procedures. In R. S. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 208–221). Westport, CT: Quorum Books/Greenwood.

Roth, P. L., BeVier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297–330.

Rushton, J. P., & Jensen, A. R. (2006). The totality of available evidence shows the race IQ gap still remains. *Psychological Science, 17*, 921–922.

Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology, 7*, 261–277.

Sackett, P. R., Schmitt, N., Ellingson, J., & Kabin, M. B. (2001). High stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56*, 302–318.

Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior, 33*, 272–292.

Schmidt, F. L., Gast-Rosenberg, I. F., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology, 65*, 643–661.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research. *American Psychologist, 36*, 1128–1137.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology, 66*, 166–185.

Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. *Personnel Psychology, 33*, 705–724.

Schulte, M., Ree, M. J., & Carretta, T. R. (2004). Emotional intelligence: Not much more than *g* and personality. *Personality and Individual Differences, 37*, 1059–1068.

Society of Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., et al. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.

Taylor, P., Keelty, Y., & McDonnell, B. (2002) Evolving personnel selection practices in New Zealand organisations and recruitment firms. *New Zealand Journal of Psychology, 31*, 8–18.

Thompson, J. R. (2006). The Federal Civil Service: The demise of an institution. *Public Administration Review, 66*, 496–503.

U.S. Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (codified in various sections of 42 U.S.C.) (Suppl. III 1992).

Verive, J. M., & McDaniel, M. A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. *Intelligence, 23*, 15–32.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982). *Ability testing: Use, consequences, and controversies*. Washington, DC: National Academy Press.