DOCUMENT RESUME

ED 317 709 CE 054 125

AUTHOR Rudner, Lawrence M.
TITLE Preemployment Testing and Employee Productivity. Background Paper No. 40.
SPONS AGENCY Department of Labor, Washington, DC. Commission on Workforce Quality and Labor Market Efficiency.
PUB DATE Sep 89
CONTRACT 99-9-4782-75-038-04
NOTE 35p.; In "Investing in People: A Strategy to Address America's Workforce Crisis" (CE 054 080).
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Employer Attitudes; *Employer Employee Relationship; Evaluation; Expectation; *Job Applicants; *Occupational Tests; *Personnel Selection; *Predictive Validity; Productivity; Test Bias; Testing Problems; Test Reliability; *Test Use
IDENTIFIERS *Honesty Tests

ABSTRACT

Although frequently attacked as invalid, demeaning, biased, illegal, and irrelevant, preemployment testing procedures appear to be increasing in popularity. Many prominent companies and organizations are making extensive use of tests. Part of the resurgence of testing is attributable to clearer definitions of acceptable practice. Legal precedents and federal and professional guidelines help both the test developer and its users. Under the right conditions, preemployment testing can vastly improve corporate productivity, but there is little evidence to indicate that companies can properly implement a testing program or evaluate its effectiveness. Tests should only be used to enhance an employment decision, not to replace professional judgment in making decisions. The utility of a testing program can be estimated as a function of three factors: (1) the predictive validity of the test; (2) the selection ratio; and (3) the base rate. Issues in testing include bias, legal rulings, validity generalization, exaggerated expectations, test quality, misuse of tests, publishers' claims, alternative assessment techniques, and the use of honesty tests. The paper concludes with recommendations to business and to the Department of Labor. A 50-item reference list is included. (CML)

40. PRE-EMPLOYMENT TESTING AND EMPLOYEE PRODUCTIVITY

Lawrence M. Rudner

LMP Associates and
American Institutes for Research

This project was funded under Purchase Order No. 99-9-4782-75-038-04 from the U.S. Department of Labor, Commission on Workforce Quality and Labor Market Efficiency. Opinions stated in this document do not necessarily represent the official position or policy of the U.S. Department of Labor, Commission on Workforce Quality and Labor Market Efficiency.

40. PRE-EMPLOYMENT TESTING AND EMPLOYEE PRODUCTIVITY

Lawrence M. Rudner
LMP Associates and
American Institutes for Research

Although frequently attacked as invalid, demeaning, biased, illegal, and irrelevant, pre-employment testing procedures appear to be increasing in popularity. The American Society for Personnel Administration found 39 percent of 360 companies surveyed were testing more in 1985 than in 1980, and 44 percent were considering even more testing. A 1988 survey of 245 human resource executives by the Bureau of National Affairs (a publisher) found that 63 percent of surveyed companies ask applicants to supply work samples or take performance tests, while 30 percent require ability tests, and 25 percent test for job knowledge. A new referral system being considered by the U.S. Employment Service, a part of the Department of Labor, could result in the testing of several million applicants annually.

Many prominent companies and organizations are making extensive use of tests. The Illinois Department of Employment Security used a written multiple choice test to screen 50,000 blue collar applicants at Diamond Star. At American Telephone and Telegraph, testing is a routine part of hiring and promotion through the second layer of management. International Business Machines uses skill and aptitude tests to evaluate applicants for about 75 percent of entry level jobs. Manpower expects to test over 700,000 applicants this year. Corporate executives, state officials, and federal policymakers are discovering that the judicious use of formal assessment procedures may lead to

increased efficiency and productivity. The benefits of testing appear to outweigh its costs and concerns.

Part of the resurgence of testing is attributable to clearer definitions of acceptable practice. The landmark case of Griggs v. Duke Power (1971) resulted in a legal precedent requiring defendants to demonstrate adequate validity. In 1978, the Equal Employment Opportunity Commission established "Uniform Guidelines on Employee Selection Procedures." In 1974 and again in 1986, the American Psychological Association, the National Council on Measurement in Education, and the American Educational Research Association adopted professional standards for educational and psychological tests. And in 1987, the Society for Industrial and Organizational Psychology issued the third edition of its own principles for the validation and use of personnel selection procedures.

Legal precedents and federal and professional guidelines help both the test developer and its users. The developer can conduct appropriate studies and prepare necessary documentation. When assured that tests meet legal and professional standards, potential customers can use them with greater confidence.

Under the right conditions, pre-employment testing can vastly improve corporate productivity. But, testing is marked with issues that employers are often ill-equipped to handle. What does an employer do about black applicants who, on average, score lower than whites on standardized tests? How does an employer demonstrate that a test is job-related? Failure to have good answers to these questions could easily result in litigation. However, readily available "good" answers are lacking. Once a testing program is found to adversely impact a

protected group, the burden of defending the program rests with the employer. The measurement community, the courts, and professional associations are divided on these and other issues. Further, there are no groups dedicated to providing employers with objective information regarding testing issues and practice.

While testing can lead to increased productivity, there is little to indicate that companies can properly implement a testing program or evaluate its effectiveness. Many reputable test publishers quickly point out that the average consumer places too much value on testing (Deutsch, 1988). At best, tests only estimate a person's ability or the extent to which a person possesses some attribute. Tests should only be used to enhance an employment decision. Too often, test results are treated as scientific evidence that inappropriately replaces professional judgment in making decisions.

This paper describes the conditions under which pre-employment testing can improve productivity. It identifies special problems and issues associated with employment testing and makes appropriate recommendations for federal action.

THE UTILITY OF FORMAL ASSESSMENT

The utility of a selection procedure may be defined as the increase in productivity as a result of incorporating that procedure. Taylor and Russell (1939) and Brogden (1949) have shown that the utility of a testing program can be estimated as a function of just three factors:

1. the correlation between test scores and job productivity (the predictive validity of the test),
2. the percentage of applicants being hired (the selection ratio), and
3. the level of performance necessary for someone to be considered successful -- defined as the proportion of all applicants who would be classified as successful (base rate).

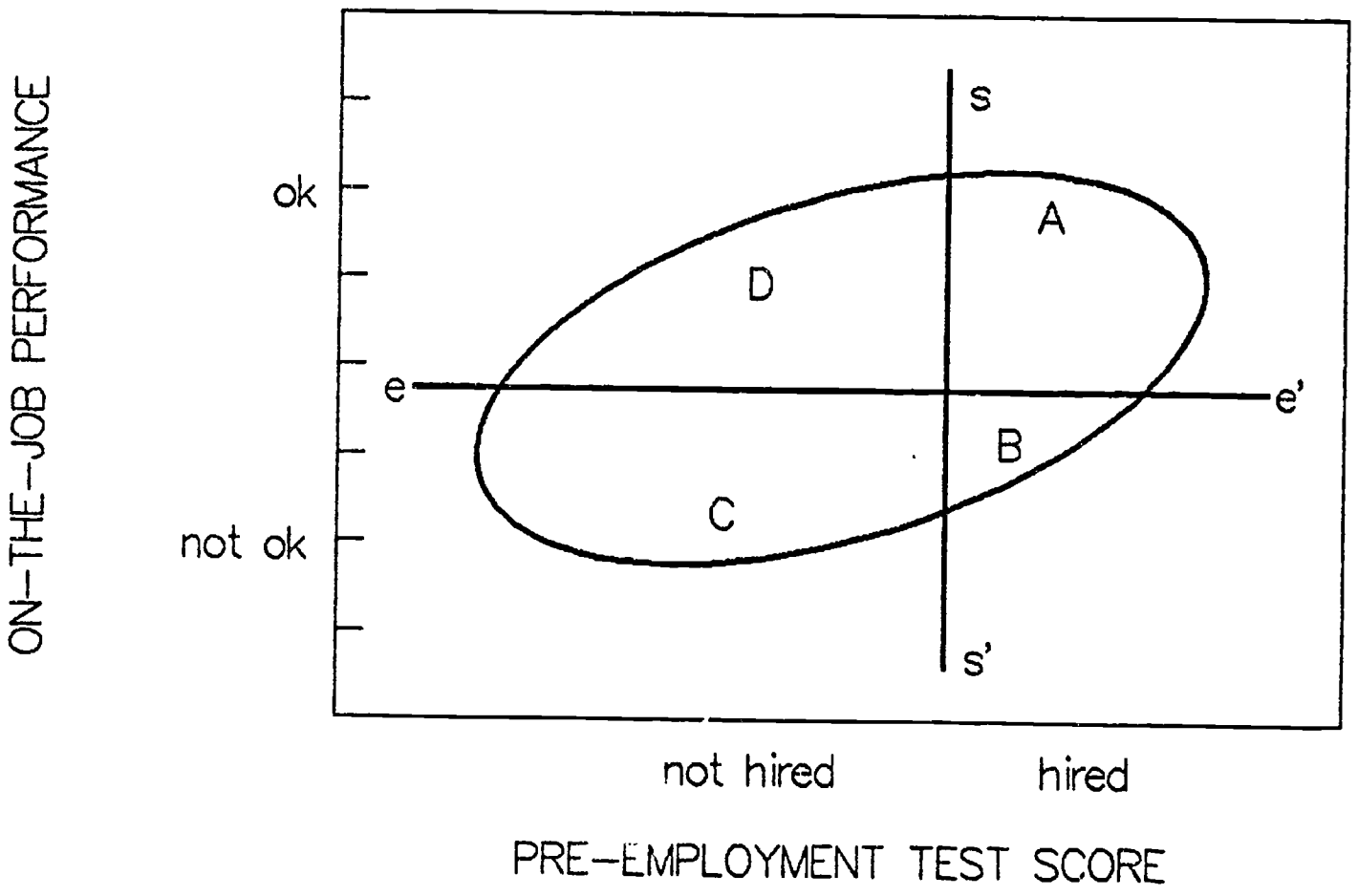
Figure 1 illustrates the relationship between test scores and performance. For example, suppose we have a large group of examinees and each examinee has two scores -- one pre-employment test score and one measure of on-the-job performance. If we plot test scores along the x-axis and performance levels along the y-axis, the sets of scores would result in a scatterplot in the shape of an ellipse. The orientation of the ellipse reflects the correlation between testing and job performance. The closer the orientation is to 45°, the greater the validity of the test. In Figure 1, the correlation between testing and job performance, that is, the validity coefficient, is .35 -- the value Ghiselli (1973) found to be average for proficiency criteria.

Now suppose, the employer uses test scores to hire new applicants. Those scoring above a certain cut-score are hired; those below that value are not. Here, the cut score is shown by line s-s' and the selection ratio is .2.

Finally, suppose we can define a satisfactory level of performance. Individuals whose performance is above that level are considered satisfactory. Those below that level re considered unsatisfactory. Here, line e-e' denotes a satisfactory level of

Figure 1

Testing and Productivity



performance and 60 percent of the current employees are working at a satisfactory level (i.e., the base rate is .6).

From an increased productivity viewpoint, the goal is to maximize the success rate -- the proportion of hired individuals who are qualified. Mathematically this can be expressed as the number of individuals in quadrant A divided by the number of individuals in quadrants A+B. A new testing program is effective when its success rate exceeds the current base rate, that is, when the proportion of qualified new hires exceeds the proportion of currently qualified employees.

Success rate will increase as

1. The job becomes less difficult for the applicants. This raising of the base rate can be visualized by moving line e-e' down.
2. Fewer individuals are hired from the applicant pool. This lowering of the selection ratio can be visualized by moving line s-s' to the right.
3. Better tests are used. This use of a test with a higher validity can be visualized by orienting the ellipse more toward 45°.

Of the options, improving validity has the least effect on success rate. Increasing hiring selectivity. i.e., decreasing the selection ratio, has the greatest effect. From a productivity viewpoint, recruitment is far more effective than using a better test.

In Figure 1, the success rate was .78. Different combinations of validity, selection rate, and base rate can result in the same success rate (Taylor and Russell, 1939). Following Linn (1984), Table 1 shows various combinations of selection ratio and validity computed by Taylor and Russell, yielding a success rate of .7, when the base rate is .6.

Table 1

Some Base Rates and Validities Yielding
A Success Rate of .7
(Base Rate = .6)

Selection Ratio	Validity
.10	.15
.20	.19
.40	.29
.60	.40
.80	.65

If only ten percent of applicants are to be hired (selection ratio=.10), even a relatively poor test -- e.g., one with a validity coefficient of .15 -- can lead to an improvement in productivity. On the other hand, if 80 percent of the applicants are to be hired, then a test with a high validity coefficient, .65, is needed to yield the same improvement.

The Taylor and Russell analysis is most applicable where job performance may be classified satisfactory or unsatisfactory (such is the case with many production jobs). By showing how to estimate average job performance level as a function of selection ratio and test validity, Brogden (1949) provided a way to determine utility without making such classifications.

By Brogden's method, the dollar value (U) of increased output attributable to pre-employment assessment is

$$U = N * T * r_{xy} * SD_y * M$$

where

N is the number of workers hired;

T is the average tenure in years;

r_{xy} is the correlation between the predictor and job performance;

SD_y is the standard deviation of performance in dollars; and

M is the mean predictor score of those hired, expressed as a standard score.

N, T, and M are determined by the individual organization. M is a function of the selection ratio: the fewer applicants hired, the larger the value of M. SD_y quantifies on-the-job performance. Hunter and Schmidt estimate SD_y to be 40 percent of the annual wage when better estimates are lacking.

Illustrating Brogden's method is an example of pre-employment screening of budget analysts provided by Schmidt, Hunter, McKenzie, and Muldrow (1979). They estimate the SD_y as \$11,327. If 20 percent of 200 applicants are hired using a test with r_{xy} = .53 and the mean tenure is 6 years, then

$$U = (200 * .20) * 6 * .53 * 11,326 * 1.40 = \$2,017,112$$

Schmidt, Hunter, McKenzie, and Muldrow cite this as a big improvement over the utility of using just an interview. With a validity coefficient of .14, the interview would provide a utility of just

$$U = (200 * .20) * 6 * .14 * 11,326 * 1.40 = \$532,822$$

which has only 1/4 of the utility.

In evaluating his now classic utility formula, Brogden concluded that

1. "A low selection ratio can be as important or even more important than high test validity in achieving savings."
2. "Even a test with very low validity can produce substantial savings if it is possible to select only a small percentage of those who apply."
3. "Even highly valid selection procedures are of little value if nearly all those who apply must be hired."

These conclusions are consistent with those of Taylor and Russell.

For years Brogden's paper was widely recognized for its theoretical value. However, because of difficulty in estimating a value for SD_y, few researchers were able to apply his derivation to actual data. Using recently developed methods to estimate SD_y, prominent researchers F.L. Schmidt and John Hunter have applied Brogden's equation and made rather startling claims about the benefit of pre-employment tests. They provided the example above where productivity increased to $2 million over a six-year period. Hunter (1983) estimated that testing could lead to $15 billion worth of increased productivity per year for the federal government, and with Schmidt (1982), calculated that the gross national product would increase from $80 to $100 billion if improved selection procedures were introduced throughout the economy.

However attractive and appealing, these estimates are not realistic. Levin (1989) asserts that Hunter and Schmidt overgeneralize the applicability of the Brogden formula, use unrealistic estimates of r_{xy} and the selection ratio, and use questionable estimates of SD_y.

Brogden's formula is based on the correlation between the assessment procedure and on-the-job performance. Yet on-the-job performance is often hard to measure. While some studies have used supervisor ratings, a common proxy for on-the-job performance has been paper-and-pencil tests of job knowledge. Levin argues convincingly that using such measures will result in inflated validity coefficients.

Another critic (Cronbach, 1984) calls these projections "a fairy tale." He claims that the projections are based on the untenable assumption that the selection ratio will remain a constant when the formula is applied to large numbers. When one starts to consider the universe of new hires in a given field, however, hiring becomes less selective. While a prestigious company may be able to be highly selective, there is little basis for assuming highly selective hiring when estimating utility across several institutions.

ISSUES

Bias

Test bias often means different things to different people. Flaugher (1978) has shown that the term has been used to refer to differences between groups in average scores, language demand, validity, content relevance, content offensiveness, and selection rates. Two definitions are particularly relevant in employment testing -- differences in validity and differences in selection. Does a particular test predict on-the-job performance of minority applicants relative to white applicants and does it result in unequal hiring rates?

The majority of tests developed by reputable companies do predict job performance equally well for minority and white applicants (Gottfredson, 1988). Little evidence exists that tests are biased using the first definition. Given that test scores of minority applicants are on average below those of whites (Jensen, 1981), the use of an employment test will often result in bias, or unfairness, by the second definition. The selection ratio for minorities will generally be lower than the selection ratio for white applicants.

Employers, then, face two difficult technical and legal challenges: how can they use tests to increase productivity while they strive for greater equity in hiring? and how can they defend their testing program against litigation?

The technical issues were addressed by the National Academy of Science at the request of the Department of Labor. They conducted a thorough, scientific evaluation of a proposed test-based employee referral system based on the Department of Labor's General Aptitude Test Battery.

In their report, Hartigan and Wigdor (1989) evaluated six selection rules:

Raw-Score, Top-Down Selection -- Applicants are selected in order of their scores on the test, from high to low. Use of this rule will result in the highest utility and, given group differences in average test scores, this approach will have the greatest adverse impact on minority group applicants. Employers using this approach should be prepared to demonstrate that they did not have discriminatory intent.

Within-Group Percentile, Top-Down Selection -- a percentile score is computed for each applicant using norms for his or her racial group. Applicants are then selected in order of their percentile scores, from high to low. This method is equivalent to the raw-score, top-down method with a constant added to the scores of minority applicants. Compared to the raw-score, top down method, this will result in a slight loss in utility and substantially increase minority referrals. When this approach was adopted by the U.S. Employment Service for an experimental referral plan, the U.S. Assistant Attorney General for Civil Rights stated that this approach "not only classifies job applicants on the basis of their race or national origin, but... requires job service offices to prefer some and disadvantage other individuals based on their membership in racial or ethnic groups. Such a procedure constitutes intentional racial discrimination." (Reynolds, 1986).

Minimum Competency Selection -- Applicants with a raw score exceeding some cut-score are randomly selected. Of the non-race-conscious rules, this rule results in the highest proportion of minority selections. The utility of this approach is generally much lower than that of other approaches. It is most applicable to jobs where most satisfactory workers have similar performance levels. This approach has been advocated by the Equal Employment Opportunity Commission.

Zone Score, Random Within-Zone Selection -- The test score range is divided into interval zones containing the same number of applicants. All applicant scores within a zone are converted to the same zone score. Applicants are then selected based on their zone scores, top down. Applicants in the lowest acceptable zone are randomly selected. The

utility of this approaches decreases as the selection ratio decreases. Minority representation increases negligibly.

Zone Score, Preferential Within-Zone Selection -- This is identical as the Zone Score, Random Within-Zone Selection method except minority applicants in the lowest acceptable zone are selected first. This procedure has the same characteristics as the Zone Score, Random Within-Zone Selection method with a slight decrease in utility and a slight increase in minority representation.

Expected Performance Ratio Selection -- An applicant's test score is converted to an expected level of performance. Hiring is then top-down, based on the expected score. This approach corrects the disadvantage to a minority group caused by a less than perfect match between a test and job performance. The higher the test validity, the closer this method is to the raw-score, top-down method. The lower the validity, the closer this approach is to within-group percentiles.

Regardless of the procedures used, employers incorporating testing in their hiring practices are placed in a difficult position. If they use a color-blind procedure, minorities will often be adversely impacted and the employer may be accused of intentional discrimination. Using a race-conscious procedure may decrease the utility of the assessment and result in charges of reverse discrimination.

Noting problems with each of these approaches, the National Academy of Science has developed a selection policy "that would allow employers to strike an appropriate compromise between the interests of productivity and racial balance in the workforce."

In their interim report, Wigdor and Hartigan (1989) conclude that

> "If the will of society is to pursue both high levels of productivity and a racially balanced workforce and if a valid test that produces adverse impact is used in the referral process, than a race-conscious referral policy is necessary."

The National Academy of Science calls for race-conscious selection, even though the practice is contrary to the nation's efforts to establish non-discriminatory color-blind hiring practices. Thus we must turn to our legal system to provide guidance to the bias issue.

Legal Issues

Since Title VII of the Civil Rights Act in 1964, our legal system has shifted the rules with regard to bias in testing at least three times. Additional shifts are also likely in our nation's continuing struggle to achieve high levels of productivity and a racially balanced workforce. Scharf (1988), Bolick (1988), and Seymour (1988) discuss three landmark events defining the shifts to date:

1. In 1964, Congress defined employment discrimination in terms of "evil intent". Plaintiffs were able to cite disparate impact as evidence of such intent.
2. In 1971, the Supreme Court concluded in *Griggs v. Duke Power Company* that if a test produces adverse impact and is not job related, then it is reasonable to infer that it is being maintained for some other reason. As a result, employers have been compelled to prove that their test predicts a reasonable measure of job

performance and, of the alternatives, that it has the least adverse impact.

3. In 1988, the Supreme Court shifted the burden of proof in *Watson v. Fort Worth Bank and Trust*. Under *Watson*, the plaintiff must specify the criteria that result in adverse impact and the employer must offer a "legitimate business reason" for a testing program. Further, Justice O'Conner emphasized that "employers are not required...to introduce formal validity studies showing that a particular criteria predicts actual on-the-job performances." The Court also ruled that adverse impact precedents also apply to subjective criteria and methods, such as interviews.

An elaboration of the legal issues with regard to employment testing are well beyond the scope of this paper. The interested reader is referred to the December 1988 issue of the *Journal of Vocational Behavior* which was dedicated to the issue of fairness in employment testing. This excellent volume brings into focus the changing nature and importance of the current legal, scientific, and social debate over fairness in employment testing. With regard to race-conscious selection, precedents are presented on both sides of the debate.

Validity Generalization

One of the most impressive and controversial bodies of testing research in recent years was sponsored by the U.S. Department of Labor. In part of it, Hunter and Schmidt claimed that the validity of a test predicting success in some occupations may make the test applicable to a

much larger number of occupations than had previously been thought. Using 515 research studies of the General Aptitude Test Battery (GATB) of the U.S Employment Service, Hunter (1983) claimed that the GATB is valid for up to 12,000 different jobs.

This concept of validity generalization can markedly effect an employer's responsibility with regard to test use. Prior to Justice O'Conner's opinion in *Watson*, virtually all employers using a test were expected to conduct local validity studies to ascertain the appropriateness of a test in their situation. With *Watson* and the concept of validity generalization, employers can cite other studies as evidence that their testing program is valid. Under this logic, employers are relieved of the burden of conducting their own validation studies to document the appropriateness of their testing activities. Yet to be resolved are what constitutes a "compelling" body of evidence and to what extent that body of evidence may be generalized to a local situation.

The extent to which validity generalizes is a function of what a test measures and what is involved in the job. A performance test that adequately predicts on-the-job performance of clerical workers in one state, for example, will probably also predict the performance of clerical workers in another state. The jobs do not markedly differ across state boundaries. However, Hunter takes the concept of validity generalization much further: if a massive data base consistently shows a high correlation between a given test and different jobs, then the test is valid for all jobs. In *EEOC v. Atlas Paper Box Company*, Hunter testified that since general intelligence tests are valid for all jobs and since the Wonderlic IQ test is a good measure of general cognitive

ability, the test is valid for clerical jobs at the Atlas Paper Box Company.

Although aptitude and intelligence are important in any job -- a fact supported by impressive statistical evidence -- Hunter's conclusions are not totally accepted by the legal and research communities. In *Van Aken v. Young*, the court rejected the concept that a general intelligence test is automatically valid for selecting firefighters, Levin (1989) challenged the evidence used in the original validity generalization studies, and Linn and Dunbar (1986) raised questions about statistical biases. Sackett et al. (1985) claimed that Schmidt and Hunter exaggerated the magnitude, conclusiveness, and policy relevance of their findings; Cronbach (1984) pointed out that variations in validity are far from "minute, decimal dust," as claimed by Schmidt and Hunter.

Nonetheless, despite debate in the research community, the concept of validity generalization has markedly influenced state and federal testing policy. By 1987, the public employment service systems in 37 states were using validity generalization to justify employment tests, thus allowing publishers of commercial tests to claim that validity generalization obviates the need to conduct local validation efforts.

Exaggerated Expectations

Levin (1989) notes the extensive writings on the relationship between various worker attributes and worker productivity. The literature includes

- cognitive dimensions such as verbal and mathematics ability;
- physical attributes such as perceptual skills and strength;

- social/affective characteristics, such as interpersonal skills and temperament; and
- personality traits, such as diligence (Dunnette, 1983; Fleishmann and Quaintance, 1984; and McCormick, 1979).

The best workers are not necessarily the ones with the most skill or knowledge. As the promotional literature for the Wonderlic Personnel Test states:

> "While we may be dazzled by the performance of the exceptionally bright employee and frustrated by the slowness of a dull employee, the real work of an organization is done by those with sufficient mental ability coupled with punctuality, cooperation, leadership, consistency, and persistency."

No assessment program can measure all relevant traits, many of which lack clear definition. In other cases we simply do not have instruments of sufficient quality for testing, and in any event the interplay of traits varies greatly between workers and their jobs. To be sure, tests can be useful prospective instruments, especially when a small percent of the applicants are to be selected, yet tests of human characteristics must always misclassify significant numbers of individuals. (See Figure 1, where many misclassified individuals appear in regions B and D.) Moreover, it is doubtful that many employers understand the conditions under which tests are useful, or that they properly select and use assessment instruments. As we will see below, the problem for employers using screening tests is a matter of quality, quantity, validity, and consideration of required alternatives.

The Matter of Quality

While there are at least 3,000 different tests sold commercially by at least 450 vendors, the objective information about psychological testing available to American companies is quite limited. Notable sources of testing information are the Buros Institute, Test Corporation of America, and the Educational Resources Information Center Clearinghouse on Tests, Measurement, and Evaluation (ERIC/TM). Buros Institute and Test Corporation of American publish extensive descriptions and reviews of commercially available tests (see Mitchell, 1983, 1985; Sweetland and Keyser, 1986; Keyser and Sweetland, 1985-1987). Concentrating on educational tests, the ERIC Clearinghouse prepares a database of published and unpublished literature and offers a range of information products.

Technical reviews of employment tests are available. There are, however, no organizations dedicated to improving employment testing practices. Research in the area is sporadic. Employers seeking objective, balanced information regarding technical, legal, and practical issues do not have a central source for information.

Test Users

In 1988, the American Psychological Association and five other professional associations published the "Code of Fair Testing Practices" (Joint Committee, 1988). Endorsed by major test publishers, the code specified the responsibilities of test developers and users. For the latter it outlined specific responsibilities regarding the selection of appropriate tests, score interpretation, fairness, and notification of test takers.

While test publishers and professional associations do not sanction violators of the code, test users are not held harmless by the courts. As a result of Griggs v. Duke Power, any company administering a psychological test had to demonstrate that the test was valid and necessary to fill a specific job. In 1988, the Supreme Court extended Griggs to interviews and less formal employee testing.

There is little evidence to indicate that users understand their tests or that they are meeting their responsibilities. Even an agency as well respected as the New York State Board of Regents was found to be misusing the SAT (a college admissions test) as the basis for scholarship awards. The court held that scholarships should be based on academic achievement, not aptitude for college.

Publishers' Claims

The testing industry is full of many specialized small companies catering to special markets. It is an attractive, unregulated growth industry: test publishers are given credit when their products support sound employment decisions, and they are usually held harmless when employers make wrong decisions on the basis of tests.

The *New York Times* (Deutsch, 1988) points out that testing is a multi-million dollar industry. While few large companies have revenues in excess of $100 million, many smaller companies nevertheless do very well. London House, Inc., a purveyor of honesty tests, posted sales of $37 million. As test use increases, the future looks bright for the industry.

Much of the promotional literature from larger, well established companies warns of the limitations of all tests, their's in particular.

The literature from smaller, highly specialized companies, however, is often full of exaggerated claims and poor recommendations. For example, in reviewing tests for an educational accrediting agency, this author found incorrect calculations, the use of data from different tests, unjustified (and ridiculously low) recommended passing scores, conflicting statements, improper interpretations of data, and grossly exaggerated claims. Companies without personnel who are trained in testing could easily fall prey to incompetent and dishonest test publishers.

Alternative Assessment Techniques

The Uniform Guidelines on Employee Selection Procedures require employers to investigate and use alternatives to conventional tests. Little guidance, however, is available to help employers evaluate or implement alternatives. The search for viable alternatives has focused on unassembled examinations, biodata banks, assessment centers, reference checks, and interviews.

Unassembled examinations: Unassembled examinations, also called experience and training exams, E&T examinations, and Traex exams, are structured evaluations of an applicant's job-related experiences. Such items as work experience, relevant education, and related achievements are scored. While they are often used to evaluate applicants for white-collar federal and state jobs, Davey (1984) noted that very little research exists on this approach.

Biodata banks: Biodata banks involve the weighted scoring of a wide range of background items that have been empirically shown to relate to performance. While E&T examinations are strictly job-related,

biodata banks usually contain a wide range of life history data that are necessarily clearly job-related. The literature strongly supports using biodata (Owens, 1976; Asher, 1972; Reilly and Chao, 1982).

Assessment Centers: Assessment Centers use a variety of work simulations, such as in-basket tests and job-related activities that are scored by multiple raters. Bray, Campbell, and Grant (1974) and Moses and Byham (1977) have found Assessment Centers to be effective for selecting managerial candidates. Davey (1984) however, noted that there is little published research on the validity of Assessment Centers designed to evaluate non-managerial candidates.

Reference Checks: Reference Checks refer to obtaining assessments of previous performance. While reference givers can supply potentially valuable information, negative references are relatively rare. Summarizing the literature, Reilly and Chao (1982) conclude that under most circumstances, this approach is not effective.

Interviews: Interviews can range from an unstructured, non-directed set of questions to a defined set of questions that is administered orally. Recognized as the most widely used method of personnel selection (Arvey, 1979), researchers have consistently concluded that interviews lack sufficient reliability and validity (Wagner, 1949; Mayfield, 1964; Arvey, 1979).

After examining over 170 studies of assessments, Reilly and Chao (1982) concluded that additional research on each of these forms of assessment is needed. Much of the research viewed the different forms of assessment only as alternatives to written cognitive tests. There are very few studies evaluating the potential gain of combining approaches or identifying the circumstances under which different approaches would be most effective.

Honesty Tests

Written and oral tests designed to measure an applicant's honesty (or truthfulness) have long been a popular alternative to the polygraph. Typically composed of 50-100 statements with which the applicant agrees or disagrees, honesty tests are relatively inexpensive ($12 v. $40 for a polygraph) and can be administered by a telephone interview. Bean (1988) reports that approximately 2.5 million honesty tests were administered in 1987. Since the use of the polygraph was prohibited in 1988, honesty tests have become a high growth industry.

While honesty tests predict employee theft as well as cognitive tests predict productivity, they raise a host of ethical issues. They raise the same issues as the polygraph -- a good number of individuals are always misclassified. They also raise questions about what should be permissible in an interview. The increased interest in honest tests may be a leading indicator of decreasing employer confidence in the integrity of American workers.

RECOMMENDATIONS TO BUSINESS

The *Watson* decision suggests that employers should carefully evaluate their hiring practices. Interviews are subject to the same professional standards as formal paper-and-pencil instruments. The questions asked must be job-related and serve legitimate business objectives.

In examining hiring practices, employers should consider the potential of professionally developed and validated assessment procedures. Such properly designed instruments can lead to increased

productivity, reduced turn-over, and greater employee satisfaction. Properly implemented, these instruments can withstand legal challenges.

Such instruments, however, will have limited utility for companies struggling to find qualified employees. Companies fortunate to have a large applicant pool, on the other hand, stand to benefit from improved selection procedures. A shrinking labor force will compound the selection problem for everyone.

In order to strike a balance between increased productivity and a racially balanced workforce, selection procedures will have to be race conscious. The Lawyer's Committee for Civil Rights Under Law points out that significant precedents exist for race-conscious hiring practices (Hartigan and Wigdor, 1989, page 50). However, just as racial equity may not be sacrificed for the sake of increased productivity, productivity may not be sacrificed for the sake of racial equity. The two goals are not mutually exclusive, although employers will need to be flexible in establishing and maintaining their workforce.

In order to defend hiring practices against possible litigation, employers should be prepared to document that

1) be it an interview or more formal instrument, the content of the assessment instrument must be related to the job,

2) the instrument serves a legitimate business purpose,

3) the instrument was developed to meet professional standards,

4) application of the instrument meets with professional standards.

Businesses not familiar with these standards should contact the American Psychological Association or the International Personnel Management Association for more information. Businesses lacking measurement expertise should use consultants to help them evaluate the claims of test publishers.

RECOMMENDATIONS TO THE DEPARTMENT OF LABOR

A variety of techniques can be used effectively to help employers assess how well applicants will fit within their organizations. Crowded with vendors of paper-and-pencil tests, the test marketplace has concentrated almost exclusively on that form of assessment. Businesses that are interested in using tests for assessment can readily turn to any of a number of testing companies for vendor advice and information.

However, businesses that want to get information from an objective source and that want to consider other forms of assessment have few, if any, resources available. Consumer-oriented materials are not available and research isn't being conducted. Without federal action, these areas will remain undeveloped.

Research, development, and dissemination activities, in

- understanding current practice
- improving test use
- improving test quality

are recommended. Activity in the first two areas will improve testing practice in American businesses and promote better use of existing instruments. Activity in the third area will address quality and improve the methodology of testing. Contracts, grants, regional technical assistance centers, and a central information clearinghouse are envisioned. This clearinghouse would build a bibliographic database of contract reports, conference papers, planning documents, validation studies, and other unpublished reports and make them available to business. The clearinghouse would proactively disseminate concise, clearly written information regarding testing practices. In short, the clearinghouse would serve as a central source of quality information concerning employment testing.

Understanding Current Practice

Aside from occasional small surveys by professional associations and journalists, there is little hard data about current testing practices. Efforts to improve practice through testing must begin with answers to questions about test use: how many and what types of tests are given annually? what types of tests are given? what are employer attitudes toward these tests? are they being used properly?

There is, then, a need for both large scale surveys and intensive case studies. Large scale studies can provide basic non-evaluative and descriptive information. Case studies stemming from the large scale surveys can help identify cause and effect relationships and identify areas requiring concentrated effort.

Improving Test Use

Activity in this area is needed to improve the ability of employers to use tests and their results appropriately. Material describing test selection, use, and evaluation should be developed and offered to American businesses. Applied research studies focusing on test selection and interpretation are also strongly recommended. Specific topics include

- making test information available and useful to employers
- identifying problems in using tests
- applying validity generalization
- identifying job requirements
- evaluating test utility
- establishing standards
- selecting employees and establishing equity
- addressing legal issues in employment testing
- using computers and testing
- providing feedback to applicants
- developing and documenting company-prepared tests

Improving Tests

Activities in this area will contribute to the development of new methods of assessing job potential. Research on alternative assessment techniques, interviewing practices, and test methodology is strongly recommended. Specific topics include

- creating and using job simulations
- targeting interviewing

- improving the diagnostic value of tests
- improving test efficiency
- assessing unskilled labor

REFERENCES

American Psychological Association (1974) Standards for Educational and Psychological Tests and Manuals. Washington, DC: author.

American Psychological Association (1986) Standards for Educational and Psychological Tests and Manuals. Washington, DC: author

Arvey, R.D. (1979) Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin*, 86, 736-765.

Asher, J.J. (1972) The biographical item: Can it be improved? *Personnel Psychology* 25, 251-269.

Bean, E. (February 27, 1987) More firms use 'attitude tests' to keep thieves off the payroll. *Wall Street Journal*, 37.

Bray, D.W., Campbell, R.J., & D.L. Grant (1974) Formative years of business: A long term study of managerial lives. New York: Pergamon.

Bolick, C. (1988) Legal and Policy Aspects of Testing, *Journal of Vocational Behavior*, 33, 320-330.

Brogden, H.E. (1949) When testing pays off. *Personnel Psychology*, 2, 171-183.

Brogden, H.E. (1951) Increased efficiency of selection resulting from replacement of a single predictor with several different predictors, *Educational and Psychological Measurement*, 173-195.

Cronbach, L.J. (1980) Selection Theory for a Political World, *Public Personnel Management*, 9,1, 37-50.

Cronbach, L.J. (1984) Essentials of Psychological Testing, New York: Harper and Row, fourth edition.

Davey, B.W. (1984) Personnel testing and the search for alternatives, Public *Personnel Management Journal*, 13, 4, 361-374.

Deutsch, C.H. (1988) A mania for testing spells money, *New York Times*, October 16, p.4.

Dunnette, M.D. (1983) Aptitudes, abilities, and skills. In M.D. Dunnette (ed) Handbook of Industrial and Organizational Psychology, New York: John E. Wiley and Sons, 473-520.

Equal Employment Opportunity Commission (1978) Uniform Guidelines on Employee Selection Procedures, *Federal Register* 43, 116, 38295 - 38309.

Flaugher, R.L. (1978) The many definitions of test bias. *American Psychologist*, 33, 7, 671-679.

Fleishman, E. & M. Quaintance (1984) Taxonomies of human performance, New York: Academic Press.

Ghiselli, E.E. (1973) The validity of aptitude tests in personnel selection, *Personnel Psychology*, 26, 461-477.

Gottfredson, L.S. (1988) Reconsidering fairness: A matter of social and ethical priorities, *Journal of Vocational Behavior*, 31, 3, 295.

Griggs v. Duke Power Co (1971) 401 U.S. 424.

Hartigan, J.A. and A.K. Wigdor, eds (1989) *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery*, Washington, DC: National Academy Press.

Hunter, J.E. (1980) Validity Generalization for 12,000 Jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB). Washington, DC: U.S. Employment Service, U.S. Department of Labor.

Hunter, J.E. (1983) The economic benefits of personnel selection using ability tests: A state of the art review including a detailed analysis of the dollar benefit of U.S. employment service placements and a critique of the low-cutoff method of test use. Washington, DC: Employment and Training Administration, U.S. Department of Labor.

Hunter, J.E. & F.L. Schmidt (1982) Fitting people to jobs: The impact of personnel selection on national productivity. In E.A. Fleishman & M.D. Dunnette (eds) Human Performance and Productivity: Vol 1 Human Capability Assessment. Hillsdale, NJ: Erlbaum.

Hunter, J.E. & F.L. Schmidt (1983) Quantifying the effects of psychological interventions on employee job performance and work force productivity. *American Psychologist*, 38, 473-478.

Jensen, A.R. (1980) Bias in Mental Testing, New York: The Free Press.

Joint Committee on Testing Practices, Code of Fair Testing Practices, Washington, DC: author.

Keyser, D.J., & R.C. Sweetland, eds., (1985-1987) *Test Critiques*. Kansas City, MO: Test Corporation of America, Volumes I-VI.

Levin, H. (1989) Ability testing for job selection: Are the economic claims justified? In B. Gifford Testing and the Allocation of Opportunity, Boston: Kluwer Academic Publishers, in press.

Linn, R.L (1982) Ability testing: Individual differences and differential prediction. In Ability Testing: Uses, Consequences, and Controversies, part II. Washington, DC: National Academy Press 335-388.

Linn, R.L. & S.B. Dunbar (1986) Validity generalization and predictive bias. In R.A. Berk (ed) Performance Assessment: Methods and Applications. Baltimore: Johns Hopkins Press.

Mayfield, E.C. (1964) The selection interview: A re-evaluation of published research. *Personnel Psychology*, 17, 239-260.

McCor mk, E. (1979) Job analysis: Methods and Applications. New York: AMA-COM.

Mitchell, J.V., Jr. (ed.), (1983) *Tests in Print III (TIP III): An Index to Tests, Test Reviews, and the Literature on Specific Tests*. Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska Press

Mitchell, J.V. Jr. (ed.) (1986) *The Ninth Mental Measurement Yearbook*. Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska Press

Owens, W.A. (1976) Background data. In M.D. Dunnette (ed) Handbook of Industrial and Organizational Psychology, Chicago: Rand McNally.

Reilly, R.R. & G.T. Chao (1982) Validity and fairness of some alternate employee selection procedures. *Personnel Psychology* 35, 1-62.

Reynolds, W.B. (Nov 10, 1986) Memorandum to the Director of the U.S. Employment service. Cited in Wigdor, A.K. and J.A. Hartigan, eds., (1988) Interim Report: Within Group Scoring of the general Aptitude Test Battery, Washington, DC: National Academy Press.

Sackett, P.R., Schmidt, N., Tenopyr, M.L., Kehoe, J., & S. Zedeck (1985) Commentary on forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697-798.

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

Schmidt, F.L., Hunter, J.E., & K. Pearlman (1981) Tasl differences as moderators of aptitude test validity in selection: A red herring, *Journal of Applied Psychology*, 66, 2, 166-185.

Schmidt, F.L., Hunter, J.E., & K. Pearlman (1982) Assessing the economic impact of personnel programs on work force productivity, 35, 333-347.

Seymour, R.T. (1988) Why Plaintiffs' Counsel Challenge Tests, and How They Can Successfully Challenge the Theory of "Validity Generalization", *Journal of Vocational Behavior*, 33, 331-364.

Sharf, J.C. (1988) Litigating Personnel Measurement Policy, *Journal of Vocational Behavior*, 33, 235-271.

Society for Industrial and Organizational Psychology (1987) Principles for the validation and use of personnel selection procedures, Third edition, College Park, MD: author.

Sweetland, R. C. and D.J. Keyser, eds., (1986) *Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business (2nd ed.)*. Kansas City, MO: Test Corporation of America

Taylor, H.C. & J.T. Russell (1939) The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and Tables. *Journal of Applied Psychology*, 23, 656-578.

Wagner, R. (1949) The employment interview: A critical review. *Personnel Psychology*, 2, 17-46.

Wigdor, A.K. and J.A. Hartigan, eds., (1988) Interim Report: Within Group Scoring of the general Aptitude Test Battery, Washington, DC: National Academy Press.

Wonderlic & Associates (1982), Validity of the Wonderlic Personnel Test, author.