TESTING AND ASSESSMENT:

AN EMPLOYER'S GUIDE

TO GOOD PRACTICES

Compliments of Success Performance Solutions and Ira S. Wolfe
For more information about assessments of choice for matching,managing and motivating employees, call or visit our website at:

800.803.4303 or 717.291.4640 or 410.941.2345

www.successperformancesolutions.com

U.S. Department of Labor
Employment and Training Administration
2000

Foreword

PURPOSE of the GUIDE

In today's competitive marketplace and complex legal environment, employers face the challenge of attracting, developing, and retaining the best employees. Michael Eisner, CEO of the Disney Corporation, recognized the impact of personnel decisions on a business' bottom-line when he remarked, "My inventory goes home every night."

This Guide is designed to help managers and human resource (HR) professionals use assessment practices to reach their organizations' HR goals. It conveys the essential concepts of employment testing in easy-to-understand terms so that managers and HR professionals can:

- Evaluate and select assessment tools/procedures that maximize chances for getting the right fit between jobs and employees.
- Administer and score assessment tools that are the most efficient and effective for their particular needs.
- Accurately interpret assessment results.
- Understand the professional and legal standards to be followed when conducting personnel assessment.

FORMAT of the GUIDE

This Guide is structured around a set of assessment principles and their applications. The information is organized so that readers from a variety of backgrounds will find it to be clear and useful.

- Each chapter covers a critical aspect of the assessment process. The issues involved in each aspect are outlined at the beginning of each chapter.
- Thirteen principles of assessment are explained in the Guide. The last chapter (Chapter 9) serves as a review by summarizing the main points of the thirteen principles.
- To assist readers in finding additional information, links to relevant websites are imbedded in the text throughout the Guide.
- In addition, Appendix A offers a list of resource materials for those interested in more information on a particular topic.
- Appendix B is a glossary for quick clarification of terms and concepts.

The *Guide* is designed to provide accurate and important information regarding testing as part of a personnel assessment program. It gives general guidelines and must not be viewed as legal advice.

Acknowledgments

Testing and Assessment: An Employer's Guide to Good Practices (Guide) was produced and funded by the Skills Assessment and Analysis Program in the U.S. Department of Labor, Employment and Training Administration, Office of Policy and Research (OPR) under the direction of Gerard F. Fiala, Administrator. The Skills Assessment and Analysis Program is directed by Donna Dye, Personnel Research Psychologist, who provided technical direction and support for this Guide.

The Guide was prepared under Department of Labor grants with the North Carolina Employment Security Commission, Southern Assessment Research and Development Center and National O*NET Consortium; the New York Department of Labor; and the Utah Department of Employment Security. The Guide was completed under the direction of David Rivkin and Phil Lewis. Mr. Rivkin also served as editor of the Guide. Authors of this Guide were Syed Saad, Gary W. Carter, Mark Rothenberg, and Enid Israelson. Identification and incorporation of the website links was conducted by Dr. James Rounds and Thomas Smith of the University of Illinois at Urbana-Champaign. Grateful acknowledgment is made to Laurie Koritko, Patrice Gilliam-Johnson, Jonathan Levine, and Brenda Dunn for their contribution. Thanks are also given to Ann Kump, Helen Tannenbaum, Don Kreger, Kristin Fiske, and Marilyn Silver whose valuable suggestions were very much appreciated. Grateful acknowledgment is also made to Suzan Chastain, Department of Labor, Office of the Solicitor, Division of Civil Rights, and Hilary R. Weinerand and Cynthia Misicka of the Equal Employment Opportunity Commission for consultant review and insights into the final preparation of this Guide.

Special Notice

This material will be made available to sensory impaired individuals upon request.
Voice phone: (202) 219-7664
TDD* phone: (800) 326-2577

*Telecommunications Device for the Deaf

Table of Contents

CHAPTER 1 Personnel Assessment

Personnel assessment is a systematic approach to gathering information about individuals. This information is used to make employment or career-related decisions about applicants and employees.

Information regarding organizations involved with personnel assessment can be found at the International Personnel Management Association (IPMA) sites: www.ipmaac.org/ and www.ipma-hr.org/

Assessment is conducted for some specific purpose. For example, you, as an employer, may conduct personnel assessment to select employees for a job. Career counselors may conduct personnel assessment to provide career guidance to clients.

Information and links to various career guidance instruments are available in the Resource Library of America's Career InfoNet at: http://www.acinet.org/

Chapter Highlights

1. Personnel assessment tools: tests and procedures
2. Relationship between the personnel assessment process and tests and procedures
3. What do tests measure?
4. Why do organizations conduct assessment?
5. Some situations in which an organization may benefit from testing
6. Importance of using tests in a purposeful manner
7. Limitations of personnel tests and procedures—*fallibility* of test scores

Principles of Assessment Discussed

Use assessment tools in a *purposeful manner*
Use the *whole-person approach* to assessment.

1. Personnel assessment tools: tests and procedures

Any test or procedure used to measure an individual's employment or career-related qualifications and interests can be considered a personnel assessment tool. There are many types of personnel assessment tools. These include *traditional knowledge and ability tests, inventories, subjective procedures*, and *projective instruments.* In this Guide, the term *test* will be used as a generic term to refer to any instrument or procedure that measures samples of behavior or performance.

Personnel assessment tools differ in:

- **Purpose,** e.g., selection, placement, promotion, career counseling, or training
- **What they are designed to measure,** e.g., abilities, skills, work styles, work values, or vocational interests
- **What they are designed to predict,** e.g., job performance, managerial potential, career success, job satisfaction, or tenure
- **Format,** e.g., paper-and-pencil, work-sample, or computer simulation
- **Level of standardization, objectivity, and quantifiability**—Assessment tools and procedures vary greatly on these factors. For example, there are subjective evaluations of resumes, highly structured achievement tests, interviews having varying degrees of structure, and personality inventories with no specific right or wrong answers.

All assessment tools used to make employment decisions, regardless of their format, level of standardization, or objectivity, are subject to professional and legal standards. For example, both the evaluation of a resume and the use of a highly standardized achievement test must comply with applicable laws. Assessment tools used solely for career exploration or counseling are usually **not** held to the same legal standards.

2. Relationship between the personnel assessment process and tests and procedures

A personnel test or a procedure provides only *part of the picture* about a person. On the other hand, the personnel assessment process combines and evaluates all the information gathered about a person to make career or employment-related decisions. Figure 1 on page 1-3 highlights the relationship between assessment tools and the personnel assessment process.

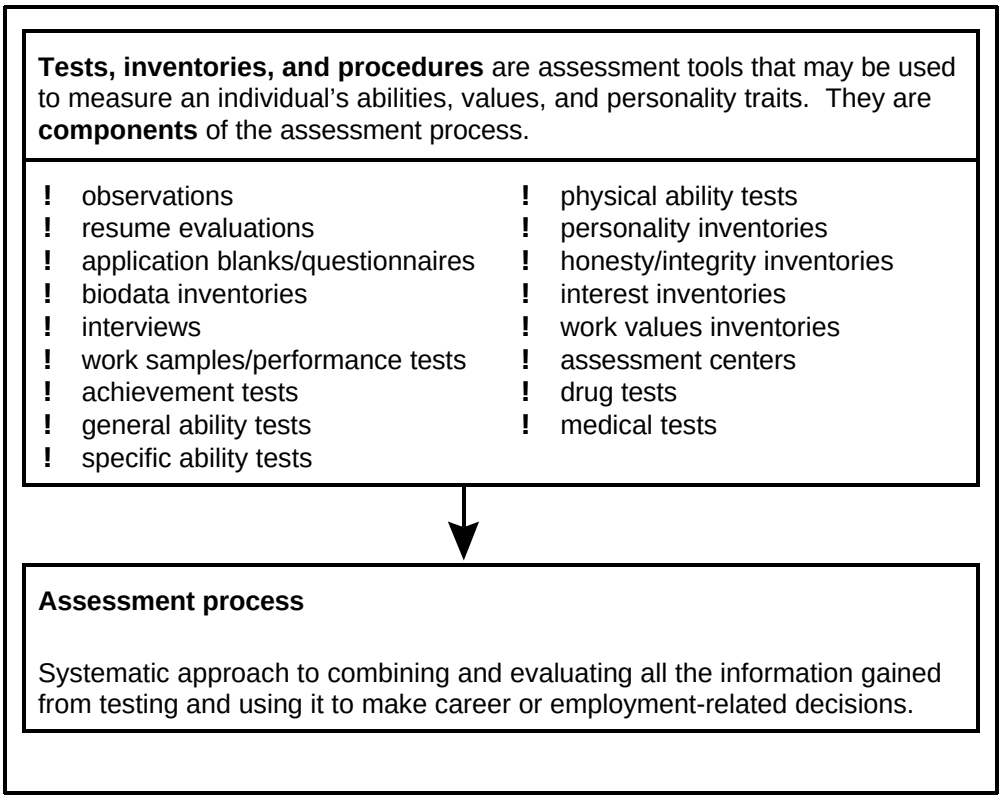

Figure 1. Relationship between assessment tools and the assessment process.

3. What do tests measure?

People differ on many psychological and physical characteristics. In testing, these characteristics are called *constructs.* For example, people skillful in verbal and mathematical reasoning are considered high on the construct *mental ability.* Those who have little physical stamina and strength are labeled low on the constructs *endurance* and *physical strength. Constructs* can be used to identify personal characteristics and to sort people in terms of these characteristics.

Constructs cannot be seen or heard, but we can observe their effects on other variables. For example, we don't observe physical strength but we can observe people with great strength lifting heavy objects and people with limited strength attempting, but failing, to lift these objects. Such differences in characteristics among people have important implications in the employment context.

Employees and applicants vary widely in their knowledge, skills, abilities, interests, work styles, and other characteristics. These differences systematically affect the way people perform or behave on the job.

These differences in characteristics are not necessarily apparent by simply observing the employee or job applicant. Employment tests can be used to gather accurate information about job-relevant characteristics. This information helps assess the fit or match between people and jobs. For example, an applicant's score on a mechanical test reflects his or her mechanical ability as measured by the test. This score can be used to predict how well that applicant is likely to perform in a job that requires mechanical ability, as demonstrated through a professionally conducted job analysis. Tests can be used in this way to identify potentially good workers.

Some tests can be used to *predict* employee and applicant job performance. In testing terms, whatever the test is designed to predict is called the *criterion.* A *criterion* can be any measure of work behavior or any outcome that can be used as the standard for successful job performance. Some commonly used *criteria* are productivity, supervisory ratings of job performance, success in training, tenure, and absenteeism. For example, in measuring job performance, *supervisory ratings* could be the *criterion* predicted by a test of mechanical ability. How well a test predicts a criterion is one indication of the usefulness of the test.

4. Why do organizations conduct assessment?

Organizations use assessment tools and procedures to help them perform the following human resource functions:

- **Selection.** Organizations want to be able to identify and hire, fairly and efficiently, the best people for the job and the organization. A properly developed and applied assessment tool may provide a way to select successful sales people, concerned customer service representatives, and effective workers in many other occupations.
- **Placement.** Organizations also want to be able to assign people to the appropriate job level. For example, an organization may have several managerial positions, each having a different level of responsibility. Assessment may provide information that helps organizations achieve the best fit between employees and jobs.
- **Training and development.** Tests are used to find out whether employees have mastered training materials. They can help identify those applicants and employees who might benefit from either remedial or advanced training. Information gained from testing can be used to design or modify training programs. Test results also help individuals identify areas in which self-development activities would be useful.

Additional information on workforce training and development can be found at the Employment and Training Administration's Workforce Development site: www.doleta.gov/employer/wd.htm

- **Promotion.** Organizations may use tests to identify employees who possess managerial potential or higher level capabilities, so that these employees can be promoted to assume greater duties and responsibilities.
- **Career exploration and guidance.** Tests are sometimes used to help people make educational and vocational choices. Tests may provide information that helps individuals choose occupations in which they are likely to be successful and satisfied.

The Occupational Information Network (or O*NET; www.doleta.gov/programs/onet/ and www.onetcenter.org/) is the nation's primary source for occupational exploration and information. O*NET provides comprehensive information on job requirements and worker competencies for over 1100 occupations. This Guide (*Testing and Assessment: An Employer's Guide to Good Practices*), as well as other guides in this series are available for downloading from the above Web sites. A variety of career exploration links and workforce development initiatives are also accessible.

- **Program evaluation.** Tests may provide information that the organization can use to determine whether employees are benefiting from training and development programs.

The American Evaluation Association (www.eval.org) is an international professional association concerned with various types of evaluation, including program evaluation.

5. Some situations in which an organization may benefit from testing

Some examples of these situations include the following:

- Current selection or placement procedures result in poor hiring decisions.
- Employee productivity is low.
- Employee errors have serious financial, health, or safety consequences.
- There is high employee turnover or absenteeism.
- Present assessment procedures do not meet current legal and professional standards.

6. Importance of using tests in a purposeful manner

Assessment instruments, like other tools, can be extremely helpful when used properly, but counterproductive when used inappropriately. Often inappropriate use stems from not having a clear understanding of what you want to measure and why you want to measure it. Having a clear understanding of the purpose of your assessment system is important in selecting the appropriate assessment tools to meet that purpose. This brings us to an important principle of assessment.

Principle of Assessment

Use assessment tools in a *purposeful manner*. It is critical to have a clear understanding of what needs to be measured and for what purpose.

Assessment strategies should be developed with a clear understanding of the knowledge, skills, abilities, characteristics, or personal traits you want to measure. It is also essential to know what each assessment tool you are considering using is designed to measure.

7. Limitations of personnel tests and procedures—fallibility of test scores

Professionally developed tests and procedures that are used as part of a planned assessment program may help you select and hire more qualified and productive employees. However, it is essential to understand that *all assessment tools are subject to errors*, both in measuring a characteristic, such as verbal ability, and in predicting performance criteria, such as success on the job. This is true for all tests and procedures, regardless of how objective or standardized they might be.

- **Do not** expect any test or procedure to measure a personal trait or ability with perfect accuracy for every single person.
- **Do not** expect any test or procedure to be completely accurate in predicting performance.

There will be cases in which a test score or procedure will predict someone to be a good worker, who, in fact, is not. There will also be cases in which an individual receiving a low score will be rejected, when he or she would actually be a capable and good worker. Such errors in the assessment context are called *selection errors*. Selection errors cannot be completely avoided in any assessment program.

Why do organizations conduct testing despite these errors? The answer is that appropriate use of professionally developed assessment tools, on average, enables organizations to make more effective employment-related decisions than does the use of simple observations or random decision making.

Using a single test or procedure will provide you with a limited view of a person's employment or career-related qualifications. Moreover, you may reach a mistaken conclusion by giving too much weight to a single test result. On the other hand, using a variety of assessment tools enables you to get a more complete picture of the individual. The practice of using a variety of tests and procedures to more fully assess people is referred to as the *whole-person approach* to personnel assessment. This approach will help reduce the number of selection errors and boost the effectiveness of your decision making. This leads to an important principle of assessment.

Principle of Assessment

> Do not rely too much on any one test to make decisions. Use the *whole-person approach* to assessment.

CHAPTER 2 Understanding the Legal Context of Assessment—Employment Laws and Regulations with Implications for Assessment

The number of laws and regulations governing the employment process has increased over the past four decades. Many of these laws and regulations have important implications for conducting employment assessment. This chapter discusses what you should do to make your practices consistent with legal, professional, and ethical standards.

Chapter Highlights

1. Title VII of the Civil Rights Act (CRA) of 1964, as amended in 1972; Tower Amendment to Title VII
2. Age Discrimination in Employment Act of 1967 (ADEA)
3. Equal Employment Opportunity Commission (EEOC) - 1964
4. *Uniform Guidelines on Employee Selection Procedures* - 1978; adverse or disparate impact, approaches to determine existence of adverse impact, four-fifths rule, job-relatedness, business necessity, biased assessment procedures
5. Title I of the Civil Rights Act (CRA) of 1991
6. Americans with Disabilities Act (ADA) - 1990
7. Record keeping of adverse impact and job-relatedness of tests
8. The Standards for Educational and Psychological Testing[1] - 1985; The Principles for the Validation and Use of Personnel Selection Procedures - 1987
9. Relationship between federal, state, and local employment laws

Principles of Assessment Discussed

> Use only assessment instruments that are *unbiased* and *fair* to all groups.

The general purpose of employment laws and regulations is to prohibit *unfair discrimination* in employment and provide equal employment opportunity for all. Unfair discrimination occurs when employment decisions are based on race, sex, religion, ethnicity, age, or disability rather than on job-relevant knowledge, skills, abilities, and other characteristics. Employment practices that unfairly discriminate against people are called *unlawful* or *discriminatory employment practices*.

[1]Currently under revision by the American Psychological Association.

The summaries of the laws and regulations in this chapter focus on their impact on employment testing and assessment. Before you institute any policies based on these laws and regulations, read the specific laws carefully, and consult with your legal advisors regarding the implications for your particular assessment program.

1. Title VII of the Civil Rights Act (CRA) of 1964 (as amended in 1972); Tower Amendment to Title VII

Title VII is landmark legislation that prohibits unfair discrimination in all terms and conditions of employment based on race, color, religion, sex, or national origin. Other subsequent legislation, for example, ADEA and ADA, has added age and disability, respectively, to this list. Women and men, people age 40 and older, people with disabilities, and people belonging to a racial, religious, or ethnic groups are protected under Title VII and other employment laws. Individuals in these categories are referred to as members of a **protected group.** The employment practices covered by this law include the following:

- recruitment
- hiring
- job classification
- transfer
- training
- promotion
- performance appraisal
- compensation
- union or other membership
- disciplinary action
- termination
- fringe benefits

Employers having 15 or more employees, employment agencies, and labor unions are subject to this law.

The *Tower Amendment* to this act stipulates that professionally developed workplace tests can be used to make employment decisions. However, only instruments that do not discriminate against any protected group can be used. Use only tests developed by experts who have demonstrated qualifications in this area.

Title VII of the CRA of 1964, as amended, can be found at
www.dol.gov/dol/oasam/public/regs/statutes/2000e-16.htm

2. Age Discrimination in Employment Act of 1967 (ADEA)

This act prohibits discrimination against employees or applicants age 40 or older in all aspects of the employment process. Individuals in this group must be provided equal employment opportunity; discrimination in testing and assessment is prohibited. If an older worker charges discrimination under the ADEA, the employer may defend the practice if it can be shown that the job requirement of age is a matter of *business necessity*. Employers must have documented support for the argument they use as a defense.

ADEA covers employers having 20 or more employees, employment agencies, and labor unions. Certain groups of employees are exempt from ADEA coverage, including public law enforcement

personnel, such as police officers and firefighters. Uniformed military personnel also are exempt from ADEA coverage.

The Age Discrimination in Employment Act is available online at the U.S. Equal Employment Opportunity Commission (EEOC) site: www.eeoc.gov/laws/adea.html

3. Equal Employment Opportunity Commission (EEOC)—1964

The EEOC is responsible for enforcing federal laws prohibiting employment discrimination, including Title VII, the ADEA, and the ADA. It receives, investigates, and processes charges of unlawful employment practices of employers filed by an individual, a group of individuals, or an EEOC commissioner. If the EEOC determines that there is "reasonable cause" that an unlawful employment practice has occurred, it is also authorized to sue on behalf of the charging individual(s) or itself. The EEOC participated in developing the *Uniform Guidelines on Employee Selection Procedures*.

As indicated earlier, the EEOC maintains a web site at www.eeoc.gov/ It includes general information, as well as specific information on laws, regulations and policy guidance, and is directed towards both employers and employees.

4. *Uniform Guidelines on Employee Selection Procedures*—1978; adverse or disparate impact, approaches to determine existence of adverse impact, four-fifths rule, job-relatedness, business necessity, biased assessment procedures

In 1978, the EEOC and three other federal agencies—the Civil Service Commission (predecessor of the Office of Personnel Management) and the Labor and Justice Departments—jointly issued the *Uniform Guidelines on Employee Selection Procedures*. The *Guidelines* incorporate a set of principles governing the use of employee selection procedures according to applicable laws. They provide a framework for employers and other organizations for determining the proper use of tests and other selection procedures. The *Guidelines* are legally binding under a number of civil rights laws, including Executive Order 11246 and the Civil Rights Requirements of the National Job Training Partnership Act and the Wagner Peyser Act. In reviewing the testing practices of organizations under Title VII, the courts generally give great importance to the *Guidelines*' technical standards for establishing the job-relatedness of tests. Also, federal and state agencies, including the EEOC, apply the *Uniform Guidelines* in enforcing Title VII and related laws.

The *Guidelines* cover all employers employing 15 or more employees, labor organizations, and employment agencies. They also cover contractors and subcontractors to the federal government and organizations receiving federal assistance. They apply to all tests, inventories and procedures used to make employment decisions. Employment decisions include hiring, promotion, referral, disciplinary action, termination, licensing, and certification. Training may be included as an employment decision if it leads to any of the actions listed above. The *Guidelines* have significant implications for personnel assessment.

The *Uniform Guidelines* are available on-line at www.acd.ccac.edu/hr/DOL/60_3_toc.htm, as are the Civil Rights Laws under which they are bound:
www.dol.gov/dol/esa/public/regs/statutes/ofccp/eo11246.htm and
www.wdsc.org/dinap/html/jtpa-toc-title.html

One of the basic principles of the *Uniform Guidelines* is that it is unlawful to use a test or selection procedure that creates adverse impact, unless justified. Adverse impact occurs when there is a substantially different rate of selection in hiring, promotion, or other employment decisions that work to the disadvantage of members of a race, sex, or ethnic group.

Different approaches exist that can be used to determine whether adverse impact has occurred. Statistical techniques may provide information regarding whether or not the use of a test results in adverse impact. Adverse impact is normally indicated when the selection rate for one group is less than 80% (4/5) that of another. This measure is commonly referred to as the four-fifths or 80% rule. However, variations in sample size may affect the interpretation of the calculation. For example, the 80% rule may not be accurate in detecting substantially different rates of selection in very large or small samples. When determining whether there is adverse impact in very large or small samples, more sensitive tests of statistical significance should be employed.

When there is no charge of adverse impact, the *Guidelines* do not require that you show the job-relatedness of your assessment procedures. However, you are strongly encouraged to use only job-related assessment tools.

If your assessment process results in adverse impact, you are required to eliminate it or justify its continued use. The *Guidelines* recommend the following actions when adverse impact occurs:

- Modify the assessment instrument or procedure causing adverse impact.
- Exclude the component procedure causing adverse impact from your assessment program.
- Use an alternative procedure that causes little or no adverse impact, assuming that the alternative procedure is substantially equally valid.
- Use the selection instrument that has adverse impact if the procedure is job-related and valid for selecting better workers, and there is no equally effective procedure available that has less adverse impact.

Note that for the continued use of assessment instruments or procedures that cause adverse impact, courts have required justification by business necessity as well as validity for the specific use. The issue of business necessity is specifically addressed in Title I of the Civil Rights Act of 1991 (see next section).

Additional information regarding the issue of business necessity may be found at
http://www.smallbiz.findlaw.com/text/

An assessment procedure that causes adverse impact may continue to be used only if there is evidence that:

- It is job-related for the position in question.
- Its continued use is justified by business necessity.

Demonstrating job-relatedness of a test is the same as establishing that the test may be validly used as desired. Chapter 3 discusses the concept of test validity and methods for establishing the validity or job-relatedness of a test.

Demonstrating the business necessity of using a particular assessment instrument involves showing that its use is essential to the safe and efficient operation of the business and there are no alternative procedures available that are substantially equally valid to achieve the business objectives with a lesser adverse impact.

Specific discussion of adverse impact within the *Uniform Guidelines* can be found at www2.dol.gov/dol/esa/public/regs/cfr/41cfr/toc_Chapt60/60_3.4.htm

Another issue of importance discussed in the *Uniform Guidelines* relates to test fairness. The *Uniform Guidelines* define **biased** or **unfair assessment procedures** as those assessment procedures on which one race, sex, or ethnic group characteristically obtains lower scores than members of another group and the differences in the scores are not reflected in differences in the job performance of members of the groups.

The meaning of scores on an unfair or biased assessment procedure will differ depending on the group membership of the person taking the test. Therefore, using biased tests can prevent employers from making equitable employment decisions. This leads to the next principle.

Principle of Assessment

> Use only assessment instruments that are *unbiased* and *fair to* all groups.

Use of biased tools may result in unfair discrimination against members of the lower scoring groups. However, use of fair and unbiased tests can still result in adverse impact in some cases. If you are developing your own test or procedure, expert help may be advisable to make sure your procedure is fair to all relevant groups. If you are planning to purchase professionally developed assessment tools, first evaluate the fairness of those you are considering by reading the test manuals and consulting independent reviews.

The National Library of Education has developed an Educational Resources Information Center (ERIC). The ERIC Clearinghouse on Assessment and Evaluation provides links to a variety of sources concerned with test fairness. Locate the link “Fairness in Testing” at www.ericae.net/

5. Title I of the Civil Rights Act of 1991

Title I of the CRA of 1991 reaffirms the principles developed in Title VII of the CRA of 1964, but makes several significant changes.

As noted previously, the Act specifically requires demonstration of both the job-relatedness and business necessity of assessment instruments or procedures that cause adverse impact. The business necessity requirement, set forth in Title I of the CRA of 1991, is harder to satisfy in defending challenged practices than a business purpose test suggested by the Supreme Court earlier.

Another important provision relates to the use of group-based test score adjustments to maintain a representative work force. The CRA prohibits score adjustments, the use of different cut-off scores for different groups of test takers, or alteration of employment-related test results based on the demographics of the test takers. Such practices, which are referred to as *race norming* or *within-group norming*, were used by some employers and government agencies in the past to avoid adverse impact.

The CRA also makes compensatory and punitive damages available as a remedy for claims of intentional discrimination under Title VII and the ADA.

Title I of the Civil Rights Act is available online at www.eeoc.gov/laws/cra91.html

6. Americans with Disabilities Act (ADA) - 1990

Under the ADA, qualified individuals with disabilities must be given equal opportunity in all aspects of employment. The law prohibits employers with 15 or more employees, labor unions, and employment agencies from discriminating against qualified individuals with disabilities. Prohibited discrimination includes failure to provide reasonable accommodation to persons with disabilities when doing so would not pose undue hardship to the organization.

A qualified individual with a disability is one who can perform the essential functions of a job, with or without reasonable accommodation. In other words, an individual with a disability, who is able to perform the essential functions of a job, is considered qualified, even if the employer has to make reasonable accommodation to enable the individual to perform the job.

- *Disability* is defined broadly to include any physical or mental impairment that substantially limits one or more of an individual's major life activities, such as caring for oneself, walking, talking, hearing, or seeing. Some common examples include visual, speech, and hearing disabilities; epilepsy; specific learning disabilities; cancer; serious mental illness; AIDS and HIV infection; alcoholism; and past drug addiction. Noteworthy among conditions **not** covered are current illegal use of drugs, sexual behavior disorders, compulsive gambling, kleptomania, and pyromania.

- *Essential functions* are the primary job duties that are fundamental, and not marginal to the job. Factors relevant to determining whether a function is essential include written job descriptions, the amount of time spent performing the function, the consequences of not requiring the function, and the work experiences of employees who hold the same or similar jobs.
- *Reasonable accommodation* is defined as a change in the job application and selection process, a change in the work environment or the manner in which the work is performed, that enables a qualified person with a disability to enjoy equal employment opportunities. Under the CRA, qualified individuals with disabilities must be provided reasonable accommodation so they can perform the essential job functions, as long as this does not create undue hardship to the employer.
- *Undue hardship* is defined as significant difficulty or additional expense and is determined based on a number of factors. Some factors that are considered are the nature and net cost of the accommodation, the financial resources of the facility, the number employed at the facility, the effect on resources and operations, the overall financial resources of the entire organization, and the fiscal relationship of the facility with the organization. An accommodation that is possible for a large organization may pose an undue hardship for a small organization.

The ADA has major implications for your assessment practices:

- In general, it is the responsibility of the individual with a disability to inform you that an accommodation is needed. However, you may ask for advance notice of accommodations required, for the hiring process only, so that you may adjust your testing program or facilities appropriately. When the need for accommodation is not obvious, you may request reasonable documentation of the applicant's disability and functional limitations for which he or she needs an accommodation.
- Reasonable accommodation may involve making the test site accessible, or using an alternative assessment procedure. Requiring individuals with disabilities to use their impaired abilities for an employment test is prohibited unless the test is intended to measure one of these abilities. For example, under the ADA, when a test screens out one or more individuals with a disability, its use must be shown to be job-related for the position in question and justified by business necessity.
- One possible alternative procedure, if available, would be to use a form of the test that does not require use of the impaired ability. Another possibility is to use a procedure that compensates for the impaired ability, if appropriate. For example, allowing extra time to complete certain types of employment tests for someone with dyslexia or another learning disability, or providing a test with larger print or supplying a reader to a visually impaired individual where appropriate, would be considered reasonable accommodation.
- The ADA expressly prohibits making medical inquiries or administering medical examinations prior to making a job offer. Before making medical inquiries, or requiring medical exams, you must make an offer of employment to the applicant. You may make medical inquiries or require medical exams of an employee only when doing so is work-related and justified by business necessity. All medical information you obtain about your applicants and employees is

strictly confidential and must be treated as such. Access to and use of this information is also greatly restricted. For a more detailed discussion of medical examinations see Chapter 4.

Your organization should develop a written policy on conducting testing and assessment of individuals with disabilities. This will help ensure compliance with the provisions of the ADA.

If you need assistance in complying with the ADA, there are several resources you may contact:

- The Job Accommodation Network: (800) 526-7234

  Web site: janweb.icdi.wvu.edu/

- Industry-Labor Council on Employment and Disability: (516) 747-6323

- The American Foundation for the Blind: (202) 408-0200, (800) 232-5463

  Web site: www.afb.org/

- The President's Committee on Employment of People with Disabilities: (202) 376-6200

  Web site: www50.pcepd.gov/pcepd/ztextver/

- Disability and Business Technical Assistance Centers: (800) 949-4232

  Web site: www.adata.org/

- EEOC Enforcement Guidance (clarifies the rights and responsibilities of employers and individuals with disabilities regarding reasonable accommodation and undue hardship):

  Web site: www.eeoc.gov/docs/accommodation.html

The Americans with Disabilities Document Center web site (janweb.icdi.wvu.edu/kinder/) includes the ADA statute, regulations, ADAAG (Americans with Disabilities Act Accessibility Guidelines), federally reviewed tech sheets, and other assistance documents.

Additional sources of information relating to the ADA and accommodations include the following:

- ABLEDATA- Sponsored by the National Institute on Disability and Rehabilitation Research, U.S. Department of Education. ABLEDATA's primary mission is to provide information on assistive technology and rehabilitation equipment.

  Web site: www.abledata.com/

- Trace Research and Development Center at the University of Wisconsin. Focuses on making off-the-shelf technologies and systems more accessible for everyone.

  Web site: www.trace.wisc.edu/about/

- National Business and Disability Council (NBDC) is a leading national corporate resource on issues related to the successful employment and integration of individuals with disabilities into America's workforce.

  Web site: www.business-disability.com/home.asp

- The Access Board, United States Architectural and Transportation Barriers Compliance Board- The federal agency that develops the minimum guidelines and requirements for standards issued under the ADA.

  Web site: www.access-board.gov/adaag/html/adaag.htm

7. Record keeping of adverse impact and job-relatedness of tests

The *Uniform Guidelines* and subsequent regulations[2] require that all employers maintain a record of their employment-related activities, including statistics related to testing and adverse impact. Filing and record-keeping requirements for large employers (those with over 100 employees) are generally more extensive than those for employers with 100 or fewer employees. To learn more about the specific requirements, refer to EEOC regulations on record-keeping and reporting requirements under Title VII, and the ADA, 29 CFR part 1602, and the *Uniform Guidelines*.

8. The Standards for Educational and Psychological Testing - 1985; The Principles for the Validation and Use of Personnel Selection Procedures—1987

There are two resource guides published by major organizations in the testing field that will help you set up and maintain an assessment program. The principles and practices presented in these publications set the standards for professional conduct in all aspects of assessment.

- *The Standards for Educational and Psychological Testing*. This publication was developed jointly by the American Psychological Association (APA; www.apa.org), the National Council on Measurement in Education (NCME; ncme.ed.uiuc.edu), and the American Educational Research Association (AERA; www.aera.net). The *Standards* are an authoritative and comprehensive source of information on how to develop, evaluate, and use tests and other assessment procedures in educational, employment, counseling, and clinical settings. Although developed as professional guidelines, they are consistent with applicable regulations and are frequently cited in litigation involving testing practices.
- *The Principles for the Validation and Use of Personnel Selection Procedures*. This publication was developed by the Society for Industrial and Organizational Psychology (SIOP; www.siop.org). Like the *Standards*, the *Principles* are also an excellent guide to good practices in the choice, development, evaluation, and use of assessment tools. However, their main focus is on tools used in the personnel assessment context. The *Principles* explain their relationship to the *Standards* in the following way:

[2] 29 CFR part 1602, as amended by 56 Fed. Reg. 35,753 (July 26, 1991); previously, record- keeping requirements did not apply to temporary and seasonal positions.

> The *Standards* primarily address psychometric issues while the *Principles* primarily address the problems of making decisions in employee selection, placement, promotion, etc. The major concern of the *Standards* is general; the primary concern of the *Principles* is that performance on a test . . . is related to performance on a job or other measures of job success.

Compatibility of the Standards and the Principles with the Uniform Guidelines
The *Uniform Guidelines* were intended to be consistent with generally accepted professional standards for validating and evaluating standardized tests and other selection procedures. In this regard, the *Guidelines* specifically refer to the *Standards.*

It is strongly encouraged that you develop familiarity with both the *Standards* and the *Principles* in addition to the *Uniform Guidelines*. Together, they can help you conduct personnel assessment in a manner consistent with legal and professional standards.

The *Standards* may be ordered from the APA web site (www.apa.org/books/standard.html) and the *Principles* is available for purchase at webm8426.ntx.net/order.html

9. Relationship between federal, state, and local employment laws

Some states and localities have issued their own fair employment practices laws, and some have adopted the federal *Uniform Guidelines*. These state and local laws may be more stringent than corresponding federal laws. When there is a contradiction, federal laws and regulations override any contradictory provisions of corresponding state or local laws. You should become thoroughly familiar with your own state and local laws on employment and testing before you initiate and operate an assessment program.

CHAPTER 3 Understanding Test Quality—Concepts of Reliability and Validity

Test *reliability* and *validity* are two technical properties of a test that indicate the quality and usefulness of the test. These are the two most important features of a test. You should examine these features when evaluating the suitability of the test for your use. This chapter provides a simplified explanation of these two complex ideas. These explanations will help you to understand reliability and validity information reported in test manuals and reviews and use that information to evaluate the suitability of a test for your use.

Chapter Highlights

1. What makes a good test?
2. Test reliability
3. Interpretation of reliability information from test manuals and reviews
4. Types of reliability estimates
5. Standard error of measurement
6. Test validity
7. Methods for conducting validation studies
8. Using validity evidence from outside studies
9. How to interpret validity information from test manuals and independent reviews

Principles of Assessment Discussed

Use only *reliable* assessment instruments and procedures.

Use only assessment procedures and instruments that have been demonstrated to be valid for the specific purpose for which they are being used.

Use assessment tools that are appropriate for the target population.

1. What makes a good test?

An employment test is considered "good" if the following can be said about it:

- The test measures what it claims to measure consistently or reliably. This means that if a person were to take the test again, the person would get a *similar* test score.
- The test measures what it claims to measure. For example, a test of mental ability does in fact measure mental ability, and not some other characteristic.
- The test is job-relevant. In other words, the test measures one or more characteristics that are important to the job.
- By using the test, more effective employment decisions can be made about individuals. For example, an arithmetic test may help you to select qualified workers for a job that requires knowledge of arithmetic operations.

The degree to which a test has these qualities is indicated by two technical properties: *reliability* and *validity*.

2. Test reliability

Reliability refers to how dependably or consistently a test measures a characteristic. If a person takes the test again, will he or she get a similar test score, or a much different score? A test that yields similar scores for a person who repeats the test is said to measure a characteristic reliably.

How do we account for an individual who does not get exactly the same test score every time he or she takes the test? Some possible reasons are the following:

- **Test taker's temporary psychological or physical state.** Test performance can be influenced by a person's psychological or physical state at the time of testing. For example, differing levels of anxiety, fatigue, or motivation may affect the applicant's test results.
- **Environmental factors.** Differences in the testing environment, such as room temperature, lighting, noise, or even the test administrator, can influence an individual's test performance.
- **Test form.** Many tests have more than one version or form. Items differ on each form, but each form is supposed to measure the same thing. Different forms of a test are known as ***parallel forms*** or ***alternate forms***. These forms are designed to have similar measurement characteristics, but they contain different items. Because the forms are not exactly the same, a test taker might do better on one form than on another.
- **Multiple raters.** In certain tests, scoring is determined by a rater's judgments of the test taker's performance or responses. Differences in training, experience, and frame of reference among raters can produce different test scores for the test taker.

These factors are sources of chance or random measurement error in the assessment process. If there were no random errors of measurement, the individual would get the same test score, the

individual's "true" score, each time. The degree to which test scores are unaffected by measurement errors is an indication of the **reliability of the test**.

Reliable assessment tools produce dependable, repeatable, and consistent information about people. In order to meaningfully interpret test scores and make useful employment or career-related decisions, you need reliable tools. This brings us to the next principle of assessment.

Principle of Assessment

> Use only *reliable* assessment instruments and procedures. In other words, use only assessment tools that provide dependable and consistent information.

3. Interpretation of reliability information from test manuals and reviews

Test manuals and independent review of tests provide information on test reliability. The following discussion will help you interpret the reliability information about any test.

The reliability of a test is indicated by the ***reliability coefficient***. It is denoted by the letter "r," and is expressed as a number ranging between 0 and 1.00, with r = 0 indicating no reliability, and r = 1.00 indicating perfect reliability. Do not expect to find a test with perfect reliability. Generally, you will see the reliability of a test as a decimal, for example, r = .80 or r = .93. The larger the reliability coefficient, the more repeatable or reliable the test scores. Table 1 serves as a general guide-line for interpreting test reliability. However, do **not** select or reject a test solely based on the size of its reliability coefficient. To evaluate a test's reliability, you should consider the type of test, the type of reliability estimate reported, and the context in which the test will be used.

Table 1. General Guidelines for Interpreting Reliability Coefficients

Reliability coefficient value	Interpretation
.90 and up	excellent
.80 - .89	good
.70 - .79	adequate
below .70	may have limited applicability

4. Types of reliability estimates

There are several types of reliability estimates, each influenced by different sources of measurement error. Test developers have the responsibility of reporting the reliability estimates that are relevant for a particular test. Before deciding to use a test, read the test manual and any independent reviews to determine if its reliability is acceptable. The acceptable level of reliability will differ depending on the type of test and the reliability estimate used.

The discussion in Table 2 should help you develop some familiarity with the different kinds of reliability estimates reported in test manuals and reviews.

Table 2. Types of Reliability Estimates

#	***Test-retest reliability*** indicates the repeatability of test scores with the passage of time. This estimate also reflects the stability of the characteristic or construct being measured by the test. Some constructs are more stable than others. For example, an individual's reading ability is more stable over a particular period of time than that individual's anxiety level. Therefore, you would expect a higher test-retest reliability coefficient on a reading test than you would on a test that measures anxiety. For constructs that are expected to vary over time, an acceptable test-retest reliability coefficient may be lower than is suggested in Table 1.
#	***Alternate*** or ***parallel form reliability*** indicates how consistent test scores are likely to be if a person takes two or more forms of a test. A high parallel form reliability coefficient indicates that the different forms of the test are very similar which means that it makes virtually no difference which version of the test a person takes. On the other hand, a low parallel form reliability coefficient suggests that the different forms are probably **not** comparable; they may be measuring different things and therefore cannot be used interchangeably.
#	***Inter-rater reliability*** indicates how consistent test scores are likely to be if the test is scored by two or more raters. On some tests, raters evaluate responses to questions and determine the score. Differences in judgments among raters are likely to produce variations in test scores. A high inter-rater reliability coefficient indicates that the judgment process is stable and the resulting scores are reliable. Inter-rater reliability coefficients are typically lower than other types of reliability estimates. However, it is possible to obtain higher levels of inter-rater reliabilities if raters are appropriately trained.
#	***Internal consistency reliability*** indicates the extent to which items on a test measure the same thing. A high internal consistency reliability coefficient for a test indicates that the items on the test are very similar to each other in content (homogeneous). It is important to note that the length of a test can affect internal consistency reliability. For example, a very lengthy test can spuriously inflate the reliability coefficient. Tests that measure multiple characteristics are usually divided into distinct components. Manuals for such tests typically report a separate internal consistency reliability coefficient for each component in addition to one for the whole test. Test manuals and reviews report several kinds of internal consistency reliability estimates. Each type of estimate is appropriate under certain circumstances. The test manual should explain why a particular estimate is reported.

5. Standard error of measurement

Test manuals report a statistic called the **standard error of measurement (SEM)**. It gives the margin of error that you should expect in an individual test score because of imperfect reliability of the test. The SEM represents the degree of confidence that a person's "true" score lies within a particular range of scores. For example, an SEM of "2" indicates that a test taker's "true" score probably lies within 2 points in either direction of the score he or she receives on the test. This means that if an individual receives a 91 on the test, there is a good chance that the person's "true" score lies somewhere between 89 and 93.

The SEM is a useful measure of the accuracy of individual test scores. The smaller the SEM, the more accurate the measurements.

When evaluating the reliability coefficients of a test, it is important to review the explanations provided in the manual for the following:

- **Types of reliability used.** The manual should indicate why a certain type of reliability coefficient was reported. The manual should also discuss sources of random measurement error that are relevant for the test.
- **How reliability studies were conducted.** The manual should indicate the conditions under which the data were obtained, such as the length of time that passed between administrations of a test in a test-retest reliability study. In general, reliabilities tend to drop as the time between test administrations increases.
- **The characteristics of the sample group.** The manual should indicate the important characteristics of the group used in gathering reliability information, such as education level, occupation, etc. This will allow you to compare the characteristics of the people you want to test with the sample group. If they are sufficiently similar, then the reported reliability estimates will probably hold true for your population as well.

For more information on reliability, consult the APA Standards, the SIOP Principles, or any major textbook on psychometrics or employment testing. Appendix A lists some possible sources.

6. Test validity

Validity is the most important issue in selecting a test. Validity refers to *what characteristic* the test measures and *how well* the test measures that characteristic.

- Validity tells you if the characteristic being measured by a test is related to job qualifications and requirements.
- Validity gives *meaning* to the test scores. Validity evidence indicates that there is linkage between test performance and job performance. It can tell you what you may conclude or predict about someone from his or her score on the test. If a test has been demonstrated to be a valid predictor of performance on a specific job, you can conclude that persons scoring high

on the test are more likely to perform well on the job than persons who score low on the test, all else being equal.

- Validity also describes the *degree* to which you can make specific conclusions or predictions about people based on their test scores. In other words, it indicates the usefulness of the test.

It is important to understand the differences between *reliability* and *validity*. Validity will tell you how good a test is for a particular situation; reliability will tell you how trustworthy a score on that test will be. You cannot draw valid conclusions from a test score unless you are sure that the test is reliable. Even when a test is reliable, it may not be valid. You should be careful that any test you select is both reliable and valid for your situation.

A test's validity is established in reference to a specific purpose; the test may not be valid for different purposes. For example, the test you use to make valid predictions about someone's technical proficiency on the job may not be valid for predicting his or her leadership skills or absenteeism rate. This leads to the next principle of assessment.

Principle of Assessment

Use only assessment procedures and instruments that have been demonstrated to be valid for the specific purpose for which they are being used.

Similarly, a test's validity is established in reference to specific groups. These groups are called the reference groups. The test may not be valid for different groups. For example, a test designed to predict the performance of managers in situations requiring problem solving may not allow you to make valid or meaningful predictions about the performance of clerical employees. If, for example, the kind of problem-solving ability required for the two positions is different, or the reading level of the test is not suitable for clerical applicants, the test results may be valid for managers, but not for clerical employees.

Test developers have the responsibility of describing the reference groups used to develop the test. The manual should describe the groups for whom the test is valid, and the interpretation of scores for individuals belonging to each of these groups. You must determine if the test can be used appropriately with the particular type of people you want to test. This group of people is called your *target population* or *target group*.

Principle of Assessment

Use assessment tools that are appropriate for the target population.

Your target group and the reference group do **not** have to match on all factors; they must be sufficiently similar so that the test will yield meaningful scores for your group. For example, a writing ability test developed for use with college seniors may be appropriate for measuring the

writing ability of white-collar professionals or managers, even though these groups do not have identical characteristics. In determining the appropriateness of a test for your target groups, consider factors such as occupation, reading level, cultural differences, and language barriers.

Recall that the *Uniform Guidelines* require assessment tools to have adequate supporting evidence for the conclusions you reach with them in the event adverse impact occurs. A valid personnel tool is one that measures an important characteristic of the job you are interested in. Use of valid tools will, on average, enable you to make better employment-related decisions. Both from business-efficiency and legal viewpoints, it is essential to only use tests that are valid for your intended use.

In order to be certain an employment test is useful and valid, evidence must be collected relating the test to a job. The process of establishing the job relatedness of a test is called **validation**.

7. Methods for conducting validation studies

The *Uniform Guidelines* discuss the following three methods of conducting validation studies. The *Guidelines* describe conditions under which each type of validation strategy is appropriate. They do not express a preference for any one strategy to demonstrate the job-relatedness of a test.

- ***Criterion-related validation*** requires demonstration of a correlation or other statistical relationship between test performance and job performance. In other words, individuals who score high on the test tend to perform better on the job than those who score low on the test. If the criterion is obtained at the same time the test is given, it is called concurrent validity; if the criterion is obtained at a later time, it is called predictive validity.
- ***Content-related validation*** requires a demonstration that the content of the test represents important job-related behaviors. In other words, test items should be relevant to and measure directly important requirements and qualifications for the job.
- ***Construct-related validation*** requires a demonstration that the test measures the construct or characteristic it claims to measure, and that this characteristic is important to successful performance on the job.[3]

The three methods of validation—criterion-related, content, and construct—should be used to provide validation support depending on the situation. These three general methods often overlap, and, depending on the situation, one or more may be appropriate. French (1990) offers situational examples of when each method of validity may be applied.

First, as an example of criterion-related validity, take the position of millwright. Employees' scores (predictors) on a test designed to measure mechanical skill could be correlated with their performance in servicing machines (criterion) in the mill. If the correlation is high, it can be said

[3]Current thinking in psychology is that construct validity encompasses all other forms of validity; validation is the cumulative and on-going process of giving meaning to test scores.

that the test has a high degree of validation support, and its use as a selection tool would be appropriate.

Second, the content validation method may be used when you want to determine if there is a relationship between behaviors measured by a test and behaviors involved in the job. For example, a typing test would be high validation support for a secretarial position, assuming much typing is required each day. If, however, the job required only minimal typing, then the same test would have little content validity. Content validity does not apply to tests measuring learning ability or general problem-solving skills (French, 1990).

Finally, the third method is construct validity. This method often pertains to tests that may measure abstract traits of an applicant. For example, construct validity may be used when a bank desires to test its applicants for "numerical aptitude." In this case, an aptitude is not an observable behavior, but a concept created to explain possible future behaviors. To demonstrate that the test possesses construct validation support, ". . . the bank would need to show (1) that the test did indeed measure the desired trait and (2) that this trait corresponded to success on the job" (French, 1990, p. 260).

Professionally developed tests should come with reports on validity evidence, including detailed explanations of how validation studies were conducted. If you develop your own tests or procedures, you will need to conduct your own validation studies. As the test user, you have the ultimate responsibility for making sure that validity evidence exists for the conclusions you reach using the tests. This applies to all tests and procedures you use, whether they have been bought off-the-shelf, developed externally, or developed in-house.

Validity evidence is especially critical for tests that have adverse impact. When a test has adverse impact, the *Uniform Guidelines* require that validity evidence for that specific employment decision be provided.

The particular job for which a test is selected should be very similar to the job for which the test was originally developed. Determining the degree of similarity will require a **job analysis**. Job analysis is a systematic process used to identify the tasks, duties, responsibilities and working conditions associated with a job and the knowledge, skills, abilities, and other characteristics required to perform that job.

Job analysis information may be gathered by direct observation of people currently in the job, interviews with experienced supervisors and job incumbents, questionnaires, personnel and equipment records, and work manuals. In order to meet the requirements of the *Uniform Guidelines*, it is advisable that the job analysis be conducted by a qualified professional, for example, an industrial/organizational psychologist or other professional well trained in job analysis techniques. Job analysis information is central in deciding what to test for and which tests to use.

The President's Committee on Employment of People with Disabilities web site contains information on conducting and using a job analysis. See www50.pcepd.gov/pcepd/pubs/fact/analysis.htm

8. Using validity evidence from outside studies

Conducting your own validation study is expensive, and, in many cases, you may not have enough employees in a relevant job category to make it feasible to conduct a study. Therefore, you may find it advantageous to use professionally developed assessment tools and procedures for which documentation on validity already exists. However, care must be taken to make sure that validity evidence obtained for an "outside" test study can be suitably "transported" to your particular situation.

The *Uniform Guidelines*, the *Standards*, and the SIOP *Principles* state that evidence of transportability is required. Consider the following when using outside tests:

- ***Validity evidence.*** The validation procedures used in the studies must be consistent with accepted standards.
- ***Job similarity.*** A job analysis should be performed to verify that your job and the original job are substantially similar in terms of ability requirements and work behavior.
- ***Fairness evidence.*** Reports of test fairness from outside studies must be considered for each protected group that is part of your labor market. Where this information is not available for an otherwise qualified test, an internal study of test fairness should be conducted, if feasible.
- ***Other significant variables.*** These include the type of performance measures and standards used, the essential work activities performed, the similarity of your target group to the reference samples, as well as all other situational factors that might affect the applicability of the outside test for your use.

To ensure that the outside test you purchase or obtain meets professional and legal standards, you should consult with testing professionals. See Chapter 5 for information on locating consultants.

A brief essay, available online at the ERIC web site (ericae.net/db/edo/ED338699.htm), discusses the issue of validity generalization.

9. How to interpret validity information from test manuals and independent reviews

To determine if a particular test is valid for your intended use, consult the test manual and available independent reviews. (Chapter 5 offers sources for test reviews.) The information below can help you interpret the validity evidence reported in these publications.

- In evaluating validity information, it is important to determine whether the test can be used in the specific way you intended, and whether your target group is similar to the test reference group.

  Test manuals and reviews should describe:

— Available validation evidence supporting use of the test for specific purposes. The manual should include a thorough description of the procedures used in the validation studies and the results of those studies.

— The possible valid uses of the test. The purposes for which the test can legitimately be used should be described, as well as the performance criteria that can validly be predicted.

— The sample group(s) on which the test was developed. For example, was the test developed on a sample of high school graduates, managers, or clerical workers? What was the racial, ethnic, age, and gender mix of the sample?

— The group(s) for which the test may be used.

- The *criterion-related validity* of a test is measured by the *validity coefficient*. It is reported as a number between 0 and 1.00 that indicates the magnitude of the relationship, "r," between the test and a measure of job performance (criterion). The larger the validity coefficient, the more confidence you can have in predictions made from the test scores. However, a single test can never fully predict job performance because success on the job depends on so many varied factors. Therefore, validity coefficients, unlike reliability coefficients, rarely exceed r = .40.

  As a general rule, the higher the validity coefficient the more beneficial it is to use the test. Validity coefficients of r=.21 to r=.35 are typical for a single test. Validities for selection systems that use multiple tests will probably be higher because you are using different tools to measure/predict different aspects of performance, where a single test is more likely to measure or predict fewer aspects of total performance. Table 3 serves as a general guideline for interpreting test validity for a single test. Evaluating test validity is a sophisticated task, and you might require the services of a testing expert. In addition to the magnitude of the validity coefficient, you should also consider at a minimum the following factors:

Table 3. General Guidelines for Interpreting Validity Coefficients

Validity coefficient value	Interpretation
above .35	very beneficial
.21 - .35	likely to be useful
.11 - .20	depends on circumstances
below .11	unlikely to be useful

  — level of adverse impact associated with your assessment tool

  — selection ratio (number of applicants versus the number of openings)

  — cost of a hiring error

  — cost of the selection tool

  — probability of hiring qualified applicant based on chance alone.

Here are three scenarios illustrating why you should consider these factors, individually and in combination with one another, when evaluating validity coefficients:

Scenario One
You are in the process of hiring applicants where you have a high selection ratio and are filling positions that do not require a great deal of skill. In this situation, you might be willing to accept a selection tool that has validity considered "likely to be useful" or even "depends on circumstances" because you need to fill the positions, you do not have many applicants to choose from, and the level of skill required is not that high.

Now, let's change the situation.

Scenario Two
You are recruiting for jobs that require a high level of accuracy, and a mistake made by a worker could be dangerous and costly. With these additional factors, a slightly lower validity coefficient would probably not be acceptable to you because hiring an unqualified worker would be too much of a risk. In this case you would probably want to use a selection tool that reported validities considered to be "very beneficial" because a hiring error would be too costly to your company.

Here is another scenario that shows why you need to consider multiple factors when evaluating the validity of assessment tools.

Scenario Three
A company you are working for is considering using a very costly selection system that results in fairly high levels of adverse impact. You decide to implement the selection tool because the assessment tools you found with lower adverse impact had substantially lower validity, were just as costly, and making mistakes in hiring decisions would be too much of a risk for your company. Your company decided to implement the assessment given the difficulty in hiring for the particular positions, the "very beneficial" validity of the assessment, and your failed attempts to find alternative instruments with less adverse impact. However, your company will continue trying to find ways of reducing the adverse impact of the system.

Again, these examples demonstrate the complexity of evaluating the validity of assessments. Multiple factors need to be considered in most situations. You might want to seek the assistance of a testing expert (for example, an industrial/organizational psychologist) to evaluate the appropriateness of particular assessments for your employment situation.

When properly applied, the use of valid and reliable assessment instruments will help you make better decisions. Additionally, by using a variety of assessment tools as part of an assessment program, you can more fully assess the skills and capabilities of people, while reducing the effects of errors associated with any one tool on your decision making.

CHAPTER 4 Assessment Tools and Their Uses

This chapter briefly describes different types of assessment tools and procedures that organizations commonly use to conduct personnel assessment. Included are techniques such as employment interviews and reference checks, as well as various types of professionally developed assessment instruments. This chapter also includes a discussion of the use of medical tests and drug and alcohol testing in the workplace. Table 4, which appears at the end of this chapter, contains a brief description of the advantages and disadvantages of different types of assessment instruments.

Chapter Highlights

1. Mental and physical ability tests
2. Achievement tests
3. Biodata inventories
4. Employment interviews
5. Personality inventories
6. Honesty and integrity measures
7. Education and experience requirements (including licensing and certification)
8. Recommendations and reference checks
9. Assessment centers
10. Medical examinations
11. Drug and Alcohol tests

It takes a good deal of knowledge and judgment to properly use assessment tools to make effective employment-related decisions. Many assessment tools and procedures require specialized training, education, or experience to administer and interpret correctly. These requirements vary widely, depending on the specific instruments being used. Check with the test publisher to determine whether you and your staff meet these requirements. To ensure that test users have the necessary qualifications, some test publishers and distributors require proof of qualifications before they will release certain tests.

1. Mental and physical ability tests

When properly applied, ability tests are among the most useful and valid tools available for predicting success in jobs and training across a wide variety of occupations. Ability tests are most commonly used for entry-level jobs, and for applicants without professional training or advanced degrees. Mental ability tests are generally used to measure the *ability* to learn and perform particular job responsibilities. Examples of some mental abilities are verbal, quantitative, and

spatial abilities. Physical ability tests usually encompass abilities such as strength, endurance, and flexibility.

- **General ability tests** typically measure one or more broad mental abilities, such as verbal, mathematical, and reasoning skills. These skills are fundamental to success in many different kinds of jobs, especially where cognitive activities such as reading, computing, analyzing, or communicating are involved.
- **Specific ability tests** include measures of distinct physical and mental abilities, such as reaction time, written comprehension, mathematical reasoning, and mechanical ability, that are important for many jobs and occupations. For example, good mechanical ability may be important for success in auto mechanic and engineering jobs; physical endurance may be critical for fire fighting jobs.

Although mental ability tests are valid predictors of performance in many jobs, use of such tests to make employment decisions often results in adverse impact. For example, research suggests that mental abilities tests adversely impact some racial minority groups and, if speed is also a component of the test, older workers may be adversely impacted. Similarly, use of physical ability tests often results in adverse impact against women and older persons. See Chapter 7 for strategies to minimize adverse impact in your assessment program.

Issues involved with physical ability tests, especially as pertains to the ADA, are discussed at http://www.smallbiz.findlaw.com/text/

2. Achievement tests

Achievement tests, also known as proficiency tests, are frequently used to measure an individual's current knowledge or skills that are important to a particular job. These tests generally fall into one of the following formats:

- **Knowledge tests** typically involve specific questions to determine how much the individual knows about particular job tasks and responsibilities. Traditionally they have been administered in a paper-and-pencil format, but computer administration is becoming more common. Licensing exams for accountants and psychologists are examples of knowledge tests. Knowledge tests tend to have relatively high validity.
- **Work-sample or performance tests** require the individual to actually demonstrate or perform one or more job tasks. These tests, by their makeup, generally show a high degree of job-relatedness. For example, an applicant for an office-machine repairman position may be asked to diagnose the problem with a malfunctioning machine. Test takers generally view these tests as fairer than other types of tests. Use of these tests often results in less adverse impact than mental ability tests and job knowledge tests. However, they can be expensive to develop and administer.

A broad list of various achievement tests is available from the Buros Institute at www.unl.edu/buros/index01.html Additional information regarding achievement tests may be found at http://www.smallbiz.findlaw.com/text/

3. Biodata inventories

Biodata inventories are standardized questionnaires that gather job-relevant biographical information, such as amount and type of schooling, job experiences, and hobbies. They are generally used to predict job and training performance, tenure, and turnover. They capitalize on the well-proven notion that *past behavior is a good predictor of future behavior*.

Some individuals might provide inaccurate information on biodata inventories to portray themselves as being more qualified or experienced than they really are. Internal consistency checks (checking for consistent responses to items of similar content) can be used to detect whether there are discrepancies in the information reported. In addition, reference checks and resumes can be used to verify information.

A brief essay on the use of biodata inventories as selection instruments is available online at the ERIC web site ericae.net/db/edo/ED338702.htm

4. Employment interviews

The employment interview is probably the most commonly used assessment tool. The interview can range from being totally unplanned, that is, *unstructured*, to carefully designed beforehand, that is, completely *structured*. The most structured interviews have characteristics such as standardized questions, trained interviewers, specific question order, controlled length of time, and a standardized response evaluation format. At the other end of the spectrum, a completely unstructured interview would probably be done "off the cuff," with untrained interviewers, random questions, and with no consideration of time. A structured interview that is based on an analysis of the job in question is generally a more valid predictor of job performance than an unstructured interview. Keep in mind that interviews may contain both structured and unstructured characteristics.

Regardless of the extent to which the interview is structured or unstructured, the skill of the interviewer can make a difference in the quality of the information gathered. A skillful, trained interviewer will be able to ask job-relevant follow-up questions to clarify and explore issues brought up during the interview.

According to the Americans with Disabilities Act, it is unlawful to ask questions about medical conditions and disability before a conditional job offer. Even if a job applicant volunteers such information, the interviewer is not permitted to pursue inquiries about the nature of the medical condition or disability. Instead, the interview should be refocused so that emphasis is on the

ability of the applicant to perform the job, not on the disability. In some limited circumstances, the interviewer may ask about the need for reasonable accommodation.

Where disability is concerned, the law requires that employers provide reasonable accommodations (meaning a modification or adjustment) to a job, the work environment or the way things are usually done so that qualified individuals with a disability are not excluded from jobs that they can perform. These legal requirements apply to all selection standards and procedures, including questions and rating systems used during the interview process.

Following a structured interview format can help interviewers avoid unlawful or inappropriate inquiries where medical conditions, disability, and age are concerned. For additional information on the ADA, see the *EEOC Technical Assistance Manual on the Employment Provisions of the Americans with Disabilities Act* (janweb.icdi.wvu.edu/kinder/pages/tam1.htm) and the EEOC *ADA Enforcement Guidance: Preemployment Disability - Related Questions and Medical Examinations* (janweb.icdi.wvu.edu/kinder/pages/Medical_Exams.html) .

It is important to note that inquiries about race, ethnicity, or age generally are not expressly prohibited under the law, but usually serve no credible purpose in an interview. These types of questions are also closely scrutinized by organizations, including regulatory agencies, interested in protecting the civil rights of applicants.

Information on effectively interviewing applicants can be found at http://www.smallbiz.findlaw.com/text/

5. Personality inventories

In addition to abilities, knowledge, and skills, job success also depends on an individual's personal characteristics. Personality inventories designed for use in employment contexts are used to evaluate such characteristics as motivation, conscientiousness, self-confidence, or how well an employee might get along with fellow workers. Research has shown that, in certain situations, use of personality tests with other assessment instruments can yield helpful predictions.

Some personality inventories have been developed to determine the psychological attributes of an individual for diagnostic and therapeutic purposes. These clinical tools are not specifically designed to measure job-related personality dimensions. These tests are used in only very limited employment situations, primarily with jobs where it is critical to have some idea about an applicant's state of mind, such as in the selection of law enforcement officers or nuclear power plant workers. This distinction between clinical and employment-oriented personality inventories can be confusing. Applicants asked to take personality tests may become concerned even though only employment-oriented personality inventories will be administered.

If a personality inventory or other assessment tool provides information that would lead to identifying a mental disorder or impairment, the tool is considered a medical exam under the ADA. The ADA permits medical examinations of applicants and employees only in limited circumstances.

There are a few additional concerns about personality tests. Since there are usually no right or wrong answers to the test items, test takers may provide socially desirable answers. However, sophisticated personality inventories often have "lie-scales" built in, which allow such response patterns to be detected. There is also a general perception that these tests ask personal questions that are only indirectly relevant to job performance. This may raise concern on the part of test takers that such tests are an invasion of privacy. Some of these concerns can be reduced by including personality tests only as one part of a broader assessment program.

The Buros Institute maintains a listing of various personality tests. It is available online at www.unl.edu/buros/index12.html Additional information about personality tests may be found at http://www.smallbiz.findlaw.com/text/

The Oregon Research Institute maintains the International Personality Item Pool—a collaborative effort that includes raw data and items for a variety of personality tests. The web site may be accessed at ipip.ori.org/ipip/

6. Honesty and integrity measures

Honesty tests are a specific type of personality test. There has been an increase in the popularity of honesty and integrity measures since the Employee Polygraph Protection Act (1988) prohibited the use of polygraph tests by most private employers. Honesty and integrity measures may be broadly categorized into two types:

- **Overt integrity tests** gauge involvement in and attitudes toward theft and employee delinquency. Test items typically ask for opinions about frequency and extent of employee theft, leniency or severity of attitudes toward theft, and rationalizations of theft. They also include direct questions about admissions of, or dismissal for, theft or other unlawful activities.
- **Personality-based measures** typically contain disguised-purpose questions to gauge a number of personality traits. These traits are usually associated with a broad range of counterproductive employee behaviors, such as insubordination, excessive absenteeism, disciplinary problems, and substance abuse.

All the legitimate concerns and cautions of personality testing apply here. For instance, test takers may raise privacy concerns or question the relevance of these measures to job performance. If you choose to use an honesty test to select people for a particular job, you should document the business necessity of such a test. This would require a detailed job analysis, including an assessment of the consequences of hiring a dishonest individual. Make certain that your staff have the proper training and qualifications to administer and interpret integrity tests.

It is generally recommended that these tests be used only for pre-employment screening. Using the test with present employees could create serious morale problems. Using current employees' poor scores to make employment decisions may have legal repercussions when not substantiated by actual counterproductive behavior.

All honesty and integrity measures have appreciable prediction errors. To minimize prediction errors, thoroughly follow up on poor-scoring individuals with retesting, interviews, or reference checks. In general, integrity measures should not be used as the sole source of information for making employment decisions about individuals.

A number of states currently have statutes restricting the use of honesty and integrity measures. At least one state has an outright ban on their use. Consult regulations in your state that govern the use of honesty and integrity tests before using them.

Additional information on honesty tests and lie detector tests can be found at http://www.smallbiz.findlaw.com/text/ and http://www.smallbiz.findlaw.com/text/

Information on the application of the Employee Polygraph Protection Act is available online at www.dol.gov/dol/allcfr/ESA/Title_29/Part_801/toc.htm

7. Education and experience requirements (including licensing and certification)

Most jobs have some kind of education and experience requirements. For example, they may specify that only applicants with college degrees or equivalent training or experience will be considered. Such requirements are more common in technical, professional, and higher-level jobs. Certain licensing, certification, and education requirements are mandated by law, as in the case of truck drivers and physicians. This is done to verify minimum competence and to protect public safety.

Requirements for experience and education should be job-related. If the requirements you set result in adverse impact, you will have to demonstrate that they are job-related and justified by business necessity. However, in some cases job-relatedness might be difficult to demonstrate. For example, it is difficult to show that exactly 3 years of experience is necessary or demonstrate that a high school degree is required for a particular job.

8. Recommendations and reference checks

Recommendations and reference checks are often used to verify education, employment, and achievement records already provided by the applicant in some other form, such as during an interview or on a resume or application form. This is primarily done for professional and high-level jobs.

These verification procedures generally do not help separate potentially good workers from poor workers. This is because they almost always result in positive reports. However, use of these measures may serve two important purposes:

- they provide an incentive to applicants to be more honest with the information they provide
- they safeguard against potential negligent hiring lawsuits.

9. Assessment centers

In the assessment center approach, candidates are generally assessed with a wide variety of instruments and procedures. These could include interviews, ability and personality measures, and a range of standardized management activities and problem-solving exercises. Typical of these activities and exercises are in-basket tests, leaderless group discussions, and role-play exercises. Assessment centers are most widely used for managerial and high level positions to assess managerial potential, promotability, problem-solving skills, and decision-making skills.

- **In-basket tests** ask the candidates to sort through a manager's "in-basket" of letters, memos, directives, and reports describing problems and scenarios. Candidates are asked to examine them, prioritize them, and respond appropriately with memos, action plans, and problem-solving strategies. Trained assessors then evaluate the candidates' responses.
- **Leaderless group discussions** are group exercises in which a group of candidates is asked to respond to various kinds of problems and scenarios, without a designated group leader. Candidates are evaluated on their behavior in the group discussions. This might include their teamwork skills, their interaction with others, or their leadership skills.
- In **role-play exercises**, candidates are asked to pretend that they already have the job and must interact with another employee to solve a problem. The other employee is usually a trained assessor. The exercise may involve providing a solution to a problem that the employee presents, or suggesting some course of action regarding a hypothetical situation. Candidates are evaluated on the behavior displayed, solutions provided, or advice given.

Assessors must be appropriately trained. Their skills and experience are essential to the quality of the evaluations they provide. Assessment centers apply the whole-person approach to personnel assessment. They can be very good predictors of job performance and behavior when the tests and procedures making up the assessment center are constructed and used appropriately.

It can be costly to set up an assessment center. Large companies may have their own assessment centers; mid-size and smaller firms sometimes send candidates to private consulting firms for evaluation.

10. Medical examinations

Medical examinations are used to determine if a person can safely and adequately perform a specific job. Medical exams may also be part of a procedure for maintaining comprehensive employee health and safety plans. In some limited circumstances, medical exams may be used for evaluating employee requests for reasonable accommodation for disabilities.

The Americans with Disabilities Act outlines when and in what manner medical exams can be used in employment-related situations. For additional information on the ADA, see Chapter 2 of the Guide, the *EEOC Technical Assistance Manual on the Employment Provisions of the Americans*

*with Disabilities Act (*janweb.icdi.wvu.edu/kinder/pages/tam1.htm), *the EEOC ADA Enforcement Guidance: Preemployment Disability - Related Questions and Medical Examinations (*janweb.icdi.wvu.edu/kinder/pages/Medical_Exams.html), and the *EEOC Uniform Guidelines on Employee Selection Procedures (*www.acd.ccac.edu/hr/DOL/60_3_toc.htm) Some major points regarding medical exams are described below:

- Administering medical exams to job applicants or asking questions related to disability prior to making a job offer is prohibited.
- Once you make a job offer to an applicant, you may require a medical exam, as long as you require the exam of **all** persons entering the same job category. You may require a medical exam even if it bears no relevance to job performance. However, if you refuse to hire based on the results of the medical exam, the reasons for refusing to hire must be founded on issues of job-relevance and business necessity. In addition, you must demonstrate that no *reasonable accommodation* was available or possible without imposing *undue hardship* on your business.
- A medical exam may disqualify an individual who is deemed to be a direct threat to the health and safety of self or others. The EEOC has provided an explanation of what constitutes a direct threat. When an individual is rejected as a direct threat to health and safety:
    - — the employer must be prepared to show a significant current risk of substantial harm (not a speculative or remote risk);
    - — the specific risk must be identified;
    - — consideration of the risk must be based on objective medical or other factual evidence regarding the particular individual; and
    - — even if a genuine significant risk of substantial harm exists, the employer must consider whether it can be eliminated or reduced below the level of a direct threat by reasonable accommodation.
- Stricter rules apply for medical exams or inquiries of current employees. Unlike the rules for applicants, these exams or inquiries must be justified based on job relevance and business necessity. The need for a medical exam may arise as a result of some problems with job performance or safety caused by a medical condition or it may be mandated by federal law for certain job categories.[4]
- Your organization may conduct voluntary medical exams and inquiries of employees as part of an employee health program. However, the ADA imposes limitations on the use of this information. Medical records of all applicants and employees must be kept separate from all other personnel information.

If your organization uses medical information to make personnel decisions, you should develop a written policy on medical testing to ensure compliance with relevant federal, state, and local laws. For additional information on the ADA, see the EEOC *Technical Assistance Manual on the*

[4]Federal law (Occupational Safety and Health Act - OSHA) mandates medical monitoring of employees with exposure to specific occupational health hazards, e.g., exposure to toxic chemicals, carcinogens, or workplace sound levels exceeding 85 decibels on average.

Employment Provisions of the Americans with Disabilities Act (janweb.icdi.wvu.edu/kinder/pages/tam1.htm) and the *EEOC ADA Enforcement Guidance: Preemployment Disability - Related Questions and Medical Examinations* (janweb.icdi.wvu.edu/kinder/pages/Medical_Exams.html)

Further discussion of legal considerations of administering medical exams can be found at http://www.smallbiz.findlaw.com/text/

11. Drug and alcohol tests

An employer may prohibit the use of alcohol and illegal drugs at the workplace and may require that employees not be under the influence of either while on the job. Some commonly reported negative work behaviors and outcomes associated with alcohol and drug abuse are industrial accidents, work-related injuries, excessive absenteeism or tardiness, and workplace violence.

Current use, possession, or distribution of illicit drugs does **not** qualify as a "disability" under the ADA. You may prohibit the use of such drugs at the workplace, and you may administer drug tests to applicants and employees alike. You may deny employment to an applicant and discipline or discharge an employee currently engaged in illegal drug use. However, you may **not** discriminate against a former drug addict who has successfully undergone rehabilitation and does not currently use illicit drugs.

If your organization is in the public sector, federal courts have generally upheld the use of random drug tests only when applied to safety-sensitive positions. This federal restriction does not apply if you are a private employer. However, state or local laws and collective bargaining agreements pertaining to drug testing may impose restrictions on your drug testing policy.

Some legal medications or even some foods can produce a positive reading on a drug screening test for an individual who, in fact, has not used illegal drugs. To minimize such errors, it is advisable to have a formal appeals process, and also provisions for retesting with a more sensitive drug test when necessary.

Under the ADA, a test for the illegal use of drugs is not considered a medical exam, but a test for alcohol use is. Therefore, you must follow the ADA rules on medical exams in deciding whether and when to administer an alcohol test to applicants or employees.

Alcoholism may qualify as a disability under the ADA, and hence an individual with this condition may be extended protection. However, organizations may discipline individuals who violate conduct or performance standards that are related to the job. Organizations also may discharge, or deny employment to individuals whose use of alcohol impairs job performance or compromises safety to the extent that he or she can no longer be considered a "qualified individual with a disability."

If your organization uses drug or alcohol tests to make personnel decisions, you should develop a written policy governing such a program to ensure compliance with all relevant federal, state, and local laws. Most states require written consent of employees and applicants before drug or alcohol tests can be administered. Consult the ADA, the *EEOC Technical Assistance Manual on the Employment Provisions of the Americans with Disabilities Act* (janweb.icdi.wvu.edu/kinder/pages/tam1.htm), the *EEOC ADA Enforcement Guidance: Preemployment Disability - Related Questions and Medical Examinations* (janweb.icdi.wvu.edu/kinder/pages/Medical_Exams.html), and the *EEOC Uniform Guidelines on Employee Selection Procedures* (www.acd.ccac.edu/hr/DOL/60_3_toc.htm), as well as your state and local laws when developing a drug or alcohol testing program.

The Drug and Alcohol Testing Industry Association (DATIA) represents the drug and alcohol testing industry. Their web site is www.datia.org

Table 4. Main Advantages and Disadvantages of Different Types of Assessment Instruments

Type of assessment instrument	Advantages	Disadvantages
Ability tests	Mental ability tests • Are among the most useful predictors of performance across a wide variety of jobs • Are usually easy and inexpensive to administer	• Use of ability tests can result in high levels of adverse impact • Physical ability tests can be costly to develop and administer
Achievement/ proficiency tests	• In general, job knowledge and work-sample tests have relatively high validity • Job knowledge tests are generally easy and inexpensive to administer • Work-sample tests usually result in less adverse impact than ability tests and written knowledge tests	• Written job knowledge tests can result in adverse impact • Work-sample tests can be expensive to develop and administer
Biodata inventories	• Easy and inexpensive to administer • Some validity evidence exists • May help to reduce adverse impact when used in conjunction with other tests and procedures	• Privacy concerns may be an issue with some questions • Faking is a concern (information should be verified when possible)
Employment interviews	• Structured interviews, based on job analyses, tend to be valid • May reduce adverse impact if used in conjunction with other tests	• Unstructured interviews typically have poor validity • Skill of the interviewer is critical to the quality of interview (interviewer training can help)
Personality inventories	• Usually do not result in adverse impact • Predictive validity evidence exists for some personality inventories in specific situations • May help to reduce adverse impact when used in conjunction with other tests and procedures • Easy and inexpensive to administer	• Need to distinguish between clinical and employment-oriented personality inventories in terms of their purpose and use • Possibility of faking or providing socially desirable answers • Concern about invasion of privacy (use only as part of a broader assessment battery)
Honesty/integrity measures	• Usually do not result in adverse impact • Have been shown to be valid in some cases • Easy and inexpensive to administer	• Strong concerns about invasion of privacy (use only as part of a broader assessment battery) • Possibility of faking or providing socially desirable answers • Test users may require special qualifications for administration and interpretation of test scores • Should not be used with current employees • Some states restrict use of honesty and integrity tests

(continued)

Table 4. (continued)

Type of assessment instrument	Advantages	Disadvantages
Education and experience requirements	• Can be useful for certain technical, professional, and higher level jobs to guard against gross mismatch or incompetence	• In some cases, it is difficult to demonstrate job relatedness and business necessity of education and experience requirements
Recommendations and reference checks	• Can be used to verify information previously provided by applicants • Can serve as protection against potential negligent hiring lawsuits • May encourage applicants to provide more accurate information	• Reports are almost always positive; they do not typically help differentiate between good workers and poor workers
Assessment centers	• Good predictors of job and training performance, managerial potential, and leadership ability • Apply the *whole-person approach* to personnel assessment	• Can be expensive to develop and administer • Specialized training required for assessors; their skill is essential to the quality of assessment centers
Medical examinations	• Can help ensure a safe work environment when use is consistent with relevant federal, state, and local laws	• Cannot be administered prior to making a job offer. Restrictions apply to administering to applicants postoffer or to current employees. • There is a risk of violating applicable regulations (a *written policy*, consistent with all relevant laws, should be established to govern the entire medical testing program)
Drug and alcohol tests	• Can help ensure a safe and favorable work environment when program is consistent with relevant federal, state, and local laws	• An alcohol test is considered a medical exam and applicable law restricting medical examination in employment must be followed. • There is a risk of violating applicable regulations (a written policy, consistent with all relevant laws, should be established to govern the entire drug or alcohol testing program)

CHAPTER 5 How to Select Tests—Standards for Evaluating Tests

Previous chapters described a number of types of personnel tests and procedures and use of assessment tools to identify good workers and improve organizational performance. Technical and legal issues that have to be considered in using tests were also discussed. In this chapter, information and procedures for evaluating tests will be presented.

Chapter Highlights

1. Sources of information about tests
2. Standards for evaluating a test—information to consider to determine suitability of a test for your use
3. Checklist for evaluating a test.

Principle of Assessment

Use assessment instruments for which *understandable and comprehensive documentation* is available.

1. Sources of information about tests

Many assessment instruments are available for use in employment contexts. Sources that can help you determine which tests are appropriate for your situation are described below.

- ***Test manual.*** A test manual should provide clear and complete information about how the test was developed; its recommended uses and possible misuses; and evidence of reliability, validity, and fairness. The manual also should contain full instructions for test administration, scoring, and interpretation. In summary, a test manual should provide sufficient administrative and technical information to allow you to make an informed judgment as to whether the test is suitable for your use. You can order specimen test sets and test manuals from most test publishers.

  Test publishers and distributors vary in the amount and quality of information they provide in test manuals. The quality and comprehensiveness of the manual often reflect the adequacy of the research base behind the test. Do not mistake catalogs or pamphlets provided by test publishers and distributors for test manuals. Catalogs and pamphlets are marketing tools aimed at selling products. To get a balanced picture of the test, it is important to consult independently published critical test reviews in addition to test manuals.

- ***Mental Measurements Yearbook (MMY).*** The MMY is a major source of information about assessment tools. It consists of a continuing series of volumes. Each volume contains reviews of tests that are new or significantly revised since the publication of the previous volume. New volumes do not replace old ones; rather, they supplement them.

  The MMY series covers nearly all commercially available psychological, educational, and vocational tests published for use with English-speaking people. There is a detailed review of each test by an expert in the field. A brief description of the test covering areas such as purpose, scoring, prices, and publisher is also provided.

  The MMY is published by the Buros Institute of Mental Measurements. The Buros Institute also makes test reviews available through a computer database. This database is updated monthly via an on-line computer service. This service is administered by the Bibliographic Retrieval Services (BRS).

  Further information about the MMY, as well as an online order form is available at www.unl.edu/buros/catalog.html#mmy

- ***Tests in Print (TIP).*** TIP is another Buros Institute publication. It is published every few years and lists virtually every test published in English that is available for purchase at that time. It includes the same basic information about a test that is included in the MMY, but it does **not** contain reviews. This publication is a good starting place for determining what tests are currently available.

  Further information about TIP, including an online order form, can be found at www.unl.edu/buros/catalog.html#tip

- ***Test Critiques.*** This publication provides practical and straightforward test reviews. It consists of several volumes, published over a period of years. Each volume reviews a different selection of tests. The subject index at the back of the most recent volume directs the reader to the correct volume for each test review.

- ***Professional consultants.*** There are many employment testing experts who can help you evaluate and select tests for your intended use. They can help you design personnel assessment programs that are effective and comply with relevant laws.

  If you are considering hiring a consultant, it is important to evaluate his or her qualifications and experience beforehand. Professionals working in this field generally have a Ph.D. in industrial/organizational psychology or a related field. Look for an individual with hands-on experience in the areas in which you need assistance. Consultants may be found in psychology or business departments at universities and colleges. Others serve as full-time consultants, either working independently, or as members of consulting organizations. Typically, professional consultants will hold memberships in APA, SIOP, or other professional organizations.

Reference libraries should contain the publications discussed above as well as others that will provide information about personnel tests and procedures. The *Standards for Educational and Psychological Testing* and the *Principles for the Validation and Use of Personnel Selection Procedures* can also help you evaluate a test in terms of its development and use. In addition, these publications indicate the kinds of information a good test manual should contain. Carefully evaluate the quality and the suitability of a test before deciding to use it. Avoid using tests for which only unclear or incomplete documentation is available, and tests that you are unable to thoroughly evaluate. This is the next principle of assessment.

Principle of Assessment

Use assessment instruments for which *understandable and comprehensive documentation* is available.

2. Standards for evaluating a test—information to consider to determine suitability of a test for your use

The following basic descriptive and technical information should be evaluated before you select a test for your use. In order to evaluate a test, you should obtain a copy of the test and test manual. Consult independent reviews of the test for professional opinions on the technical adequacy of the test and the suitability of the test for your purposes.

- General information
  - **Test description.** As a starting point, obtain a full description of the test. You will need specific identifying information to order your specimen set and to look up independent reviews. The description of the test is the starting point for evaluating whether the test is suitable for your needs.
    - **Name of test.** Make sure you have the accurate name of the test. (There are tests with similar names, and you want to look up reviews of the correct instrument.)
    - **Publication date.** What is the date of publication? Is it the latest version? If the test is old, it is possible that the test content and norms for scoring and interpretation have become outdated.
    - **Publisher.** Who is the test publisher? Sometimes test copyrights are transferred from one publisher to another. You may need to call the publisher for information or for determining the suitability of the test for your needs. Is the publisher cooperative in this regard? Does the publisher have staff available to assist you?
    - **Authors.** Who developed the test? Try to determine the background of the authors. Typically, test developers hold a doctorate in industrial/organizational psychology, psychometrics, or a related field and are associated with professional organizations such

as APA. Another desirable qualification is proven expertise in test research and construction.

- **Forms.** Is there more than one version of the test? Are they interchangeable? Are forms available for use with special groups, such as non-English speakers or persons with limited reading skills?
- **Format.** Is the test available in paper-and-pencil and/or computer format? Is it meant to be administered to one person at a time, or can it be administered in a group setting?
- **Administration time.** How long does it take to administer?

— **Costs.** What are the costs to administer and score the test? This may vary depending on the version used, and whether scoring is by hand, computer, or by the test publisher.

— **Staff requirements.** What training and background do staff need to administer, score, and interpret the test? Do you have suitable staff available now or do you need to train and/or hire staff?

◗ **Purpose, nature, and applicability of the test**

— **Test purpose.** What aspects of job performance do you need to measure? What characteristics does the test measure? Does the manual contain a coherent description of these characteristics? Is there a match between what the developer says the test measures and what you intend to measure? The test you select for your assessment should relate directly to one or more important aspects of the job. A job analysis will help you identify the tasks involved in the job, and the knowledge, skills, abilities, and other characteristics required for successful performance.

— **Similarity of reference group to target group.** The test manual will describe the characteristics of the reference group that was used to develop the test. How similar are your test takers, the target group, to the reference group? Consider such factors as age, gender, racial and ethnic composition, education, occupation, and cultural background. Do any factors suggest that the test may not be appropriate for your group? In general, the closer your group matches the characteristics of the reference group, the more confidence you will have that the test will yield meaningful scores for your group.

— **Similarity of norm group to target group.** In some cases, the test manual will refer to a **norm group**. A norm group is the sample of the relevant population on whom the scoring procedures and score interpretation guidelines are based. In such cases, the norm group is the same as the reference group. If your target group differs from the norm group in important ways, then the test cannot be meaningfully used in your situation. For further discussion of norm groups, see Chapter 7.

◗ **Technical information**

— **Test reliability.** Examine the test manual to determine whether the test has an acceptable level of reliability before deciding to use it. See Chapter 3 for a discussion of how to interpret reliability information. A good test manual should provide detailed information on the types of reliabilities reported, how reliability studies were conducted, and the size and

nature of the sample used to develop the reliability coefficients. Independent reviews also should be consulted.

— **Test validity.** Determine whether the test may be validly used in the way you intended. Check the validity coefficients in the relevant validity studies. Usually the higher the validity coefficient, the more useful the test will be in predicting job success. See Chapter 3 for a discussion of how to interpret validity information. A good test manual will contain clear and complete information on the valid uses of the test, including how validation studies were conducted, and the size and characteristics of the validation samples. Independent test reviews will let you know whether the sample size was sufficient, whether statistical procedures were appropriate, and whether the test meets professional standards.

— **Test fairness.** Select tests developed to be as fair as possible to test takers of different racial, ethnic, gender, and age groups. See Chapter 7 for a discussion of test fairness. Read the manual and independent reviews of the test to evaluate its fairness to these groups. To secure acceptance by all test takers, the test should also appear to be fair. The test items should not reflect racial, cultural, or gender stereotypes, or overemphasize one culture over another. The rules for test administration and scoring should be clear and uniform. Does the manual indicate any modifications that are possible and may be needed to test individuals with disabilities?

The National Library of Education has developed an Educational Resources Information Center (ERIC). The ERIC Clearinghouse on Assessment and Evaluation provides links to a variety of sources concerned with test fairness. Locate the link "Fairness in Testing" at www.ericae.net/

The Educational Testing Service (ETS) web site (www.ets.org/fairness.html) contains information on their approach to test fairness. Much of this information may be helpful to other users/developers of tests.

— **Potential for adverse impact.** The manual and independent reviews should help you to evaluate whether the test you are considering has the potential for causing adverse impact. As discussed earlier, mental and physical ability tests have the potential for causing substantial adverse impact. However, they can be an important part of your assessment program. If these tests are used in combination with other employment tests and procedures, you will be able to obtain a better picture of an individual's job potential and reduce the effect of average score differences between groups on one test.

◗ **Practical evaluation**

— **Test tryout.** It is often useful to try the test in your own organizational setting by asking employees of your organization to take the test and by taking the test yourself. Do not compute test scores for these employees unless you take steps to ensure that results are anonymous. By trying the test out, you will gain a better appreciation of the administration procedures, including the suitability of the administration manual, test booklet, answer sheets and scoring procedures, the actual time needed, and the adequacy of the planned staffing arrangements. The reactions of your employees to the test may give you additional insight into the effect the test will have on candidates.

— **Cost-effectiveness.** Are there less costly tests or assessment procedures that can help you achieve your assessment goals? If possible, weigh the potential gain in job performance against the cost of using the test. Some test publishers and test reviews include an expectancy chart or table that you can consult to predict the expected level of performance of an individual based on his or her test score. However, make sure your target group is comparable to the reference group on which the expectancy chart was developed.

— **Independent reviews.** Is the information provided by the test manual consistent with independent reviews of the test? If there is more than one review, do they agree or disagree with each other? Information from independent reviews will prove most useful in evaluating a test.

— **Overall practical evaluation.** This involves evaluating the overall suitability of the test for your specific circumstances. Does the test appear easy to use or is it unsettling? Does it appear fair and appropriate for your target groups? How clear are instructions for administration, scoring, and interpretation? Are special equipment or facilities needed? Is the staff qualified to administer the test and interpret results or would extensive training be required?

3. Checklist for evaluating a test

It is helpful to have an organized method for choosing the right test for your needs. A checklist can help you in this process. Your checklist should summarize the kinds of information discussed above. For example, is the test valid for your intended purpose? Is it reliable and fair? Is it cost-effective? Is the instrument likely to be viewed as fair and valid by the test takers? Also consider the ease or difficulty of administration, scoring, and interpretation given available resources. A sample checklist that you may find useful appears on the following page. Completing a checklist for each test you are considering will assist you in comparing them more easily.

CHECKLIST FOR EVALUATING A TEST

Characteristic to be measured by test (skill, ability, personality trait):	
Job/training characteristic to be assessed:	
Candidate population (education, or experience level, other background):	
TEST CHARACTERISTICS	
Test name: **Version:**	
Type: (paper-and-pencil, computer) **Alternate forms available:**	
Scoring method: (hand-scored, machine-scored)	
Technical considerations:	
Reliability: r= **Validity: r=** **Reference/norm group:**	
Test fairness evidence:	
Adverse impact evidence:	
Applicability (indicate any special group):	
Administration considerations:	**Administration time:**
Materials needed (include start-up costs, operational and scoring cost):	**Costs:**
Facilities needed:	
Staffing requirements:	
Training requirements:	
Other considerations (consider clarity, comprehensiveness, utility):	
Test manual:	
Supporting documents from the publisher:	
Publisher assistance:	
Independent reviews:	
Overall evaluation:	

CHAPTER 6 Administering Assessment Instruments

Proper administration of assessment instruments is essential to obtaining *valid* or *meaningful* scores for your test takers. This chapter discusses how to administer assessment instruments so that you can be certain that the results will be valid and fair.

Chapter Highlights

1. Training and qualifications of administration staff
2. Following instructions and guidelines stated in the test manual
3. Ensuring suitable and uniform assessment conditions
4. How much help to offer test takers
5. Test anxiety
6. Alternative assessment methods for special cases
7. Providing reasonable accommodation in the assessment process to people with disabilities
8. Administering computer-based tests
9. Obtaining informed consent of test takers and a waiver of liability claims
10. Maintaining assessment instrument security
11. Maintaining confidentiality of assessment results
12. Testing unionized employees

Principles of Assessment Discussed

Ensure that administration staff are *properly trained.*
Ensure that *testing conditions are suitable* for all test takers.
Provide *reasonable accommodation* in the assessment process for people with disabilities.
Maintain assessment instrument *security*.
Maintain *confidentiality* of assessment results.

1. Training and qualifications of administration staff

The qualifications and training required for a test administrator will depend on the nature and complexity of the test. The more complex the test administration procedures, the more training an administrator will need. However, even simple-to-administer tests need trained staff to ensure valid results. Administrators should be given ample time to learn their responsibilities before they administer a test to applicants. Your staff may need professional training on test administration offered by some test publishers.

Only those staff who can administer the test in a professional and satisfactory manner should be assigned test administration duties. Test administrators should be well organized and observant, speak well, and be able to deal comfortably with people. They should also be trained to handle special situations with sensitivity. For example, they should know how to respond to a test taker's request for an accommodation and be able to calm down those who may become overly anxious about taking a test. This leads to our next principle of assessment.

Principle of Assessment

Ensure that administration staff are *properly trained*.

2. Following instructions and guidelines stated in the test manual

Staff should be thoroughly familiar with the testing procedures before administering the test. They should carefully follow **all** standardized administration and scoring procedures as outlined in the test manual. Test manuals will indicate the test materials that are needed, the order of presentation, and the instructions that must be read verbatim. They will also indicate whether there are time limits, and, if so, what those time limits are. Any special instructions noted by the test manual should be observed. This includes meeting the requirements for specific equipment or facilities. Alterations can invalidate results.

3. Ensuring suitable and uniform assessment conditions

There are various extraneous influences that may affect the reliability and validity of an assessment procedure. To maintain the integrity of results you and your staff should make sure that adverse conditions are minimized.

- **Choose a suitable testing location.** Obtain a room that is well-lit, well-ventilated, with acceptable room temperature. Make sure the room is free of noise, traffic, and other interruptions. Chairs should be comfortable and tables should be at an appropriate height, with sufficient room for test booklets and answer sheets. Furthermore, testing facilities and

conditions must be uniform for all test takers. This means that people taking the test in another room, or at a different time, should be in substantially the same testing environment. As indicated in Chapter 3, these extraneous factors can affect the reliability and validity of test results.

- **Prepare the room and test materials ahead of time.** Chairs and tables should be set up in position. Staff should check that all needed test materials and equipment are available and in good condition.
- **Test taker readiness or suitability for testing.** Be alert to problems individuals may have in taking the test. Before the assessment begins, give them an overview of the test and ask whether anyone anticipates having a problem taking the test. Some test takers may have forgotten to bring their eyeglasses; others may have bad colds or other temporary illnesses. These individuals should be rescheduled. Others may have disabilities that require accommodations or an alternate assessment arrangement (see section on ADA in Chapter 2).
- **Uniform administration.** The practices and precautions discussed above should become standard procedures in preparing testing materials, equipment, and facilities. Also, make sure that all test takers understand the directions before the test begins and are ready to follow the standard set of instructions during the test. These steps will help ensure that the results reflect real differences among individuals, and not differences in test administration. This brings us to the next principle of assessment.

Principle of Assessment

Ensure that *testing conditions are suitable* for all test takers.

To maintain the integrity of test results, administrators need to be alert to test takers' activities throughout the session. For example, some individuals may lose their place in the test booklet or put answers in the wrong column on the answer sheet. Others may try to copy answers from someone else. An alert administrator will be able to correct these situations quickly before they invalidate the test takers' responses.

4. How much help to offer test takers

The test manual usually indicates the kind of assistance and information that can be provided to test takers during the test. Administration staff should be familiar with what is and is not permissible at each stage of the assessment process.

Some instruments allow the administrator to clarify the directions and practice exercises, but prohibit help with the actual test questions. This is generally true for ability and achievement tests. However, other assessment tools, such as interest inventories or biodata instruments, may allow for more assistance with the assessment.

In general, test takers should not be coached on how best to answer test questions. Administrators should not offer more information than what is indicated in the instructions. If they do, some individuals will be given an unfair advantage.

5. Test anxiety

Most people feel some anxiety about taking a test. For some otherwise qualified individuals, test anxiety can have a paralyzing effect on their performance. There are a few things that can be done to alleviate anxiety:

- Written orientation materials are available for many tests. These materials describe the test and provide sample questions. If such materials exist, they should be made available to all test takers well in advance of the test date.
- Before the test begins, give test takers a brief orientation explaining the purpose of the test, the type of questions to expect, and how long the test will last.
- Start test sessions promptly. A long wait will raise the anxiety level among test takers. All testing materials, equipment, and facilities should be ready well in advance of the scheduled session. A well-run test session helps to reduce test anxiety.

6. Alternative assessment methods for special cases

There may be qualified individuals who, due to cultural differences, poor skills in English, or limited formal education, are unable to score satisfactorily on some of the currently available selection tests. Poor test performance may not be a reflection of their job-related knowledge, skills, or abilities, but rather may be due to the existence of a cultural or language barrier. Some of these tests may be available in appropriate foreign language versions or in a version suitable for individuals functioning at low literacy levels. Also, where appropriate, work samples and structured interviews should be considered seriously as practical alternatives to written tests. At times, individual evaluations by outside agencies or consultants may be a suitable approach.

7. Providing reasonable accommodation in the assessment process to people with disabilities

The ADA has opened up employment opportunities for a great number of qualified persons with disabilities. These opportunities have enabled persons with disabilities to apply their skills and be successful in the world of work. Under the ADA, you are required to provide reasonable accommodation in the assessment process to qualified persons with disabilities. This leads to our next principle of assessment.

Principle of Assessment

Provide *reasonable accommodation* in the assessment process for people with disabilities.

Accommodation in the assessment process may involve ensuring physical accessibility to the test site, modifying test equipment or tests, or providing qualified assistance. Giving extra time on certain kinds of tests to test takers with dyslexia or other learning disability, and administering a larger print version of a test to a person who is visually impaired are examples of reasonable accommodation. Note, however, that providing a reader for a reading comprehension test, or extra time for a speeded test could invalidate the test results. You should become familiar with what accommodations can be made for different conditions or circumstances without invalidating the test. Provide all test takers with descriptive information about the test in advance, so that they will have ample opportunity to request needed accommodations. When the need for accommodation is not obvious, you may ask for reasonable documentation of the disability functional limitations for which accommodation is needed. The test taker, test manual, the test publisher, and several professional associations (listed in Chapter 2 and Appendix A) can help you determine what the appropriate reasonable accommodations are for particular situations. If an accommodation cannot be made without invalidating the test, alternative assessment strategies, such as a review of past job experience, a review of school records, or a brief job tryout, must be considered.

8. Administering computer-based tests

Many tests are now computer-based. Computers can be used to administer and score tests and print results. A number of computerized tests also provide extensive test interpretations.

Some computer-based tests are adaptive. Adaptive tests, as opposed to conventional tests, present test questions based on the responses of the test taker to previous questions, and so adjust for his or her level of ability. This allows for a more reliable measure of ability with fewer items administered.

Advantages to computer-based testing include:

- Administration procedures are the same for all test takers.
- The need for test administrators is reduced.
- Results can be available immediately.
- The test can be administered without delay to walk-in applicants.

Disadvantages of computer-based testing include:

- A computer is needed for each test taker.
- Some test takers may feel uncomfortable using a computer; this could raise anxiety levels and adversely affect scores of these individuals.

9. Obtaining informed consent of test takers and a waiver of liability claims

When a test taker gives informed consent, it implies that he or she understands the nature of the test, the reasons for it, and how the results will be used. In applications for employment and educational admissions, informed consent is clearly implied, so obtaining permission is typically not required. However, there may be state regulations requiring that written consent of test takers be obtained before certain kinds of tests can be administered. For example, most states require written permission of test takers before drug or alcohol tests can be administered. You should also obtain similar permission when administering honesty or integrity measures and physical exams.

Obtaining written consent does not relieve the organization of legal liability if applicable laws are violated.

10. Maintaining assessment instrument security

In order to obtain fair and valid results, no test taker should have an opportunity to view the test beforehand. To ensure this, keep test materials secure at all times. Store all materials relating to the test in locked rooms or cabinets when not in use, and account for all materials that are used during the testing session. Test takers should not take any items from the testing room, including scrap paper. Limit access to testing materials to staff involved in the assessment process. This brings us to the next principle of assessment.

Principle of Assessment

> Maintain assessment instrument *security*.

Security measures are also required when you use computer-based tests. Establish a password procedure for accessing computerized test materials, and secure all related computer disks and manuals. Many computerized test developers encode test items and answer keys so that items cannot easily be read if electronic files are stolen.

When tests are used over a long period of time, it becomes increasingly likely that some test questions will leak out. To help maintain security, test developers periodically introduce new alternate forms. If alternate forms of the test are available, you can increase security by varying the form used.

11. Maintaining confidentiality of assessment results

Test results and answer sheets should be kept in a secure location. Results should only be released to those who have a legitimate need to know. This includes staff involved in making the employment decision but may exclude the candidate's first-line supervisor. Test results are confidential and should not be disclosed to another individual or outside organization without the informed consent of the test taker. This is the next principle of assessment.

Principle of Assessment

> Maintain *confidentiality* of assessment results.

As discussed in Chapters 2 and 4, under the ADA, medical information about employees and applicants is confidential and must be kept in a separate location from other personnel information.

12. Testing unionized employees

Testing may be a mandatory subject of collective bargaining between management and labor unions. Therefore, if you are a unionized employer, do not institute a testing program or revise a current program without first referring to the collective bargaining agreement. Include representatives of the union on teams or task forces charged with designing and implementing personnel assessment programs.

CHAPTER 7 Using, Scoring, and Interpreting Assessment Instruments

This chapter describes some of the most common assessment instrument scoring procedures. It also discusses how to properly interpret results, and how to use them effectively. Other issues regarding the proper use of assessment tools are also discussed.

Chapter Highlights

1. Assessment instrument scoring procedures
2. Test interpretation methods: norm and criterion-referenced tests
3. Interpreting test results
4. Processing test results to make employment decisions—rank-ordering and cut-off scores
5. Combining information from many assessment tools
6. Minimizing adverse impact

Principle of Assessment

Ensure that scores are interpreted properly.

1. Assessment instrument scoring procedures

Test publishers may offer one or more ways to score the tests you purchase. Available options may range from hand scoring by your staff to machine scanning and scoring done by the publisher. All options have their advantages and disadvantages. When you select the tests for use, investigate the available scoring options. Your staff's time, turnaround time for test results, and cost may all play a part in your purchasing decision.

- **Hand scoring.** The answer sheet is scored by counting the number of correct responses, often with the aid of a stencil. These scores may then have to be converted from the *raw score* count to a form that is more meaningful, such as a *percentile* or *standard score*. Staff must be trained on proper hand scoring procedures and raw score conversion. This method is more prone to error than machine scoring. To improve accuracy, scoring should be double checked. Hand scoring a test may take more time and effort, but it may also be the least expensive method when there are only a small number of tests to score.

- **Computer-based scoring.** Tests can be scored using a computer and test scoring software purchased from the test publisher. When the test is administered in a paper-and-pencil format, raw scores and identification information must be key-entered by staff following the completion of the test session. Converted scores and interpretive reports can be printed immediately. When the test is administered on the computer, scores are most often generated automatically upon completion of the test; there is no need to key-enter raw scores or identifying information. This is one of the major advantages of computer-based testing.

- **Optical scanning.** Machine scorable answer sheets are now readily available for many multiple choice tests. They are quickly scanned and scored by an optical mark reader. You may be able to score these answer sheets in-house or send them to the test publisher for scoring.

  — ***On-site.*** You will need a personal computer system (computer, monitor, and printer), an optical reader, and special test scoring software from the publisher. Some scanning programs not only generate test scores but also provide employers with individual or group interpretive reports. Scanning systems can be costly, and the staff must learn to operate the scanner and the computer program that does the test scoring and reporting. However, using a scanner is much more efficient than hand scoring, or key-entering raw scores when testing volume is heavy.

  — ***Mail-in and fax scoring.*** In many cases the completed machine-scannable answer sheets can be mailed or faxed to the test publisher. The publisher scores the answer sheets and returns the scores and test reports to the employer. Test publishers generally charge a fee for each test scored and for each report generated.

  For mail-in service, there is a delay of several days between mailing answer sheets and receipt of the test results from the service. Overnight mail by private or public carrier will shorten the wait but will add to the cost. Some publishers offer a scoring service by fax machine. This will considerably shorten the turn-around time, but greater care must be taken to protect the confidentiality of the results.

2. Test interpretation methods: norm and criterion-referenced tests

Employment tests are used to make inferences about people's characteristics, capabilities, and likely future performance on the job. What does the test score mean? Is the applicant qualified? To help answer these questions, consider what the test is designed to accomplish. Does the test compare one person's score to those obtained by others in the occupation, or does it measure the absolute level of skill an individual has obtained? These two methods are described below.

- **Norm-referenced test interpretation.** In norm-referenced test interpretation, the scores that the applicant receives are compared with the test performance of a particular reference group. In this case the reference group is the norm group. The norm group generally consists of large representative samples of individuals from specific populations, such as high school students, clerical workers, or electricians. It is their average test performance and the distribution of their scores that set the standard and become the test norms of the group.

The test manual will usually provide detailed descriptions of the norm groups and the test norms. To ensure valid scores and meaningful interpretation of norm-referenced tests, make sure that your target group is similar to the norm group. Compare the educational level, the occupational, language and cultural backgrounds, and other demographic characteristics of the individuals making up the two groups to determine their similarity.
For example, consider an accounting knowledge test that was standardized on the scores obtained by employed accountants with at least 5 years of experience. This would be an appropriate test if you are interested in hiring experienced accountants. However, this test would be inappropriate if you are looking for an accounting clerk. You should look for a test normed on accounting clerks or a closely related occupation.

- **Criterion-referenced test interpretation.** In criterion-referenced tests, the test score indicates the amount of skill or knowledge the test taker possesses in a particular subject or content area. The test score is not used to indicate how well the person does compared to others; it relates solely to the test taker's degree of competence in the specific area assessed. Criterion-referenced assessment is generally associated with educational and achievement testing, licensing, and certification.

  A particular test score is generally chosen as the minimum acceptable level of competence. How is a level of competence chosen? The test publisher may develop a mechanism that converts test scores into proficiency standards, or the company may use its own experience to relate test scores to competence standards.

  For example, suppose your company needs clerical staff with word processing proficiency. The test publisher may provide you with a conversion table relating word processing skill to various levels of proficiency, or your own experience with current clerical employees can help you to determine the passing score. You may decide that a minimum of 35 words per minute with no more than two errors per 100 words is sufficient for a job with occasional word processing duties. If you have a job with high production demands, you may wish to set the minimum at 75 words per minute with no more than 1 error per 100 words.

It is important to ensure that all inferences you make on the basis of test results are well founded. Only use tests for which sufficient information is available to guide and support score interpretation. Read the test manual for instructions on how to properly interpret the test results. This leads to the next principle of assessment.

Principle of Assessment

Ensure that scores are interpreted properly.

3. Interpreting test results

Test results are usually presented in terms of numerical scores, such as raw scores, standard scores, and percentile scores. In order to interpret test scores properly, you need to understand the scoring system used.

◗ Types of scores

— **Raw scores.** These refer to the unadjusted scores on the test. Usually the raw score represents the number of items answered correctly, as in mental ability or achievement tests. Some types of assessment tools, such as work value inventories and personality inventories, have no "right" or "wrong" answers. In such cases, the raw score may represent the number of positive responses for a particular trait.

Raw scores do not provide much useful information. Consider a test taker who gets 25 out of 50 questions correct on a math test. It's hard to know whether "25" is a good score or a poor score. When you compare the results to all the other individuals who took the same test, you may discover that this was the highest score on the test.

In general, for norm-referenced tests, it is important to see where a particular score lies within the context of the scores of other people. Adjusting or converting raw scores into standard scores or percentiles will provide you with this kind of information. For criterion-referenced tests, it is important to see what a particular score indicates about proficiency or competence.

— **Standard scores.** Standard scores are converted raw scores. They indicate where a person's score lies in comparison to a reference group. For example, if the test manual indicates that the average or mean score for the group on a test is 50, then an individual who gets a higher score is above average, and an individual who gets a lower score is below average. Standard scores are discussed in more detail below in the section on standard score distributions.

— **Percentile score.** A percentile score is another type of converted score. An individual's raw score is converted to a number indicating the percent of people in the reference group who scored below the test taker. For example, a score at the 70th percentile means that the individual's score is the same as or higher than the scores of 70% of those who took the test. The 50th percentile is known as the median and represents the middle score of the distribution.

◗ Score distribution

— **Normal curve.** A great many human characteristics, such as height, weight, math ability, and typing skill, are distributed in the population at large in a typical pattern. This pattern of distribution is known as the *normal curve* and has a symmetrical bell-shaped appearance. The curve is illustrated in Figure 2. As you can see, a large number of individual cases cluster in the middle of the curve. The farther from the middle or average you go, the fewer the cases. In general, distributions of test scores follow the same normal curve pattern. Most individuals get scores in the middle range. As the

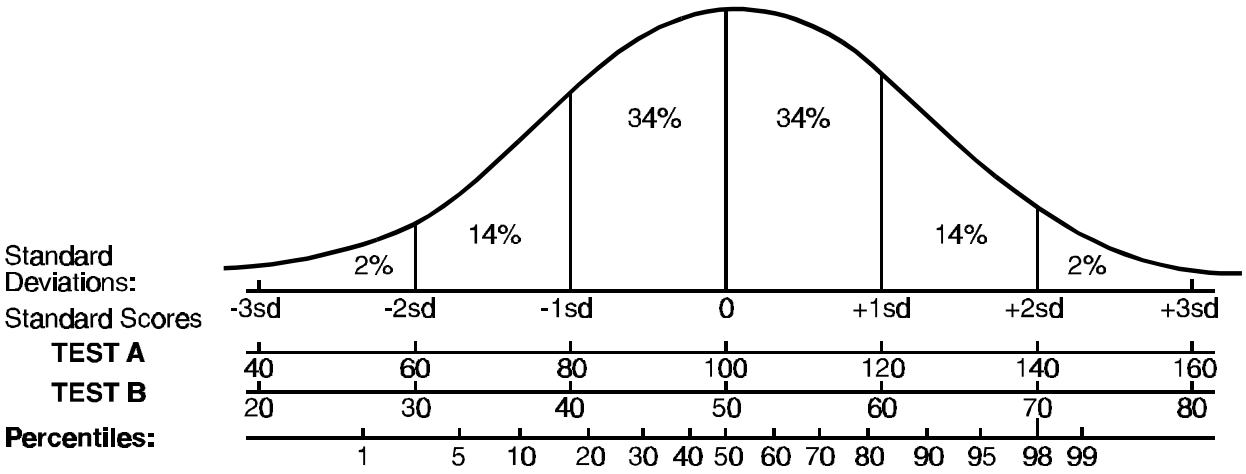

Figure 2. The normal curve illustrating standard score and percentile distribution.

extremes are approached, fewer and fewer cases exist, indicating that progressively fewer individuals get low scores (left of center) and high scores (right of center).

— **Standard score distribution.** There are two characteristics of a standard score distribution that are reported in test manuals. One is the mean, a measure of central tendency; the other is the standard deviation, a measure of the variability of the distribution.

- **Mean.** The most commonly used measure of central tendency is the mean or arithmetic average score. Test developers generally assign an arbitrary number to represent the mean standard score when they convert from raw scores to standard scores. Look at Figure 2. Test A and Test B are two tests with different standard score means. Notice that Test A has a mean of 100 and Test B has a mean of 50. If an individual got a score of 50 on Test A, that person did very poorly. However, a score of 50 on Test B would be an average score.

- **Standard deviation.** The standard deviation is the most commonly used measure of variability. It is used to describe the distribution of scores around the mean. Figure 2 shows the percent of cases 1, 2, and 3 standard deviations (sd) above the mean and 1, 2, and 3 standard deviations below the mean. As you can see, 34% of the cases lie between the mean and +1 sd, and 34% of the cases lie between the mean and -1 sd. Thus, approximately 68% of the cases lie between -1 and +1 standard deviations.

  Notice that for Test A, the standard deviation is 20, and 68% of the test takers score between 80 and 120. For Test B the standard deviation is 10, and 68% of the test takers score between 40 and 60.

— **Percentile distribution.** The bottom horizontal line below the curve in Figure 2 is labeled "Percentiles." It represents the distribution of scores in percentile units. Notice that the median is in the same position as the mean on the normal curve. By knowing the percentile score of an individual, you already know how that individual compares with others in the group. An individual at the 98th percentile scored the same or better than 98% of the

individuals in the group. This is equivalent to getting a standard score of 140 on Test A or 70 on Test B.

4. Processing test results to make employment decisions—rank-ordering and cut-off scores

The rank-ordering of test results, the use of cut-off scores, or some combination of the two is commonly used to assess the qualifications of people and to make employment-related decisions about them. These are described below.

Rank-ordering is a process of arranging candidates on a list from highest score to lowest score based on their test results. In rank-order selection, candidates are chosen on a top-down basis.

A **cut-off score** is the minimum score that a candidate must have to qualify for a position. Employers generally set the cut-off score at a level which they determine is directly related to job success. Candidates who score below this cut-off generally are not considered for selection. Test publishers typically recommend that employers base their selection of a cut-off score on the norms of the test.

5. Combining information from many assessment tools

Many assessment programs use a variety of tests and procedures in their assessment of candidates. In general, you can use a "multiple hurdles" approach or a "total assessment" approach, or a combination of the two, in using the assessment information obtained.

- ***Multiple hurdles approach.*** In this approach, test takers must pass each test or procedure (usually by scoring above a cut-off score) to continue within the assessment process. The multiple hurdles approach is appropriate and necessary in certain situations, such as requiring test takers to pass a series of tests for licensing or certification, or requiring all workers in a nuclear power plant to pass a safety test. It may also be used to reduce the total cost of assessment by administering less costly screening devices to everyone, but having only those who do well take the more expensive tests or other assessment tools.
- ***Total assessment approach.*** In this approach, test takers are administered every test and procedure in the assessment program. The information gathered is used in a flexible or *counterbalanced* manner. This allows a high score on one test to be counterbalanced with a low score on another. For example, an applicant who performs poorly on a written test, but shows great enthusiasm for learning and is a very hard worker, may still be an attractive hire.

  A key decision in using the total assessment approach is determining the relative weights to assign to each assessment instrument in the program.

Figure 3 is a simple example of how assessment results from several tests and procedures can be combined to generate a weighted composite score.

Assessment instrument	Assessment score (0-100)	Assigned weight	Weighted total
Interview	80	8	640
Mechanical ability test	60	10	600
H.S. course work	90	5	450
		Total Score: 1,690	

An employer is hiring entry-level machinists. The assessment instruments consist of a structured interview, a mechanical ability test, and high school course work. After consultation with relevant staff and experts, a weight of 8 is assigned for the interview, 10 for the test, and 5 for course work. A sample score sheet for one candidate, Candidate A, is shown above. As you can see, although Candidate A scored lowest on the mechanical ability test, the weights of all of the assessment instruments as a composite allowed him/her to continue on as a candidate for the machinist job rather than being eliminated for consideration as a result of the one low score.

Figure 3. Score-sheet for entry level machinist job: Candidate A.

6. Minimizing adverse impact

A well-designed assessment program will improve your ability to make effective employment decisions. However, some of the best predictors of job performance may exhibit adverse impact. As a test user, there are several good testing practices to follow to minimize adverse impact in conducting personnel assessment and to ensure that, if adverse impact does occur, it is not a result of deficiencies in your assessment tools:

- Be clear about what needs to be measured, and for what purpose. Use only assessment tools that are job-related and valid, and only use them in the way they were designed to be used.
- Use assessment tools that are appropriate for the target population.
- Do not use assessment tools that are biased or unfair to any group of people.
- Consider whether there are alternative assessment methods that have less adverse impact.
- Consider whether there is another way to use the test that either reduces or is free of adverse impact.

- Consider whether use of a test with adverse impact is necessary. Does the test improve the quality of selections to such an extent that the magnitude of adverse impact is justified by business necessity?
- If you determine that it is necessary to use a test that may result in adverse impact, it is recommended that it be used as only *one part* of a comprehensive assessment process. That is, apply the whole-person approach to your personnel assessment program. This approach will allow you to improve your assessment of the individual and reduce the effect of differences in average scores between groups on a single test.

CHAPTER 8 Issues and Concerns with Assessment

It is important to remember that an assessment instrument, like any tool, is most effective when used properly and can be very counterproductive when used inappropriately. In previous chapters you have read about the advantages of using tests and procedures as part of your personnel assessment program. You have also read about the *limitations* of tests in providing a consistently accurate and complete picture of an individual's employment-related qualifications and potential. This chapter highlights some important issues and concerns surrounding these limitations. Careful attention to these issues and concerns will help you produce a fair and effective assessment program.

Chapter Highlights

1. Deciding whether to test or not to test
2. Viewing tests as threats and invasions of privacy
3. Fallibility of test scores
4. Appeals process and retesting
5. Qualifications of assessment staff
6. Misuse or overuse of tests
7. Ensuring both efficiency and diversity
8. Ethnic, linguistic, and cultural differences and biases
9. Testing people with disabilities

1. Deciding whether to test or not to test

How successful is your current assessment program? Is it in need of improvement? The decision to use a test is an important one. You need to carefully consider several technical, administrative, and practical matters.

Sometimes a more vigorous employee training program will help to improve individual and organizational performance without expanding your current selection procedures. Sometimes a careful review of each candidate's educational background and work history will help you to select better workers, and sometimes using additional tests will be beneficial.

Consider how much additional time and effort will be involved in expanding your assessment program. As in every business decision, you will want to determine whether the potential benefits outweigh the

expenditure of time and effort. Be sure to factor in all the costs, such as purchase of tests and staff time, and balance these against all the benefits, including potential increases in productivity.

In summary, before expanding your assessment program, it is important to have a clear picture of your organization's needs, the benefits you can expect, and the costs you will incur.

2. Viewing tests as threats and invasions of privacy

Many people are intimidated at the mere thought of taking a test. Some may fear that testing will expose their weaknesses, and some may fear that tests will not measure what they really can do on the job. Also, some people may view certain tests as an invasion of privacy. This is especially true of personality tests, honesty tests, medical tests, and tests that screen for drug use.

Fear or mistrust of tests can lower the scores of some otherwise qualified candidates. To reduce these feelings, it is important to take the time to explain a few things about the testing program before administering a test. Any explanation should, at a minimum, cover the following topics:

- why the test is being administered
- fairness of the test
- confidentiality of test results
- how the test results will be used in the assessment process.

3. Fallibility of test scores

All assessment tools and procedures are subject to measurement errors. This means that a test neither measures a characteristic with perfect accuracy for all people, nor fully accounts for their job performance. Thus, there will always be some errors in employment decisions made based on assessment results. This is true of all assessment procedures, regardless of how objective or standardized they might be.

It is, therefore, important not to rely entirely on any one assessment instrument in making employment decisions. Using a variety of assessment tools will help you obtain a fuller and more accurate picture of an individual. Consider such information as an evaluation of a person's education, work experience and other job-relevant factors in addition to standardized test results.

4. Appeals process and retesting

Every test taker should have a fair chance to demonstrate his or her best performance on an assessment procedure. However, at times this might not occur. If the results may not be valid for an individual, consider retesting or using alternative assessment procedures before screening the individual.

There are external circumstances or conditions that could invalidate the test results. These may include the test taker's state of mind or health at the time of the test, the conditions under which the test is given, and his or her familiarity with particular questions on the test. To give some specific examples, a person who has a child at home with the measles may not be able to concentrate on taking a vocabulary test. Someone sitting next to a noisy air conditioner may also not be able to concentrate on the test questions. On another day, under different circumstances, these individuals might obtain a different score.

If you believe that the test was not valid for an individual, you should consider a retest. If other versions of the test are not available, consider alternative means of assessment. Check the test manual for advice from the publisher regarding retesting. It is advisable to develop a policy on handling complaints regarding testing and appeals for retesting, so that these concerns can be resolved fairly and consistently.

5. Qualifications of assessment staff

Test results may not be accurate if the tests have not been administered and scored properly, or if the results are not interpreted appropriately. The usefulness of test results depends on proper administration, scoring and interpretation. Qualified individuals must be chosen to administer and score tests and interpret test results. These individuals must be trained appropriately. Test manuals will usually specify the qualifications and training needed to administer and score the tests and interpret results.

6. Misuse or overuse of tests

A single test cannot be expected to be valid in all situations and for all groups of people. A test generally is developed to measure specific characteristics and to predict specific performance criteria for a particular group. For example, a test with items designed to select salespersons may not be valid for identifying good sales managers.

In addition, test results usually provide specific information that is valid for a specific amount of time. Therefore, it is unlikely to be appropriate to consider an employee for a promotion based on his or her test scores on a proficiency test taken 5 years earlier.

The test manual and independent reviews of the test remain your best guides on administering, scoring, and interpreting the test.

7. Ensuring both efficiency and diversity

Use of reliable and valid assessment tools can result in improved performance of your workforce. However, when designing an assessment system, it is also important to consider how to ensure a diverse workforce that can help your organization be successful in todays diverse marketplace. To encourage diversity in your organization, consider how different types of people perform on different types of tests. Some research has indicated that older workers and members of a variety of racial and ethnic groups do not do as well on certain types of tests as members of other groups. For example, older people and women tend to do less well on physical ability and endurance tests. Members of some ethnic and racial groups, on average, may do less well on ability tests. Older people tend not to score as high as younger people on timed tests. Even though these groups perform less well on certain tests, they may still perform on the job successfully. Thus by using certain types of assessments, or relying heavily on one type of test, you may limit the diversity of your workforce and miss out on some very productive potential employees (e.g., if you used only physical ability tests, you may unnecessarily exclude older workers). You might also be violating federal, state, and local equal employment opportunity laws.

To help ensure both efficiency and diversity in your workforce, apply the whole-person approach to assessment. Use a variety of assessment tools to obtain a comprehensive picture of the skills and capabilities of applicants and employees. This approach to assessment will help you make sure you don't miss out on some very qualified individuals who could enhance your organization's success.

8. Ethnic, linguistic, and cultural differences and biases

The American workforce is made up of a diverse array of ethnic and cultural groups, including many persons for whom English is not the primary language. Some of these individuals may experience difficulty on standardized tests due to cultural differences or lack of mastery of the English language. Depending on the nature of the job for which they are applying, this could mean that their test scores will not accurately predict their true job potential.

Before selecting new tests, consider the composition of your potential candidate population. Are the tests appropriate for all of them? The test manuals may provide assistance in determining this. If you need further clarification, contact the test publisher.

There may be cases where appropriate standardized tests are not available for certain groups. You may have to rely on other assessment techniques, such as interviews and evaluations of education and work experience, to make your employment decisions.

9. Testing people with disabilities

Many people with disabilities are productive workers. The ADA protects qualified individuals with disabilities from discrimination in all aspects of employment, including personnel assessment. Your staff should be trained to evaluate requests for reasonable accommodation and provide these accommodations if they are necessary and would not cause "undue hardship." These situations must be handled with professionalism and sensitivity. Properly handled, this can be accomplished without compromising the integrity of the assessment process.

Accommodation may involve ensuring physical accessibility to the test site, modifying test equipment or tests, or providing other forms of assistance. Giving extra time for certain kinds of tests to test takers with dyslexia or other learning disabilities and administering a braille version of a test for the blind may be examples of reasonable accommodation. See Chapters 2 and 6 for further discussions on testing people with disabilities.

CHAPTER 9 A Review—Principles of Assessment

Employers can effectively use personnel assessment instruments to measure job-relevant skills and capabilities of applicants and employees. These tools can help to identify and select better workers and can help improve the quality of an organization's overall performance. To use these tools properly, employers must be aware of the inherent limitations of any assessment procedure, as well as the legal issues involved in assessment.

The guide is organized around 13 important assessment principles and their applications. This final chapter brings all the principles together. They are listed below in the order of their appearance in the text, with the chapter number in parentheses. Together, the 13 principles provide a comprehensive framework for conducting an effective personnel assessment program.

◗ Use assessment tools in a *purposeful manner* (Chapter 1)

Assessment instruments, like other tools, are helpful when used properly but can be useless, harmful, or illegal when used inappropriately. Often, inappropriate use results from not having a clear understanding of what you want to measure and why you want to measure it. As an employer, you must first be clear about what you want to accomplish with your assessment program in order to select the proper tools to achieve those goals.

Your assessment strategies should be based on both an understanding of the kind of employment decisions to be made and the population to be assessed. Once you are clear on your purpose, you will be better able to select appropriate assessment tools, and use those tools in an effective manner. Only use tests that are appropriate for your particular purpose.

◗ Use the *whole-person approach* to assessment (Chapter 1)

An assessment instrument may provide you with important employment-related information about an individual. However, no assessment tool is 100% reliable or valid; all are subject to errors, both in measuring job-relevant characteristics and in predicting job performance. Moreover, a single assessment instrument only provides you with a limited view of a person's qualifications. Using a variety of tools to measure skills, abilities, and other job-relevant characteristics provides you with a solid basis upon which to make important career and employment-related decisions and minimizes adverse impact.

◗ Use only assessment instruments that are *unbiased* and *fair* to all groups (Chapter 2)

Using unbiased and fair tests will help you select a qualified and diverse workforce. Employment decisions based on tests that are biased are likely to lead to unfair and illegal discrimination against members of the lower scoring groups. You should review the fairness evidence associated with

assessment instruments before selecting tools by examining the test manual and independent test reviews.

◗ Use only *reliable* assessment instruments and procedures (Chapter 3)

If a person takes the same test again, will he or she get a similar score, or a very different score? A reliable instrument will provide accurate and consistent scores. To meaningfully interpret test scores and make useful career or employment-related decisions, use only reliable tools. Test manuals will usually provide a statistic, known as the reliability coefficient, giving you an indication of a test's reliability. The higher the reliability coefficient, the more confidence you can have that the score is accurate.

◗ Use only assessment procedures and instruments that have been demonstrated to be *valid for the specific purpose* for which they are being used (Chapter 3)

Validity is the most important issue in selecting assessment tools. It refers to (1) the characteristic the assessment instrument measures, and (2) how well the instrument measures the characteristic. Validity is not a property of the assessment instrument itself; it relates to how the instrument is being used.

A test's validity is established in reference to a specific purpose; it may not be valid for different purposes. For example, a test that may be valid for predicting someone's "job knowledge," may not be valid for predicting his or her "leadership skills." You must be sure that the instrument is valid for the purpose for which it is to be used. Selecting a commercially developed instrument does not relieve you of this responsibility.

The test manual usually provides a statistic, the validity coefficient, that will give an indication of the test's validity for a specific purpose under specific circumstances. It measures the degree of relationship between test performance and job performance (i.e., job-relatedness of the test).

◗ Use assessment tools that are appropriate for the *target population* (Chapter 3)

An assessment tool is usually developed for use with a specific group; it may not be valid for other groups. For example, a test designed to predict the performance of office managers may not be valid for clerical workers. The skills and abilities required for the two positions may be different, or the reading level of the test may not be suitable for clerical workers. Tests should be appropriate for the individuals you want to test, that is, your target population.

The manual should indicate the group or groups the test is designed to assess. Your target population should be similar to the group on which the test was developed, or normed. In determining the appropriateness of an instrument for your target group, also consider such factors as reading levels, cultural backgrounds, and language barriers.

- **Use assessment instruments for which *understandable and comprehensive documentation* is available (Chapter 5)**

Are the instructions for administration and interpretation understandable? Is the information sufficiently comprehensive to evaluate the suitability of the instrument for your needs? Carefully evaluate the documentation provided by the test publisher to be sure that the tools you select do the job you want them to do and furnish you with the information you need. If the documentation is not understandable or complete, you run the risk of selecting inappropriate instruments.

Test manuals should provide information about both the development and psychometric characteristics of tests. They should cover topics such as procedures for administration, scoring and interpretation, the recommended uses of an instrument, the groups for whom the test is appropriate, and test norms. They should also include a description of the validation procedures used, and evidence of validity, reliability, and test fairness.

- **Ensure that administration staff are *properly trained* (Chapter 6)**

Assessment instruments must be administered properly to obtain valid results. Consult the test publisher and administration manual for guidelines on the qualifications and training required for test administrators. These requirements will vary depending on the nature and complexity of the test. Only suitable staff should be selected. Administrators should be given ample time to learn their responsibilities and should practice by administering tests to other staff before administering tests to applicants. Some test publishers may run training sessions for test administration and interpretation.

Administration staff should also be trained to handle special situations with sensitivity. An example would be responding to a request for accommodation based on a disability.

- **Ensure that *testing conditions are suitable* for all test takers (Chapter 6)**

There are various extraneous influences that may affect the reliability and validity of an assessment procedure. For example, noise in the testing room, poor lighting, inaccurate timing and damaged test equipment may adversely affect test takers. Staff should ensure that the testing environment is suitable and that administration procedures are uniform for all test takers.

- **Provide *reasonable accommodation* in the assessment process for people with disabilities (Chapter 6)**

To ensure that qualified individuals with disabilities have an equal chance to demonstrate their potential, accommodations in the assessment process may be necessary. Under the ADA,

reasonable accommodation may involve ensuring physical accessibility to the test site, modifying test equipment or the testing process, or providing qualified assistance to the test taker. For example, administering a braille version of a test, allowing extra time to complete the test, or supplying a reader may be appropriate. It is important to become familiar with the types of accommodations that can be made without invalidating test results. If reasonable accommodation involving test administration cannot be made, consider alternative assessment strategies.

◗ Maintain assessment instrument *security* (Chapter 6)

All materials used in the assessment process, whether paper-and-pencil or computer-based, must be kept secure. Lack of security may result in some test takers having access to test questions beforehand, thus invalidating their scores. To prevent this, test users should, for example, keep testing materials in locked rooms or cabinets and limit access to those materials to staff involved in the assessment process. Security is also the responsibility of test developers. The security of a test may become compromised over time. To protect security, test developers periodically introduce new forms of tests.

◗ Maintain *confidentiality* of assessment results (Chapter 6)

Assessment results are highly personal. Employers must respect the test taker's right to confidentiality. Assessment results should only be shared with those who have a legitimate need to know. This would include staff involved in interpreting assessment results and making employment decisions. Personal information should not be released to other organizations or individuals without the informed consent of the test taker.

◗ Ensure that scores are *interpreted properly* (Chapter 7)

Tests are used to make inferences about people's characteristics, capabilities, and future performance. The inferences should be reasonable, well-founded, and not based upon stereotypes. If test scores are not interpreted properly, the conclusions drawn from them are likely to be invalid, thus leading to poor decision making.

Ensure that there is solid evidence to justify your test score interpretations and the employment decisions you make based on those scores. The test manual should provide instructions on how to properly interpret test results.

APPENDIX A: Sources of Additional Information on Personnel Assessment

The following list of reference material provides sources of information on specific topics and issues relating to personnel testing and assessment. The main text has referred to many of the publications listed below. Others are included as general reference documents and as recommended readings.

Publications on testing and assessment may be ordered from various online sources, including: http://www.ericae.net/bstore/home2.htm, http://www.unl.edu/buros/catalog.html and http://www.apa.org/books/topiclist-2.html

The Educational Testing Service (ETS) web site www.ets.org/research/index.html contains a variety of downloadable files relating to testing and assessment.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. 1985. *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association.

Anastasi, A. 1988. *Psychological Testing* (6th edition). New York: Macmillan.

Arvey, R.D., and R.H. Faley. 1988. *Fairness in Selecting Employees.* Reading, MA: Addison-Wesley.

Boudreau, J. 1996. *Cumulative Supplement to Employment Testing Manual.* Boston: Warren, Gorham & Lamont.

Bruyère, S.M., and J. O'Keeffe (eds.). 1994. *Implications of the Americans with Disabilities Act for Psychology.* Washington, DC: American Psychological Association.

Bureau of National Affairs. 1990. *The Americans with Disabilities Act: A Practical and Legal Guide to Impact, Enforcement, and Compliance.* Washington, DC: Author.

Bureau of National Affairs Policy and Practice Series. 1992-1995. *Fair Employment Practices Manual #8.* Washington, DC: Author.

Buros Institute of Mental Measurements. Various. *Mental Measurements Yearbook.* Lincoln, NE: University of Nebraska Press.

Buros Institute of Mental Measurements. Various. *Tests in Print.* Lincoln, NE: University of Nebraska Press.

Douglas, J.A., D.E. Feld, and N. Asquith. 1989. *Employment Testing Manual.* Boston, MA: Warren, Gorham & Lamont.

Equal Employment Opportunity Commission. 1978. The Office of Personnel Management, U.S. Department of Justice and U.S. Department of Labor (1979). *Questions and Answers Clarifying and Interpreting the Uniform Guidelines on Employee Selection Procedures.* 29 CFR Part 1607 (1988).

Equal Employment Opportunity Commission. 1978. The Office of Personnel Management, U.S. Department of Justice and U.S. Department of Labor (1979). *Uniform Guidelines on Employee Selection Procedures.* 41 CFR Part 603 (1978).

Equal Employment Opportunity Commission. 1992. *A Technical Assistance Manual on the Employment Provisions (Title I) of the Americans with Disabilities Act.* Washington, DC: U.S. Government Printing Office.

Equal Employment Opportunity Commission. 1992. EEOC Technical Assistance Manual on Employment Provisions of the Americans with Disabilities Act; *ADA Enforcement Guidance: Preemployment Disability Related Questions and Medical Examinations.*

Equal Employment Opportunity Commission and U.S. Department of Justice. 1991. *Americans with Disabilities Act Handbook.* Washington, DC: Author.

French, W.L. 1990. *Human resources management* (2nd edition). Houghton Mifflin Co.: Boston, MA.

Guion, R.M. 1997. *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hogan J., and R. Hogan (eds.) 1984-1990. *Business and Industry Testing: Current Practices and Test Reviews.* Austin, TX: Pro-ed.

Keyser, D.J., and R.C. Sweetland (eds.). 1984-1988. *Test Critiques. (Vols. 1-7).* Kansas City, MO: Test Corporation of America.

Murphy, K.R., and C.O. Davidshofer. 1988. *Psychological Testing: Principles and Applications.* Englewood Cliffs, NJ: Prentice Hall.

Nunnally, J.C. 1978. *Psychometric Theory* (2nd edition). New York: McGraw-Hill.

Society for Industrial and Organizational Psychology, Inc. 1987. *Principles for the Validation and Use of Personnel Selection Procedures* (3rd edition). College Park, MD: Author.

U.S. Department of Justice. 1991. *The Americans with Disabilities Act: Questions and Answers.* Washington, DC: Civil Rights Division, U.S. Department of Justice.

U.S. Department of Labor, Employment and Training Administration. 1993. *JTPA: Improving Assessment: A Technical Assistance Guide.* Washington, DC: Author.

APPENDIX B: Glossary of Assessment Terms

ability test
A test that measures the current performance or estimates future performance of a person in some defined area of cognitive, psychomotor, or physical functioning.

achievement test
A test that measures acquired knowledge or skills, usually as the result of previous instruction.

adverse impact
A situation in which members of a particular race, sex, or ethnic group have a substantially lower rate of selection in hiring, promotion, or other employment decisions.

alternate forms
Two or more forms of a test that are similar in nature and intended to be used for the same purpose.

assessment
Any test or procedure used to measure an individual's employment or career-related qualifications or characteristics.

basic skills test
Assessments of competence in reading, simple mathematics, and other skills that are widely required in training and employment settings.

coaching
Instructional activities designed to improve the test performance of prospective test takers.

compensatory approach
See *counterbalanced approach.*

concurrent validity
See *criterion-related validity*.

construct
A theoretical characteristic or concept (e.g., numerical ability, conscientiousness) that has been constructed to explain observable patterns of behavior.

construct-related validity
The extent to which a test measures a specific theoretical construct, characteristic, or trait. In employment testing, this characteristic should be important for job success. Examples of constructs are mechanical ability and physical endurance.

content-related validity
The extent to which the content of a test samples or represents the subject area or behavior it is intended to measure.

converted score
A raw score that has been converted by numerical transformation (for example, to percentile ranks or standard scores) to facilitate comparison of individual scores with group norms.

correlation
A statistic that indicates the degree to which two variables relate to each other, such as a test score and job performance, or one test with another test.

counterbalanced approach
An approach to personnel assessment that allows high scores in one or more areas to be counterbalanced with low scores in another area.

criterion
A measure of performance, such as productivity rate, accident rate, or supervisory ratings. Test scores are used to predict criteria.

criterion-related validity
The degree to which scores on an assessment instrument correlate with some external criterion, such as job performance. When the assessment instrument and the criterion are measured at about the same time, it is called *concurrent validity*; when the criterion is measured at some future time, it is called *predictive validity*.

derived score
See *converted score*.

equivalent forms
See *alternate forms*.

expectancy table
A table that shows the probability of different criterion outcomes for each test score.

hurdles approach
See *multiple hurdles approach*.

inventory
A questionnaire or checklist that elicits information about an individual in such areas as work values, interests, attitudes, and motivation.

job analysis
A systematic process used to identify the tasks, duties, responsibilities and working conditions associated with a job and the knowledge, skills, abilities and other characteristics required to perform that job.

mean
The average score in a group of scores, computed by adding all the scores and dividing the sum by the number of cases.

median
The middle score in a group of ranked scores. It is the point or score that divides the group into two equal parts. The median is also known as the 50th percentile.

multiple hurdles approach
An approach to personnel assessment that requires a candidate to pass all tests in sequence in order to qualify.

normal curve
A mathematical curve that is the basis of many statistical analyses. The curve is bilaterally symmetrical, with a single bell-shaped peak in the center. Most distributions of human traits, such as height, mathematical ability, and manual dexterity, approximate the normal curve.

norms
Descriptive statistics that are used to summarize the test performance of a specified group, such as a sample of workers in a specific occupation. Norms are often assumed to represent a larger population, such as all workers in an occupation.

parallel forms
See *alternate forms*.

percentile score
The score on a test below which a given percentage of scores fall. For example, a score at the 65th percentile is equal to or higher than the scores obtained by 65% of the people who took the test.

predictive validity
See *criterion-related validity*.

rank ordering
The process of ranking individuals based on their relative test scores, from the highest to the lowest score.

raw score
The obtained score on a test, usually determined by counting the number of correct answers.

reference group
The group of individuals used to develop a test.

reliability
The degree to which test scores are consistent, dependable, or repeatable.

reliability coefficient
A coefficient of correlation that indicates the degree to which test scores are dependable, or repeatable.

standard deviation
A statistic used to describe the variability within a set of scores. It indicates the extent to which scores vary around the mean or average score.

standard error of measurement (SEM)
A statistic that gives an indication of the amount of error in a measurement system. It indicates a range within which a test taker's "true" score is likely to fall.

standard score
A score that describes the location of a person's score within a set of scores in terms of its distance from the mean in standard deviation units.

standardized test
A test developed using professionally prescribed methods that provides specific administration requirements, instructions for scoring and instructions for interpreting scores.

target group
The population or group of individuals whom the employer wishes to assess.

test
Any instrument or procedure that samples behavior or performance. A personnel or employment test is the general term for any assessment tool used to measure an individual's employment qualifications, capabilities, or characteristics.

validity
The degree to which actions or inferences based on test results are meaningful or supported by theory and empirical evidence.

validity coefficient
A numerical index that shows the strength of the relationship between a test score and a criterion, such as job performance.