# Selection Assessment Methods

A guide to implementing
formal assessments to build
a high-quality workforce

**HR**

**SHRM FOUNDATION**
SOCIETY FOR HUMAN RESOURCE MANAGEMENT

Elaine D. Pulakos

# Selection Assessment Methods

A guide to implementing
formal assessments to build
a high-quality workforce

**HR**®

**SHRM FOUNDATION**
SOCIETY FOR HUMAN RESOURCE MANAGEMENT

Elaine D. Pulakos

# Table of Contents

# FORWARD

The SHRM Foundation Board of Directors appreciates how difficult it is for HR practitioners to access current research findings and incorporate them into their own HR practices.

Human resource professionals juggle multiple responsibilities and do not have time to read long research reports, no matter how beneficial. Realistically, most HR practitioners will seek guidance from research findings only if they are presented in a clear, concise and usable format.

To make research more accessible, the SHRM Foundation created this series of reports titled *Effective Practice Guidelines*. The first report on performance management was published in 2004. The Foundation will publish new reports on different HR topics each year. You are now reading the second report in the series: *Selection Assessment Methods.*

Here is the series concept: A subject matter expert with both research and practitioner experience is selected to prepare the guidelines. The author distills the research findings and expert opinion into specific advice on how to conduct effective HR practice. To provide a convenient reference tool, a substantial annotated bibliography is included with each report. We believe this new product presents relevant research-based knowledge in an easy-to-use format. We look forward to your feedback to let us know if we've achieved that goal.

Our author is Dr. Elaine Pulakos, executive vice president and director of the Personnel Decisions Research Institutes (PDRI) Washington, D.C. office. Dr. Pulakos is one of the country's leading experts on selection techniques, both as a researcher and a consultant, and she has provided the very best guidance available on this topic.

Our vision for the SHRM Foundation is: "The SHRM Foundation maximizes the impact of the HR profession on organizational decision-making and performance, by promoting innovation, research and the use of research-based knowledge."

We are confident that this new series of *Effective Practice Guidelines* takes us one step closer to making that vision a reality.

Herbert G. Heneman III, Ph.D.
Director of Research, 2005 SHRM Foundation Board
Professor, School of Business, University of Wisconsin-Madison

# Acknowledgments

The SHRM Foundation wishes to thank the following individuals for reviewing this report, providing feedback and helping to shape the finished product:

# About the Author
# Elaine D. Pulakos, Ph.D.

Elaine Pulakos is executive vice president and director of the Washington, D.C. office of Personnel Decisions Research Institute (PDRI.) PDRI is a premier consulting firm in the field of industrial and organizational psychology. A recognized expert and researcher in the areas of selection and performance appraisal, Dr. Pulakos has over 15 years of experience conducting large-scale job analysis, selection, performance appraisal and career development projects.

A Fellow of the American Psychological Association (APA) and the Society for Industrial and Organizational Psychology (SIOP), Dr. Pulakos is a successful author and has written on the topics of staffing and performance management. She is a past president of SIOP. In addition to authoring numerous publications, Dr. Pulakos recently co-edited two books: *The Changing Nature of Performance: Implications for Staffing, Motivation, and Development* with Daniel Ilgen, and *Implementing Organizational Interventions: Steps, Processes, and Best Practices* with Jerry Hedge.

Dr. Pulakos has spent her career conducting applied research in public and private sector organizations, where she has designed, developed and successfully implemented numerous HR systems including staffing, performance management, and career development and training systems. Dr. Pulakos has also been extensively involved in providing expert advice on EEO-related legal matters and serving as an expert witness and advisor to the Department of Justice, among others. Elaine received her Ph.D. in industrial and organizational psychology from Michigan State University.

# Effective Practice Guidelines: Selection Assessment Methods

Organizations compete fiercely in the war for talent. Many invest an enormous amount of money, time and other resources in advertising and recruiting strategies to attract the best candidates. This is because today's executives understand that one of the most important resources in organizations—if not the most important—is human resources.

Yet, when it comes to actually assessing which job candidates are likely to perform most effectively and make the most significant contributions, a large number of organizations employ rudimentary and haphazard approaches to selecting their workforces. This represents a serious disconnect for organizations that purport to have a strategic focus on increasing their competitive advantage through effective talent management. The disconnect stems from the fact that many organizations fail to use scientifically proven assessments to make selection decisions, even though such assessments have been shown to result in significant productivity increases, cost savings, decreases in attrition and other critical organizational outcomes that translate into literally millions of dollars. Thus, there are real and very substantial bottom-line financial results associated with using effective assessments to guide selection decisions.

One reason why more organizations do not use rigorous assessments to select employees is because many executives and HR professionals have misconceptions about the value of using them. Some of the most common misconceptions are presented below.

---

### Common Misconceptions About Selection Tests[1]

- Myth: Screening applicants for conscientiousness will yield better performers than screening applicants for intelligence.

- Myth: Screening applicants for their values will yield better performers than screening applicants for intelligence.

- Myth: Integrity tests are not useful because job candidates misrepresent themselves on these types of tests.

- Myth: Unstructured interviews with candidates provide better information than structured assessment processes.

- Myth: Using selection tests creates legal problems for organizations rather than helps solve them.

---

[1] Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resources practices: Correspondence between research and practice. *Human Resource Management, 41,* 149-174.

Another reason why formal assessments are not used more in organizations is that there tends to be a lack of knowledge about the types of assessment methods that research has shown to be most effective for identifying who will perform best on a job.[2] This, coupled with the fact that the area of selection testing is inherently technical and difficult to understand, leads many organizational decision makers and HR professionals to shy away from using formal assessments to guide their selection decisions.

A final reason why more organizations do not use effective assessments may be attributable to the multitude of consulting firms selling different selection products and tools. It is important for organizational decision makers and HR practitioners to be educated consumers regarding these products to ensure they are bringing competently developed and effective assessment methods into their organizations.

Most organizations use a funneling approach to selection, where more informal tools and procedures are used initially to reduce the pool of candidates to a manageable number of individuals who may then be put through a more extensive assessment process. Common initial screening devices include resumes, application blanks and reference checks, which are generally used to identify and exclude obvious misfits or poor performers from further consideration. Another initial screening device is the informal meeting or phone interview, which is often used to allow organizational members direct interaction with potential candidates. While these initial screening devices have a useful place in the overall selection process, the focus of this paper is on more formal assessment methods. We specifically focus on those that research has shown to have a proven track record of helping organizations build high-quality workforces by identifying individuals who will perform effectively, achieve results and make important contributions on the job.

This report has three important goals:
■ Present and summarize what is known from the research literature about the value of different types of formal assessment methods that are used to select employees in organizations.
■ Remove some of the mystique, complexity and confusion that can drive HR professionals away from implementing formal assessment methods by providing brief tutorials on the most important technical, legal and measurement issues inherent in selection testing.
■ Provide a useful roadmap to help make decisions about what assessment methods are most useful and practical in different situations.

---

[2] Ryan, A. M., & Tippins, N. T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management, 43,* 305-318.

The report is organized into four major parts. First, a brief discussion of job analysis is presented. Job analysis is important because it provides information that is necessary to make decisions about what types of assessment methods are most appropriate for a given job. Next, to familiarize readers with the array of assessment methods that are available, the second part of the paper provides brief descriptions and examples of the most common tools that research has shown to be effective in predicting who will perform successfully on a job. The third part of the paper focuses on important criteria to consider in evaluating assessment methods and provides guidance on how to make rational choices among the available alternatives. The final part discusses other issues that are relevant to using assessments, including the mode of administration, utility and legal considerations.

## Job Analysis

There are numerous different types of formal assessments that organizations can use to select employees. The first step in developing or selecting an assessment method for a given situation is to understand what the job requires employees to do and, in turn, what knowledge, skills and abilities (KSAs) individuals must possess in order to perform the job effectively. This is typically accomplished by conducting a job analysis.[3][4] The portion of a job analysis that focuses on what the job requires individuals to do is often referred to as a job-oriented or task-based job analysis, which involves a comprehensive list of work tasks that individuals are required to perform on the job.

---

**Job-Oriented Job Analysis: Sample Tasks for an Investigator Job**

- Provide testimony by stating facts and answering questions.
- Gather and review pertinent information to obtain evidence or develop background information on subjects.
- Integrate diverse information to uncover relationships between individuals, events or evidence.
- Work in a team environment as a team member or leader.
- Calm and reassure victims or distressed others in tense situations.
- Perform a variety of public service functions to enhance the image of the organization.

---

The portion of a job analysis that focuses on the KSAs that workers must possess to be effective is often referred to as a worker-oriented or KSA-based job analysis. Typically, a job analyst first identifies the tasks that workers are required to perform on the job and

---

[3] Gael, S. (Ed.). (1988). *The job analysis handbook for business, industry, and government.* (Vols. 1 and 2). New York: Wiley.

[4] Brannick, M.T., & Levine, E. L. (2002). *Job analysis*. Thousand Oaks, CA: Sage.

then identifies the KSAs that are needed to effectively perform those tasks. For example, the first task above is "Provide testimony by stating facts and answering questions." KSAs that would be required to perform this task include (1) the ability to speak clearly, self-confidently and concisely using voice inflection, gestures and eye contact for emphasis; (2) the ability to maintain a professional demeanor and appearance at all times; and (3) the ability to remain calm and levelheaded under stress.

---

**Person-Oriented Job Analysis: Sample KSAs for an Investigator Job**

■ Ability to speak clearly, self-confidently and concisely using voice inflection, gestures and eye contact for emphasis.

■ Ability to think critically, questioning assumptions and identifying merits and deficiencies in logic.

■ Ability to gain cooperation from other individuals or organizations.

■ Ability to maintain a professional demeanor and appearance at all times.

■ Ability to remain calm and levelheaded under stress.

■ Knowledge of investigative techniques and procedures.

---

While an in-depth discussion of job analysis procedures is beyond the scope of this report, the major steps involved in performing a job analysis for the purpose of developing or selecting assessment methods are shown below.

**STEP 1**
Observe or interview job experts to develop a list of tasks performed on the job.

**STEP 2**
Observe or interview job experts to develop a list of the KSAs workers must possess to perform job tasks effectively.

**STEP 3**
Survey a sufficiently large and representative sample of job incumbents or their supervisors, asking them to rate which job tasks and KSAs are most critical for effective job performance.

**STEP 4**
Analyze the survey data to prioritize the most critical tasks and KSAs for the job.

**STEP 5**
Select or develop assessments that measure the most critical tasks or KSAs that a worker must possess upon entry to the job.

Job analysis information is used as a basis for developing assessments. Specifically, assessments are developed to measure the most critical tasks or KSAs resulting from the job analysis for a given job. Some assessments involve work samples that simulate job tasks and require candidates to demonstrate that they can perform these tasks effectively. Job-oriented or task-based job analysis data are used as a basis for developing these types of assessments because they focus directly on assessing how well job candidates can perform critical work tasks.
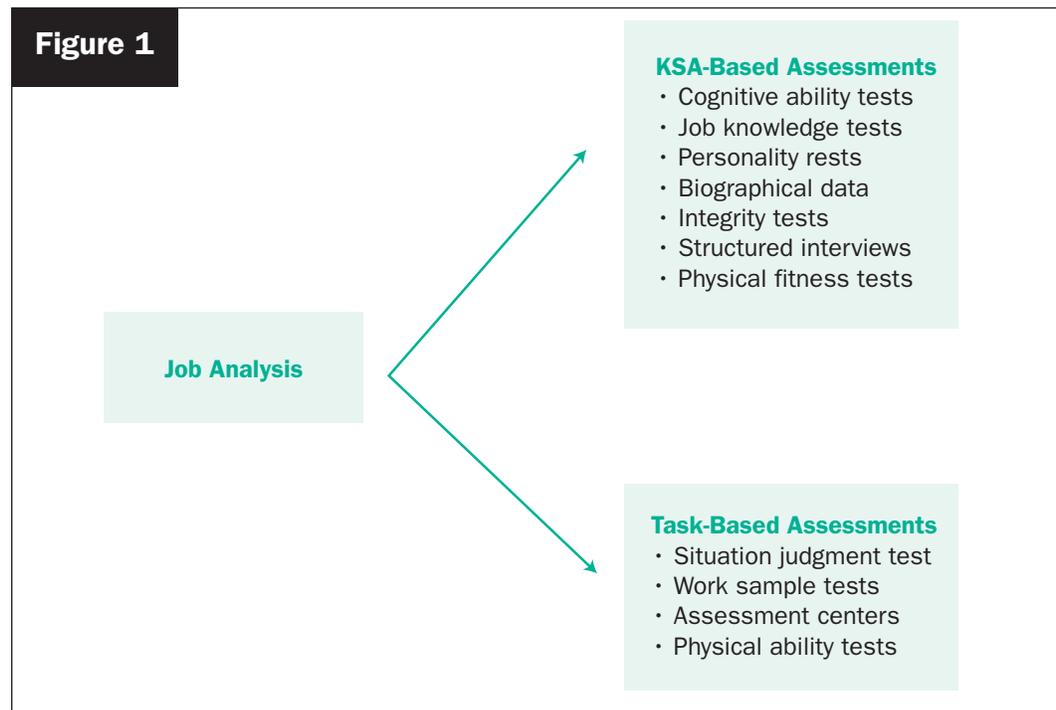
Other assessment methods focus on measuring KSAs that are required to perform job tasks effectively, such as various mental abilities, physical abilities or personality traits, depending on the job's requirements. If one were selecting a manager, for example, it would be important to assess whether candidates could solve complex business problems, be decisive and communicate effectively. Alternatively, if one were selecting an administrative assistant, KSAs such as the ability to perform work conscientiously and the ability to perform work with speed and accuracy would be much more important for identifying capable candidates. Worker-oriented or KSA-based job analysis data are used as a basis for developing assessment methods that focus on a job candidate's underlying abilities to perform important work tasks.

## Assessment Methods

This section of the paper describes the various assessment methods that can be used by organizations. Figure 1 shows which assessment methods are predominantly task-based and which are predominantly KSA-based. The methods discussed here can be used for internal or external selection. Internal selection refers to situations where an organization is hiring or promoting from within, whereas external selection refers to situations where an organization is hiring from the outside. While some assessment methods are used more commonly for external selection (e.g., cognitive ability tests, personality tests, integrity tests), there are numerous examples of organizations that have used one or more of the following tools for internal selection, external selection or both.

**Figure 1**

**Job Analysis**

**KSA-Based Assessments**
· Cognitive ability tests
· Job knowledge tests
· Personality rests
· Biographical data
· Integrity tests
· Structured interviews
· Physical fitness tests

**Task-Based Assessments**
· Situation judgment test
· Work sample tests
· Assessment centers
· Physical ability tests

**Cognitive Ability Tests.** These assessments measure a variety of mental abilities, such as verbal and mathematical ability, reasoning ability and reading comprehension. Cognitive ability tests have been shown to be extremely useful predictors of job performance and thus are used frequently in making selection decisions for many different types of jobs.[5][6][7] Cognitive ability tests typically consist of multiple-choice items that are administered via a paper-and-pencil instrument or computer.

Some cognitive ability tests contain test items that tap the various abilities (e.g., verbal ability, numerical ability, etc.) but then sum up the correct answers to all of the items to obtain a single total score. That total score then represents a measure of general mental ability. If a separate score is computed for each of the specific types of abilities, then the resulting scores represent measures of the specific mental abilities.

[5] Hunter, J. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362.

[6] Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g. Journal of Applied Psychology, 79,* 518-524.

[7] Gottredson, L. S. (Ed.). (1982). The *g* factor in employment. *Journal of Vocational Behavior, 29(3).*

**Sample Cognitive Ability Test Items**

Verbal ability

*Innocuous means the same as:*
   a. Harmless
   b. Preventative
   c. Distasteful
   d. Futile

Numerical ability

*16% of 62.5 is*
   a. .844
   b. 8.44
   c. .084
   d. 8.4

Reasoning ability

*1 3 2 4 3 5 4 6 5 ___*
   a. 4
   b. 5
   c. 6
   d. 7

Reading ability

*In American politics, the concern of each party is to win. This requires gaining the support of many people with differing views. Political parties often have to build into their programs potentially conflicting objectives that speak to the needs of these different groups in order to win. As a result, the platforms of major parties typically reflect:*
   a. Unified principles
   b. Prejudice
   c. Compromise
   d. Disagreement

**Job Knowledge Tests.** These assessments measure critical knowledge areas that are needed to perform a job effectively.[8] Typically, the knowledge areas measured represent technical knowledge. Job knowledge tests are used in situations where candidates must already possess a body of knowledge prior to job entry. Job knowledge tests are not appropriate to use in situations where candidates will be trained after selection on the knowledge areas they need to have. Like cognitive ability tests, job knowledge tests typically consist of multiple-choice items administered via a paper-and-pencil instrument or a computer, although essay items are sometimes included in job knowledge tests.

[8] Hunter, J. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362.

---

**Sample Job Knowledge Test Item**

*The principle of the "lever" is essential to the use of a:*
    a. Hydraulic jack
    b. Plow
    c. Auto steering wheel
    d. Forklift

---

**Personality Tests.** Personality tests that assess traits relevant to job performance have been shown to be effective predictors of subsequent job performance.[9][10] The personality factors that are assessed most frequently in work situations include conscientiousness, extraversion, agreeableness, openness to experience and emotional stability.[11][12] Research has shown that conscientiousness is the most useful predictor of performance across many different jobs, although some of the other personality factors have been shown to be useful predictors of performance in specific types of jobs.[13] Personality inventories consist of several multiple-choice or true/false items measuring each personality factor. Like cognitive ability and knowledge tests, they are also administered in a paper-and-pencil or computer format.

---

[9] Raymark, M. J., Schmit, M. J., & Guion, R. M. (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology, 50,* 723-736.

[10] Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

[11] Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 91,* 1-26.

[12] Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13,* 653-665.

[13] Hough, L. M. (1992). The big five personality variables-construct confusion: Description versus prediction. *Human Performance, 5,* 135-155.

---

**Sample Personality Test Items**

*It does not make sense to work hard on something if no one will notice.*
    a. Definitely true
    b. Somewhat true
    c. Neither true nor false
    d. Somewhat false
    e. Definitely false

*I tend to let others do most of the talking in conversations.*
    a. Definitely true
    b. Somewhat true
    c. Neither true nor false
    d. Somewhat false
    e. Definitely false

*I have remained calm in situations where others have become upset.*
    a. Definitely true
    b. Somewhat true
    c. Neither true nor false
    d. Somewhat false
    e. Definitely false

---

**Biographical Data.** Biographical data (biodata) inventories, which ask job candidates questions covering their background, personal characteristics or interests, have been shown to be effective predictors of job performance.[14][15] The idea is that the best predictor of future performance is past performance. Thus, biodata questions focus on assessing how effectively job candidates performed in the past in areas that are identical or highly related to what they will be required to do on the job for which they are being considered. For example, the item on page 10 about the number of volunteer organizations to which one belonged could be used to assess one's willingness to volunteer to pitch in and help others. The second item could be used to assess one's responsibility and independence. Biographical inventories consist of multiple-choice items that are also administered via paper-and-pencil or computer formats.

---

[14] Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.). (1994). *Biodata handbook.* Palo Alto, CA: CPP Books.

[15] Shoenfeldt, L. F. (1999). From dustbowl empiricism to rational constructs in biodata. *Human Resource Management Review, 9,* 147-167.

> **Sample Biographical Inventory Items**
>
> *To approximately how many volunteer organizations do you belong?*
>     a. 0
>     b. 1
>     c. 2 to 4
>     d. 5 or more
>
> *Where did most of your spending money come from during your high school years?*
>     a. Allowance from family
>     b. Own earnings
>     c. Partly allowance, partly earnings
>     d. Other sources
>     e. Had no spending money

Another form of a biodata inventory is an instrument called an "accomplishment record." With this type of assessment, candidates prepare a written account of their most meritorious accomplishments in key skill and ability areas that are required for a job (e.g., planning and organizing, customer service, conflict resolution[16]). The candidate also provides the name of an individual, such as a past supervisor, who can verify the accomplishment. Evaluators are trained to score the accomplishments in a consistent manner using standard rating criteria. These types of assessments have been shown to be effective predictors of subsequent job performance.[17]

**Integrity Tests.** Integrity tests measure attitudes and experiences that are related to an individual's honesty, trustworthiness and dependability.[18] [19] [20] [21] Like many of the tests discussed here, integrity tests are typically multiple-choice in format and administered via a paper-and-pencil instrument or a computer.

---

[16] Hough, L. M. (1984). Development and evaluation of the "accomplishment record" methods of selecting and promoting profession-als. *Journal of Applied Psychology, 69,* 135-146.

[17] Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three 'alternative' selection measures. *Personnel Psychology, 36,* 261-276.

[18] Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implica-tions for personnel selection and theories of job performance. *Journal of Applied Psychology* (monograph), 78, 531-537.

[19] Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependabil-ity, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49,* 787-829.

[20] Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist, 49,* 112-119.

[21] Goldberg, L. R., Grenier, R. M., Guion, L. B., Sechrest, L. B., & Wing, H. (1991). *Questions used in the prediction of trustworthiness in pre-employment selection decisions: An APA Task Force Report.* Washington, D.C.: American Psychological Association.

**Sample Integrity Test Items**

*It is alright to misrepresent the truth if being completely honest will create problems that small "white lies" can solve.*
 a. Strongly disagree
 b. Disagree
 c. Neither agree nor disagree
 d. Agree
 e. Strongly agree

*As long as people follow the spirit of policies and regulations, they don't need to follow them exactly.*
 a. Strongly disagree
 b. Disagree
 c. Neither agree nor disagree
 d. Agree
 e. Strongly agree

**Structured Interviews.** The interview is the most common selection device used in organizations. Most selection interviews are unstructured. That is, the questions to be asked are left up to the interviewer to decide, and there are no agreed-upon standards for evaluating an applicant's performance during the interview. Research has shown that unstructured interviews are not particularly useful for predicting job performance.

Structured interviews, on the other hand, consist of a specific set of questions that are designed to assess critical KSAs that are required for a job.[22][23][24][25] Structured interview questions can be developed to assess almost any KSA, but they are used most frequently to assess softer skills such as interpersonal skills, communication skills, leadership, planning, organizing and adaptability, among others. An important characteristic of an effective structured interview is that it provides standardized rating criteria to help interviewers judge the quality and effectiveness of the responses provided by the interviewee. To work well, interviewers must be trained in how to administer the structured interview properly, probe for additional information and apply the rating criteria accurately and systematically in evaluating job candidates.

[22] Eder, R. W., & Ferris, G. R. (Eds.). (1989). *The employment interview: Theory, research, and practice.* Newbury Park, CA: Sage Publishing.

[23] Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 48,* 289-308.

[24] Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50,* 655-702.

[25] Judge, T. A., Higgins, C. A, & Cable, D. M. (2000). The employment interview: A review of recent research and recommendations for future research. *Human Resource Management Review, 10,* 383-406.

> **Sample Structured Interview Question and Rating Criteria**
>
> *Tell me about a time when you were able to establish rapport with someone when the situation made it difficult to do so. What were the circumstances? What did you do? What were the results?*
>
> | 1 | 2 | 3 | 4 | 5 |
> |---|---|---|---|---|
> | Low | | Moderate | | High |
>
> **Low**
> - Made minimal attempts to understand the person's perspective.
> - Developed only a surface-level relationship in a simple situation.
>
> **Moderate**
> - Attempted to understand the other person's perspective.
> - Developed a positive working relationship with the person in a moderately difficult situation.
>
> **High**
> - Effectively reached out to the person and actively sought to understand the person's perspective.
> - Developed an extremely positive relationship with the person in an extremely difficult or contentious situation.

**Physical Fitness Tests.** Physical fitness tests are used in some selection situations. These tests require candidates to perform general physical activities to assess one's overall fitness, strength, endurance or other physical capabilities necessary to perform the job.

> **Sample Physical Fitness Tests**
> - Run a mile.
> - Lift 20 lb. weights for 50 repetitions.

**Situational Judgment Tests.** Situational judgment tests provide job candidates with situations that they would encounter on the job and viable options for handling the presented situations. [26][27][28] Depending on how the test is designed, candidates are asked to select the most effective or most and least effective ways of handling the situation from the response options provided. Situational judgment tests are more complicated to develop than many of the other types of assessments discussed previously. This is because there is more inherent difficulty in developing scenarios with several likely response options that are all viable, but, in fact, some are reliably rated as being more effective than others. Situational judgment tests are typically administered in written or videotaped form, with responses collected either in a paper-and-pencil test booklet or on a computer.

---

[26] Motowidlo, S. J., Dunnette, M. D., & Carter G. (1990). An alternative selection procedure: A low fidelity simulation. *Journal of Applied Psychology, 75,* 640-647.

[27] Chan, D., & Schmitt, N. (1997). Video-based versus paper and pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82,* 143-159.

[28] Weichmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410-417.

**Sample Situational Judgment Test Items**

*You are currently working on several tasks, all of which are pressing. Your supervisor asks you to work on another assignment with an immediate deadline. She asks you to phone companies to obtain financial data. The list of companies is long and not yet complete. You would…*
- a. Describe the pressing deadlines in which you are already involved and ask your supervisor to assign the new task to a less busy colleague.
- b. Complete those assignments on which you are already working, then concentrate on phoning the companies.
- c. Work on your other assignments and begin phoning companies only when you receive a complete list.
- d. Immediately phone the companies currently listed, then continue working on your other assignments; make the other phone calls as you are notified of company names.

*You have just prepared a report that you have checked and rechecked for accuracy. Before you attend a meeting at which you will submit your report, you review the typed version and note many serious errors. You would…*
- a. Show the original and the typed version to the person in charge of typing and demand that the errors be changed before the meeting.
- b. Present the report at the meeting, point out the errors and state they were due to the typist.
- c. Present the errors to the typist, ask him or her to make the corrections and explain to individuals at the meeting that your report is still being typed.
- d. Present your report at the meeting and make no mention of the errors but notify attendees of corrections after the meeting.

**Work Sample Tests.** Work sample tests consist of tasks or work activities that mirror the tasks that employees are required to perform on the job.[29][30][31] Work sample tests can be designed to measure almost any job task but are typically designed to measure technically-oriented tasks, such as operating equipment, repairing and troubleshooting equipment, organizing and planning work, and so forth.

Work sample tests typically involve having job applicants perform the tasks of interest while their performance is observed and scored by trained evaluators. Similar to job knowledge tests, work sample tests should only be used in situations where candidates are expected to know how to perform the tested job tasks prior to job entry. If training on how to perform the job will be provided after selection, work sample assessments would not be appropriate to use.

---

[29] Asher, J. J., & Sciarrino. J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology, 27,* 519-533.

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

[31] Howard, A. (1983). Work samples and simulations in competency evaluation. *Professional Psychology: Research and Practice, 14,* 780-796.

---

**Sample Work Sample Tests[32]**

Mechanic
- Repairing a problem on a car.
- Reading a blueprint.

Clerical
- Typing test.
- Proofreading.

Cashier
- Operating a cash register.
- Counting money and totaling balance sheet.

Airline Pilot
- Pilot simulator.
- Rudder control test.

Taxi Cab Driver
- Driving test.
- Street knowledge test.

Computer Programmer
- Programming and debugging test.
- Hardware replacement test.

---

**Assessment Centers.** An assessment center is a type of work sample test that is typically focused on assessing higher-level managerial and supervisory competencies.[33][34][35] Candidates are asked to complete a series of exercises that simulate actual situations, problems and tasks that they would face on the job for which they are being considered, and they are asked to handle these as if they were in the real situation. In this sense, assessment centers are similar to the work sample tests described previously.

Assessment centers usually last at least a day and up to several days. They typically include role-play exercises, in-basket exercises, analytical exercises and group discussion exercises. Trained assessors observe the performance of candidates during the assess-

---

[32] Heneman, H. G., III, & Judge, T. A. (in press). *Staffing organizations, 5E.* Middleton, WI: Mendota House, and New York: McGraw-Hill.

[33] Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analyses of assessment center validity. *Journal of Applied Psychology, 72,* 493-511.

[34] Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance.* New York: Academic Press.

[35] Thornton, G. C., III. (1992). *Assessment centers in human resources management.* Addison-Wesley.

ment process and evaluate them on standardized rating. Some assessment centers also include other types of assessment methods, such as cognitive ability, job knowledge and personality tests. It should be noted that assessment centers are not only used for selection purposes, but can also be used to provide comprehensive development feedback to participants.

---

**Sample Assessment Center Process**

Candidate Orientation

■ Role-Play: Candidate prepares for and conducts a counseling session with a subordinate who is not performing well on the job. Trained assessors play the role of the subordinate and evaluate the candidate's performance.

■ In-Basket Exercise: Candidate completes a mock "in-basket" where he or she responds to phone messages, e-mails, memoranda, reports and other items that require action to be taken. Sometimes in-basket exercises also include an interview with trained assessors where the candidate explains the rationale for his or her actions.

■ Case Analysis: A business problem is presented to the candidate that he or she has to analyze and prepare a written plan discussing the actions to be taken. Trained assessors evaluate the written products.

■ Role-Play: Candidate is given a set of facts and must prepare a persuasive oral presentation regarding a situation. Following the presentation, the candidate responds to questions. Trained assessors play the role of audience members who ask the questions and evaluate the candidate's performance.

■ Leaderless Group Discussion: A small group of candidates is given a problem to work on that is similar to what they would encounter on the job. As they work to resolve the problem, trained assessors observe and evaluate their performance.

Following all of the exercises, assessors discuss their evaluations of the candidates, decide on consensus evaluations of their performance and prepare feedback.

---

**Physical Ability Tests.** Physical ability tests are used regularly to select workers for physically demanding jobs, such as police officers and firefighters.[36][37][38] These tests are similar to work sample tests in that they typically require candidates to perform a series of actual job tasks to determine whether or not they can perform the physical requirements of a job. Physical ability tests are often scored on a pass/fail basis. To pass, the complete set of tasks that comprise the test must be properly completed within a specified timeframe.

---

[36] Hogan, J. (1991). Physical abilities. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology,* vol. 2, pp. 753-831. Palo Alto, CA: Consulting Psychologists Press.

[37] Campion, M. A. (1983). Personnel selection for physically demanding jobs: Review and recommendations. *Personnel Psychology, 36,* 527-550.

[38] Blakley, B. R., Quinones, M. S., Crawford, M. S., & Jago, I. A. (1994). The validity of isometric strength tests. *Personnel Psychology, 47,* 247-274.

While perhaps a subtle distinction, physical ability tests usually replicate actual job tasks and evaluate whether individuals can complete these within specified timeframes that mirror how quickly they would need to perform them on the job. Alternatively, physical fitness tests (discussed previously) do not replicate job tasks, per se, but rather require candidates to perform more general physical activities (e.g., running a mile) to assess their overall fitness.

---

**Sample Physical Ability Tests**

Firefighter
- Climb a ladder while carrying equipment.
- Drag a hose.
- Carry a person down from a building.

Warehouse Worker
- Lift materials of weight required on job.
- Drag materials for distances required on job.
- Roll barrels.
- Carry materials of weight and distance required on job.

---

## Criteria for Selecting and Evaluating Assessment Methods

Properly identifying and implementing formal assessment methods to select employees is one of the more complex areas for HR professionals to learn about and understand. This is because understanding selection testing requires knowledge of statistics, measurement issues and legal issues relevant to testing. Accordingly, this section provides guidelines and criteria to help HR professionals make informed decisions about what assessment methods to implement in their organizations. The table on page 17 lists the assessment methods discussed previously and shows their standing on four important criteria:

- **Validity**—the extent to which the assessment method is useful for predicting subsequent job performance.
- **Adverse impact**—the extent to which protected group members (e.g., minorities, females and individuals over 40) score lower on the assessment than majority group members.
- **Cost**—both to develop and to administer the assessment.
- **Applicant reactions**—the extent to which applicants react positively versus negatively to the assessment method.

For example, the first entry in the table is cognitive ability tests. On the positive side, this type of assessment is high on validity and low on costs. However, it is also high on adverse impact, and applicant reactions are only moderately favorable. Thus, while

cognitive tests are inexpensive and very useful for predicting subsequent job performance, minorities score significantly lower on them than whites.

It is important for HR professionals to understand the implications and tradeoffs involved in using different types of assessment methods. There is no simple, formulaic approach for selecting "one best" assessment method, because all of them have advantages and disadvantages. The following sections define and discuss the four criteria for evaluating assessment methods in detail to help HR practitioners make good decisions about which methods will be most appropriate and practical for their situations.

| Evaluation of Assessment Methods on Four Key Criteria | | | | |
|---|---|---|---|---|
| Assessment Method | Validity | Adverse Impact | Costs (Develop/ Administer) | Applicant Reactions |
| Cognitive ability tests | High | High (against minorities) | Low/low | Somewhat favorable |
| Job knowledge tests | High | High (against minorities) | Low/low | More favorable |
| Personality tests | Low to moderate | Low | Low/low | Less favorable |
| Biographical data inventories | Moderate | Low to high for different types | High/low | Less favorable |
| Integrity tests | Moderate to high | Low | Low/low | Less favorable |
| Structured interviews | High | Low | High/high | More favorable |
| Physical fitness tests | Moderate to high | High (against females and older workers) | High/high | More favorable |
| Situational judgment tests | Moderate | Moderate (against minorities) | High/low | More favorable |
| Work samples | High | Low | High/high | More favorable |
| Assessment centers | Moderate to high | Low to moderate, depending on exercise | High/high | More favorable |
| Physical ability tests | Moderate to high | High (against females and older workers) | High/high | More favorable |

Note: There was limited research evidence available on applicant reactions to situational judgment tests and physical ability tests. However, because these tests tend to appear very relevant to the job, it is likely that applicant reactions to them would be favorable.

### Validity

The most important consideration in evaluating an assessment method is its validity.

For the present purposes, validity refers to whether or not the assessment method provides useful information about how effectively an employee will actually perform once she or he is hired for a job. Validity is the most important factor in considering whether or not to use an assessment method, because an assessment that does not accurately identify who will perform effectively on a job has no value to the organization. There are two major forms of validity: criterion-related validity and content validity.

**Criterion-Related Validity.** A simple example will illustrate how criterion-related validity can be established. Assume that a sales job requires employees to have a high level of customer service orientation, and an organization decides to implement a selection test that assesses prospective applicants on their customer service skills. In order to show that the customer skills assessment is a valid predictor of performance, it must be shown that individuals who score higher on the assessment perform better on the job and individuals who score lower on the assessment perform less well on the job. Thus, validity in this case would be defined as a meaningful relationship between how well people performed on the assessment and how well they subsequently performed on the job.

To measure the extent of the relationship between performance on a test and performance on a job, a correlation coefficient is used. A correlation coefficient is a statistical measure that indicates the strength of the relationship between scores on an assessment and scores on a job performance measure, such as performance appraisal ratings. Correlations range from 0 (meaning there is no relationship between how well individuals perform on an assessment and how well they perform on the job) to 1.00 (meaning there is a perfect relationship between how well individuals perform on an assessment and how well they perform on a job—the highest scorer on the assessment receives the highest performance appraisal score, the second highest scorer on the assessment receives the second highest performance appraisal score and so forth).

Correlation coefficients that measure the validity of assessment methods never reach 1.00, because performance on an assessment can never perfectly predict how well individuals will perform on a job. Other factors inevitably influence job performance, such as an individual's motivation, relationships with co-workers and supervisors and a myriad of other factors. Assessment methods that organizations typically use tend to have validities (or correlation coefficients) in the .30 to .50 range. Although these are less than the maximum possible validity of 1.00, assessments with validities in this range nonetheless provide very useful and valuable information for making selection decisions.

To summarize criterion-related validity research, an important technique known as meta-analysis was created. Meta-analysis calculates the average criterion-related validity across different research studies for a given assessment method. The results of meta-analyses were used to denote the "low," "moderate" and "high" levels of criterion-related validity that are reported in the table on page 17 for the different assessment methods. "Low" validities would be around .20 or less, and "moderate" validities would be in the .20 to .40 range. "High" validities mean that correlations between performance on an assessment and subsequent job performance tend to be in the .40 to .50 range.

Although higher validities mean that the assessment is doing a better job of predicting subsequent job performance, even assessments with low to moderate validities provide valuable information for making selection decisions. Further, if two assessments that measure very different things are used together—for example, cognitive ability and personality—they cumulatively combine to produce a higher level of overall validity than each would yield separately. Thus, use of a measure with relatively low validity can still add substantially to the prediction of job performance when it is coupled with other measures.

**Content Validity.** The content validity approach to validation involves demonstrating that an assessment provides a direct measure of how well candidates will actually perform the job. This type of validation requires thoroughly analyzing the job to identify the tasks that are performed and the KSAs that candidates must possess to perform those tasks effectively. The job analysis information is then translated into work sample tasks that mirror the tasks candidates are required to perform on the job. Content validity is established through a series of expert judgments, which document that the assessment comprehensively measures the content of the job.

**Criterion-Related Versus Content Validation.** Criterion-related validity can be used to evaluate the validity of any assessment where individuals receive scores that reflect how well they perform on the test, and these scores are subsequently shown to relate to how well they perform on the job. Content validation can only be used to validate assessments that provide a direct measure of how well candidates perform job tasks (or, in other words, the content of the job), such as work sample tests. Content validation is not an appropriate validation strategy for cognitive ability tests, personality tests, biodata inventories or integrity tests. This is because these tests measure abilities that are thought to be related to job performance but do not measure actual job performance.

Criterion-related validity evidence is a stronger form of validity evidence than content validity. Thus, it is more desirable to obtain if it is possible to conduct a successful criterion-related study. There are various circumstances that must exist in an organization

for a criterion study to have the possibility of being successful. For example, accurate, unbiased performance measures must be available. Unfortunately, performance appraisal ratings, which are the most commonly used performance measures, can be inaccurate and often fail to effectively discriminate between an organization's most and least effective performers. Because criterion-related validity studies are complicated and their success is contingent on a number of factors, it is best to confer with an experienced professional in determining what type of validation approach will be most practical and effective in a given situation. More information regarding the various strategies for validating assessments can be found in Heneman and Judge (2006).

## Adverse Impact

Another concept that is critical to understand in order to evaluate assessment methods is adverse impact. Adverse impact can occur against protected demographic groups, such as African Americans, Hispanics, females and individuals over 40, when certain types of assessment methods are used for selection. Adverse impact results in a disproportionately small number of individuals in a protected group versus the majority group being selected for a job. For example, adverse impact against females is observed frequently when physical ability tests that measure upper body strength are used for selection in physically demanding jobs. Other types of assessments (e.g., tests of general cognitive ability), likewise, tend to systematically produce adverse impact against minority groups.

Adverse impact is examined by comparing the proportion of majority group members who are selected for a job to the proportion of protected group members who are selected. Continuing with the example of males versus females being assessed for their upper body strength, suppose that 50 females and 50 males are examined. Further assume that 35 females pass the assessment and are selected, and 45 males pass the assessment and are selected.

■ The proportion of females passing the test is 30/50 or 60%.
■ The proportion of males passing the test is 45/50 or 90%.
■ The proportion of females passing the test compared with males passing the test is 60%/90% or 67%.

Adverse impact exists if the proportion of protected group members selected is less than 80% of the proportion of majority group members selected. In this case, because the ratio of females selected to males selected is only 67%, it can be concluded that the test is producing an adverse impact against females.

While organizations are and should be interested in selecting the highest quality workforce possible, many are also concerned about selecting a diverse workforce and not

using measures that will systematically produce adverse impact against protected groups. At the same time, if there are important job requirements, such as sufficient upper body strength to perform a firefighter job, an organization would be remiss in not considering this factor in making selection decisions, even if it means that a disproportionately small proportion of females will be hired.

If an assessment method is shown to produce adverse impact and the organization wishes to continue the use of that assessment, there are legal requirements to ensure that the method must have demonstrated validity. If an organization uses an assessment that produces adverse impact without the validity evidence, the organization will be vulnerable to legal challenges against which it will not be able to prevail. While evidence of validity can be used to justify and defend the use of measures that produce an adverse impact, many organizations nonetheless attempt to mitigate the adverse impact produced by their assessment methods to the extent possible in order to minimize potential legal issues and lack of diversity concerns.

Because adverse impact analyses reflect the proportion of majority versus protected group members who are ultimately selected for a job, they cannot be computed until after the assessment process is complete and final selection decisions are made. This is obviously very late for organizational decision makers to realize that the assessment may have undesirable levels of adverse impact.

For this reason, researchers and practitioners often examine other statistics that can be calculated much earlier in the process to determine the likelihood that an assessment method will produce adverse impact. Specifically, one can compare the average scores that different demographic group members receive on an assessment that is being considered for implementation. Continuing with our upper body strength test example, this would be accomplished by calculating the average score for the group of 50 females who took the test and the average score for the group of 50 males who took the test. These average scores would then be transformed into a statistic that represents the difference between how the two groups performed on the test. This statistic is commonly referred to as either an "effect size" or a "group difference in standard deviation units."

Typical "effect sizes" range from 0, indicating no difference on average in how two groups performed on an assessment, to 1.00 or more, indicating a very large difference in how the two groups performed. Effect sizes in the .10 to .30 range are considered small, those in the .30 to .70 range are considered moderate, and those above .70 are considered large. All else being equal, an effect size of .70 to 1.00 or more on an assessment can be expected to produce a large adverse impact in the final selection decisions. Even smaller effect sizes (e.g., in the .30 to .40 range) can produce adverse impact in final selection decisions. It is important to understand how to interpret an effect size

(or group difference in standard deviation units), because these numbers are typically reported in test manuals or in the research literature when different types of assessments are evaluated.

### Tradeoffs Between Validity and Adverse Impact

HR professionals, along with organizational decision makers, need to decide what levels of validity and adverse impact are desirable in their situations. If the goal is to obtain the highest quality workforce possible, assessments should be used that provide the highest levels of validity. However, as is shown in the table on page 17, many of the assessments that produce the highest levels of validity also produce the highest levels of adverse impact. Such assessments tend to be those that assess candidates' abilities to perform technical aspects of a job, such as general mental ability tests, job knowledge tests and other similar types of assessments. Assessments that tend to produce lower levels of adverse impact and somewhat lower but still useful levels of validity are those that assess softer skills, such as communication skills, interpersonal skills, helping and supporting behaviors, and initiative, among others.

Researchers have placed a great deal of emphasis on investigating different strategies for reducing adverse impact without reducing validity, with some success. One promising strategy for reducing adverse impact is simply to recruit more qualified protected group candidates. Unfortunately, this can be easier said than done if many organizations are competing to recruit the same highly qualified individuals.

Another strategy that has been shown to be successful for mitigating adverse impact and enhancing validity is to assess the full array of KSAs that are required on a job. Sometimes professionals are tempted to focus assessments on the technical skills that are required, because these are so obviously related to effective performance. Research has shown, however, that the validity of an assessment process can be enhanced and adverse impact reduced by assessing a comprehensive array of skills and abilities that are related to both technical task performance and contextual job performance. Technical task performance and contextual job performance are two distinct components that comprise overall job performance.[39]

As suggested by the title, technical task performance refers to how well one performs the technical aspects of the job, while contextual performance concerns the extent to which employees engage in behaviors that demonstrate good organizational citizenship, such as helping co-workers when needed, putting in whatever effort is required to get the job done and so forth. Assuming a set of assessments is designed to predict

---

[39] Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W. Borman and Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

both technical and contextual performance, and if contextual performance is weighted more heavily than technical performance, then reductions in adverse impact and increases in overall validity have been observed. Because of the positive benefits for both validity and adverse impact, this is the recommended strategy to pursue in selecting or developing an assessment process.

There are several strategies for reducing adverse impact that the research literature has shown to be ineffective, and thus such strategies are not recommended. Certainly, using assessments with low or no validity may reduce adverse impact, but this will also be ineffective for identifying the best talent to hire.

Some researchers have tried providing test orientation and preparation sessions to candidates who will participate in the assessment process. These sessions almost certainly make job candidates more knowledgeable about and comfortable with an organization's assessment methods, and they may thus be beneficial for this reason. However, research on the effectiveness of assessment preparation courses has not shown them to reduce the levels of adverse impact regularly observed between majority and protected group candidates on certain types of assessments.

A third strategy that has been used to mitigate adverse impact is to examine group differences on individual test items and remove those items that have the largest group differences. Generally, this strategy has not proved to be effective, because all of the items that measure a given skill tend to have about the same amount of group differences (i.e., it is difficult to find subsets of items that have small versus large group differences for a given skill). The more typical situation is when all or most of the items on a given type of test will have a similar level of group difference associated with them.

**Strategies for Reducing Adverse Impact**

Recommended

■ Recruit more qualified minorities.

■ Focus on predicting both technical and contextual performance.[40][41]

■ Use assessments with less adverse impact early in the process and those with more adverse impact later (only tends to help when there are many applicants per position).[42]

Not Recommended

■ Use assessments that have low validity.

■ Combine a number of assessments with low adverse impact (overall adverse impact is likely to be greater than that for individual assessments).[43]

■ Omit assessments that produce adverse impact if their validity evidence is favorable.

■ Provide test orientation and preparation programs to candidates (have been shown to have no impact).[44]

■ Identify and remove individual test items on which majority and minority candidates differ (have been shown to have no impact).[45]

## Costs to Develop and Administer

Some assessment methods involve much higher costs to develop and administer than others. Consider, for example, work sample assessments that require candidates to spend a full day or more to complete actual tasks and activities they will be required to perform on the job. Such assessments are more costly to both develop and administer compared with paper-and-pencil tests. They are more costly to develop because they typically require involving job experts working in collaboration with test development experts to design the exercises and scoring protocols. They are more expensive to administer because they involve setting up equipment and facilities and paying trained evaluators to observe and score how well each candidate performs on the assessment exercises.

---

[40] Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of various conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology, 82,* 656-664.

[41] Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707-721.

[42] Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies. A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49,* 549-562.

[43] Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707-721.

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

[45] Ibid.

In contrast, a paper-and-pencil test that consists of multiple-choice items that assess a job candidate's conscientiousness can be developed by testing experts without the help of job experts and can be efficiently administered and scored. Obviously, the costs for both developing and administering this type of assessment will be significantly less.

Assessments can be relatively expensive to develop but inexpensive to administer or, conversely, relatively inexpensive to develop but costly to administer over time. One point to bear in mind, however, is that there are enormous costs to an organization of consistently hiring employees who do not perform effectively or who leave the organization after investments have been made in training them. Even the highest development and administration costs generally pale in comparison to the costs associated with unproductive or unsuccessful employees. Furthermore, implementation of effective assessment procedures has been shown to result in very substantial productivity and revenue increases as well as cost savings for organizations. Therefore, it is important to not only consider the costs associated with developing and administering effective assessments, but also see these investments in light of the financial and other benefits that will be gained. The upcoming section on utility of assessments discusses how to calculate the benefits associated with using assessment methods.

Another cost factor that HR professionals need to consider is whether the organization desires to use a commercially available assessment or prefers to develop its own customized assessment. If HR professionals choose to use a commercially available assessment, they will need to enter into a licensing agreement with the test publisher, and the organization will be charged either for each use of the test or for the duration of time the test is used. The advantages of a commercially available assessment are that it can usually be implemented quickly, it is typically maintained and updated by the publisher over time, and the data usually continue to be amassed across the different organizations using the assessment. Probably the biggest downside of commercially available assessments is that licensing agreements can be expensive and last in perpetuity.

If an organization wishes to use a commercially available assessment, it is important to identify and use a reputable test publisher. An association of test publishers created guidelines and bylaws to help with this process.[46] SHRM provides comprehensive and valuable information about available assessments and related topics through the SHRM Testing Center, www.shrm.org/testing/. It is also advisable to seek out the guidance of consultants or researchers who specialize in assessment and testing practices. These individuals usually have doctorates in industrial and organizational psychology or human resource management, and many are members of the Society for

---

[46] Wonderlic, F., Jr. (1993). Test publishers form association. *Human Resources Measurements* (Supplement to the January 1993 *Personnel Journal,* p. 3).

Industrial and Organizational Psychology (SIOP). Guidance is also available that can assist HR professionals in selecting appropriate testing vendors.[47]

As an alternative, an organization can opt to have its own customized tests developed, which it will then own but need to maintain. One advantage of a customized test is that it is uniquely designed and validated in the situation for which it was developed. Thus, items can be crafted that are ideally suited to and reflect the unique requirements and culture of a given organization or job. If an organization has undertaken a significant assessment development effort and successfully developed highly valid tests, this could also be a competitive advantage for the company. Under these circumstances, the organization would not likely want to share the tests or have them used by others.

If an organization wishes to develop and use customized tests that it will own, there will be ongoing administration and maintenance costs that accompany this decision. For instance, it will be necessary to maintain databases of test scores after the assessments are administered, analyze and report assessment data for decision-making, periodically revise items and possibly revalidate assessments, and so forth. If an organization wishes to opt for development and implementation of customized assessments, it will need access to knowledgeable testing professionals either on staff or through consultant arrangements to manage the program. Running an in-house assessment program involves sufficient specialized technical knowledge and complexity, and therefore it should not be left to untrained staff.

---

**Commercially Available Versus Custom-Developed Assessments**

**Commercial Advantages**
- Test content already available, so implementation is fast.
- Most have data on validity and group differences, often from similar jobs.
- Tests are usually upgraded, maintained and continually evaluated over time by the test developer.

**Custom-Developed Advantages**
- Once initial test development costs are incurred, the organization pays no ongoing licensing or usage fees.
- Test content can be fully customized to the organization's jobs and strategy, maximizing relevance and potential competitive advantage.

**Commercial Disadvantages**
- Licensing fees to use off-the-shelf tests over time are expensive and can be prohibitive for some organizations.
- The tests are used in many situations, raising possible test security and competitive advantage concerns.

**Custom-Developed Disadvantages**
- Proper assessment development and validation takes time.
- Organization will need to be responsible for upgrades, maintenance and ongoing evaluation.

---

[47] Azar, B. (1994). Could "policing" test use improve assessments? *APA Monitor,* p. 16.

## Applicant Reactions

A final criterion that is important to consider in developing or selecting assessment methods is the reactions of the applicants. The recruitment and selection process is a two-way interaction between the organization and prospective job candidates, where both parties are selling themselves, attempting to attract the other and assessing mutual fit. Everything that an organization does in the recruitment and selection process leaves candidates with an impression about how the organization does business, and the totality of this experience has a significant impact on whether or not they will ultimately accept employment.

While there is limited research on applicant reactions to specific assessment methods, candidates who feel positively about the selection processes used by an organization report higher levels of satisfaction with the organization, and they are more likely to recommend the organization to others.[48] Because there is such competition for highly qualified job candidates, the issue of applicant reactions is not inconsequential. There are real reasons for organizational members to be mindful and careful about the impressions candidates form as a result of their recruitment and selection processes.

> ### Factors Leading to Positive Applicant Reactions[49]
>
> ■ Use of assessment methods that are perceived as relevant to the job.
>
> ■ An opportunity to perform and demonstrate job-relevant capabilities.
>
> ■ Assessment procedures that are administered consistently and fairly.
>
> ■ Feedback provided to applicants on their performance.

In general, assessments that are viewed more favorably are those that candidates perceive as more relevant to the job. Because work samples, job knowledge tests and assessment centers mirror actual activities and knowledge areas that are required to perform the job, they tend to be viewed more positively by job candidates than multiple-choice tests that seem abstract and bear little, if any, obvious resemblance to work requirements. More specifically, job-relevant assessments are viewed more favorably than cognitive ability assessments, which, in turn, are viewed more favorably than personality inventories, integrity tests and biographical data inventories.[50] Although there is no research on applicant reactions to physical ability tests, it is reasonable to expect that reactions would be fairly favorable, especially for those assessments that mirror actual job tasks.

---

[48] Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46,* 49-76.

[49] Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3,* 11-19.

[50] Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57,* 639-683.

Another factor that leads to positive reactions is providing feedback to applicants about how they have performed on an assessment. Candidates not only want feedback on their performance, but it is not uncommon for some candidates to request feedback on how they could have performed better. While many candidates desire as much information as possible about their performance on job assessments, organizations need to take a measured approach to the feedback they provide. It is certainly reasonable to let applicants know how they scored on an assessment, and it can also be helpful to provide normative data about how they scored compared with others who were assessed or hired.

However, offering specific feedback on how the applicant might have done better is not something organizations typically do. For one reason, discussing test performance in detail can divulge information about the test or suggest correct responses that would result in test security breaches. As a practical matter as well, it could quickly become quite costly for organizations to provide in-depth feedback on assessment performance to each and every job candidate considered.

## Additional Considerations

Although not criteria for selecting and evaluating assessment methods, per se, there are three other factors that are important to know about and consider in implementing assessments for selection purposes. These are (1) mode in which the assessment will be administered; (2) the utility of the assessments used; and (3) legal requirements. These topics are discussed next.

### Mode of Assessment Administration
The issue here is whether the assessment is administered via a paper-and-pencil written format, via a computer or by live administrators, either by phone or in person. Live administrators are typically required when the assessment is an interactive work sample or interview.

More and more organizations are moving away from paper-and-pencil assessments to computer-administered assessments. An important advantage of computer-administered assessments is that responses can be automatically stored in a database and assessment scores automatically generated. Computers are also capable of easily storing multiple assessment items that can be mixed up to form different versions of an assessment that measures the same KSAs. Multiple versions of an assessment help to better ensure its security.

In recent years, organizations have begun using the Internet to administer assessments, and some assessment companies now specialize in Internet-based assessments.

Although Internet-based administration has a great deal of appeal for many practitioners, there are some important caveats with this type of assessment administration that need to be considered. First, unless the assessment is proctored, there is no easy way of knowing whether the individual taking the assessment is getting help from others or possibly even having someone take the assessment for him or her. Because of this limitation, unproctored Internet-based assessment should be used only as a preliminary screen to weed out clearly unqualified candidates. Any Internet-based assessment should be followed up with a more comprehensive assessment in a monitored and controlled assessment environment. Alternatively, candidates can be brought into testing centers where Internet-administered assessments can be properly monitored and proctored. A more complete discussion of computer and Internet-based assessment issues can be found in a paper by Potosky and Bobko.[51]

## Utility of Assessments

Another concept that is important to understand in the assessment arena is utility. Utility addresses the benefits of implementing an assessment. When one is attempting to persuade non-HR decision makers about the advantages of using assessments, the case can often be made more convincingly by discussing utility than by discussing validity. There are two types of utility that can be expected from implementing a new assessment method: hiring success gains and economic gains.[52]

Hiring success gains refer to the number of successful hires that an organization experiences subsequent to implementing an assessment method compared with the number the organization experienced prior to implementation. The greater the hiring success gain, the higher the utility of the assessment.

Economic gains refer to the monetary impact of using an assessment method on the organization. In essence, economic gain is evaluated by examining the revenue generated by hiring high-performing employees using an assessment versus the costs of using it. The greater the ratio of revenues to costs, the higher the utility of the assessment. The actual formulae used to assess economic gain are a bit more complicated than a simple comparison of revenue and costs. They consider additional factors, such as the correlation between the assessment and job performance, the number of applicants and the tenure of the selected group, among others. More information on how to calculate economic gains can be found in Heneman and Judge (2006). Suffice it to say, however, that if decision makers can be shown the positive impacts of assessment methods in terms of their bottom line hiring and economic gains, the value of the assessments is more obviously apparent.

---

[51] Potosky, D., & Bobko, P. (2004). Selection testing via the Internet: Practical considerations and exploratory empirical findings. *Personnel Psychology, 57,* 1003-1034.

[52] Heneman, H. G., III, & Judge, T. A. (in press). *Staffing organizations, 5E.* Middleton, WI: Mendota House, and New York: McGraw-Hill.

### Legal Requirements

It is important for any HR professional involved in employee selection to have an understanding of the legal issues and requirements that govern the use of assessment methods in organizations. There are two key documents with which all HR professionals should be familiar.

The first is the Uniform Guidelines on Employee Selection Procedures. This document contains a comprehensive set of federal regulations specifying requirements for selection systems of organizations covered under the Civil Rights Acts and under E.O. 11246 (see www.eeoc.gov/regs for the full text of this document). These federal regulations address the need to determine if a selection procedure is causing adverse impact and if so, the validation requirements for using the procedure. To assist practitioners in applying the Uniform Guidelines properly, several documents and articles have been published. The reader is referred to two of these in particular: *The Society for Industrial and Organizational Psychology Principles for the Validation and Use of Personnel Selection Procedures* (www.siop.org) and *the Standards for Educational and Psychological Testing* (www.apa.org/science/standards.html).

A common misconception among HR professionals is that the use of informal and unstructured selection processes in lieu of formal assessments is one way to avoid legal issues and it negates the need to comply with federal regulations, such as the Uniform Guidelines. This is not the case. If any selection practices or procedures used are shown to produce an adverse impact, irrespective of how informal or "safe" they may appear, the organization is required to demonstrate the validity of the procedures. The irony is that rather than avoiding formal assessments because of legal concerns, organizations should embrace well-developed and validated formal assessments and use them more extensively to ensure that any legal challenges can be addressed successfully.

HR professionals and especially managers sometimes react very negatively when legal concerns in hiring are raised. They feel that requirements, like those outlined in the Uniform Guidelines, are simply bureaucratic burdens imposed by the government that interfere with organizations selecting the most effective workforces. It is important to point out that the procedures outlined in the Uniform Guidelines are, in fact, the exact same procedures that an assessment expert would employ if the only goal was to identify the best qualified candidates for a job. Thus, at least in this case, compliance with legal requirements also represents best practice in developing and implementing the most effective selection practices possible.

The second document that all HR professionals engaged in selection work need to be familiar with is the Americans with Disabilities Act of 1990 (ADA). The ADA creates substantial requirements for providing reasonable accommodation to applicants with

disabilities in the selection process. If a job requires a skill that a disabled person cannot perform, the organization is not required to provide reasonable accommodation to allow that skill to be assessed. However, if there are ways to accommodate disabilities both on the job and in the assessment process, the organization must provide for this. Examples of reasonable accommodation might include substituting an oral assessment for a written one, providing extra time to complete an assessment, providing an assessment in Braille, providing items in larger print or providing an interpreter who can communicate items in sign language. Guidance is available from several sources to help HR professionals better understand their responsibilities under the ADA and provide appropriate and reasonable accommodation when warranted.[53] [54] [55] [56]

[53] Daley, L., Dolland, M., Kraft, J., Nester, M. A., & Schneider, R. (1988). *Employment testing of persons with disabling conditions.* Alexandria, VA: International Personnel Management Association.

[54] Eyde, L. D., Nester, M. A., Heaton, S. M., & Nelson, A.V. (1994). *Guide for administering written employment examinations to persons with disabilities.* Washington, D.C.: U.S. Office of Personnel Management.

[55] Williams, S. K. (Jan/Feb 2003). *Tips for minimizing abuses under the Americans with Disabilities Act.* Alexandria, VA: Society for Human Resource Management.

[56] Job applicants and the Americans with Disabilities Act. www.eeoc.gov

# Summary and Conclusions

Selecting and using effective assessment methods can greatly enhance the quality and productivity of an organization's workforce. Unfortunately, many HR professionals have misconceptions about both the value of formal assessments and the types of assessments that have proven to be most effective. This, coupled with the fact that the area of selection testing is inherently technical and difficult to understand, has led to an underutilization of formal assessments in organizations. By providing a basic under-standing of key criteria for evaluating assessment methods and resources for accessing assessment-related information and expertise, it is our hope that this report has taken a positive step toward better equipping organizational decision makers and HR profes-sionals to introduce effective assessment methods into their organizations.

# Sources and Suggested Readings

## Job Analysis

Brannick, M. T., & Levine, E. L. (2002). *Job analysis.* Thousand Oaks, CA: Sage.

> This book provides a user-friendly compilation of how to conduct a job analysis and use the results. The topics covered include the building blocks of job analysis methods, work (task)-oriented methods, worker (KSA)-methods, hybrid methods that combine both task and KSA job analysis, job analysis for management jobs and teams, job analysis and the law, doing a job analysis study, and uses of job analysis in job descriptions, performance appraisals, job evaluation, job design, staffing and training. Practical examples and summaries of research are interspersed throughout the book.

Gael, S. (Ed.). (1988). *The job analysis handbook for business, industry, and government* (Vols. 1 and 2). New York: Wiley.

> In these volumes, the author argues that job analysis is the hub of virtually all human resource administration and management activities, and it is necessary for the successful functioning of organizations. The handbook addresses job analysis topics and issues comprehensively, contains chapters written especially for job analysis by recognized authorities and presents material relevant to a wide audience of practitioners working in business, industry, labor unions, universities, the military, and federal, state and local governments.

## Cognitive Ability Tests

Gottredson, L. S. (1986). *Societal consequences of the g factor in employment. Journal of Vocational Behavior, 29,* 379-410.

> This article discusses and rebuts seven common arguments that general intelligence (*g*) is of little or no practical importance in employment and illustrates the effect that differences in intelligence in a workforce may have on the structure and functioning of whole societies. Evidence and theory are presented to support the position that the occupational status hierarchy is an intellectual complexity factor among occupations that has evolved in response to the wide dispersion in intelligence levels within populations. The mean black-white difference in intelligence is used to illustrate society-wide ramifications of individual and group differences in intelligence, particularly when social policies are based on misconceptions about intelligence or its impact.

**Hunter, J. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance.** *Journal of Vocational Behavior, 29,* 340-362.

A review of the research literature indicates that general cognitive ability (GCA) predicts supervisor ratings and training success as well as objective, rigorously content-valid work sample performance. Previous studies showed that much of this predictive power stemmed from the fact that GCA predicted job knowledge and job knowledge predicted job performance. However, GCA was shown to predict performance to a greater extent, verifying job analyses showing that most major cognitive skills were used in everyday work. Evidence showing that it is GCA—and not specific cognitive aptitude—that predicts performance is discussed. The author argues that the findings from this study support classic learning theory over behaviorist theories of learning and performance.

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

An examination of the cumulative research on various predictors of job performance showed that for entry-level jobs there was no predictor with validity equal to that of cognitive ability. For selection on the basis of current job performance, the work sample test was slightly better. For federal entry-level jobs, this research showed that substitution of an alternative predictor instead of using cognitive ability would cost from $3.12 (job tryout) to $15.89 billion/year (age). Hiring on ability had a utility of $15.61 billion/year but affected minority groups adversely. Hiring on ability by quotas would decrease utility by 5%. A third strategy—using a low cutoff score—would decrease utility by 83%. It was suggested that using other predictors in conjunction with ability tests might improve validity and reduce adverse impact.

**Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than** *g. Journal of Applied Psychology, 79,* 518-524.

The roles of general cognitive ability (*g*) and specific abilities or knowledge (*s*) were investigated as predictors of work sample job performance criteria in seven jobs for U.S. Air Force enlistees. Both *g* and *s* were defined by scores on the enlistment selection and classification test (the Armed Services Vocational Aptitude Battery) used by the armed services for selection. Analyses revealed that *g* was the best predictor of all performance measures that were examined and that *s* added a statistically significant but practically small amount to this prediction. These results were consistent with those of previous studies, most notably the Army's Project A (J. J. McHenry, L. M. Hough, J. L. Toquam, M. A. Hanson & S. Ashworth, 1990).

## Personality Tests

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 91,* 1-26.

This study examined the validity of five personality measures for predicting performance in five occupational groups (professionals, police, managers, sales and skilled or semi-skilled). Three different kinds of performance measures were used in this study: job proficiency, training proficiency and personnel data. The results showed that one of the personality measures—conscientiousness—was consistently related to all three performance measures for all of the occupational groups. Extraversion (another one of the personality measures examined) was a valid predictor for two occupations involving social interaction—managers and sales. Openness to experience and extraversion were valid predictors of the training proficiency criterion. The results for agreeableness suggested that it was not an important predictor of job performance. Overall, the results illustrated the benefits of using the five-factor model of personality in assessment research.

Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13,* 653-665.

A considerable number of research studies have shown support for a five-factor model of personality traits. These include conscientiousness, extraversion, agreeableness, openness to experience and emotional stability. The claim that these factors represent the basic dimensions of personality is based on four lines of reasoning and evidence that are discussed in this paper: (1) longitudinal and cross-observer studies have demonstrated that all five factors are enduring dispositions that are manifested in patterns of behavior; (2) traits related to each of the factors have been found in a variety of personality systems and in the natural language of trait description; (3) the factors have been found in different age, sex, race and language groups, although they may be somewhat differently expressed in different cultures; and (4) evidence of heritability has suggested that all have some biological basis. The authors also identify and address some outstanding issues regarding the five-factor model.

**Hough, L. M. (1992). The big five personality variables—construct confusion: Description versus prediction. *Human Performance, 5,* 135-155.**

This article suggests that the five-factor model of personality traits (extraversion, conscientiousness, agreeableness, emotional stability and openness to experience) is not an adequate taxonomy of personality variables for predicting important job performance measures. The author argues that the five factors are too heterogeneous and incomplete. This paper first discusses the development of personality and job performance models. A nine-factor taxonomy is then presented that includes affiliation, potency, achievement, dependability, adjustment, agreeableness, intellectance, rugged individualism and locus of control. Comparisons between different taxonomies are made, and correlational evidence demonstrating the usefulness of the nine personality measures is presented.

**Raymark, M. J., Schmit, M. J., & Guion, R. M. (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology, 50,* 723-736.**

This article describes the Personality-Related Position Requirements Form (PPRF), a job analysis form used in making hypotheses about what personality predictors will be relevant for predicting performance in different jobs. The Big Five personality factors provided an organizing framework for the PPRF. Subsequent development resulted in identifying 12 specific sets of items for facets of each of the Big Five. A study was conducted by gathering job descriptions on 260 different jobs to determine if the PPRF could reliably differentiate between jobs, and such evidence was found. The PPRF is offered to both researchers and practitioners for use, refinement and further testing of its technical merits and intended purposes.

**Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.**

This study examined past research to (1) assess the overall validity of personality measures as predictors of job performance; (2) investigate moderating effects of several study characteristics on personality scale validity; and (3) investigate the predictability of job performance as a function of eight personality measures: neuroticism, extraversion, openness, agreeableness, conscientiousness, locus of control, Type A and miscellaneous. Based on the review of 494 studies, usable results were identified for 97 independent samples. Consistent with predictions, studies using confirmatory research strategies produced corrected mean personality scale validity that was more than twice as high as that based on studies adopting exploratory strategies. An even higher mean validity was obtained based on studies using job analysis explicitly in selection of personality measures.

## Biodata Tests

**Hough, L. M. (1984). Development and evaluation of the "accomplishment record" methods of selecting and promoting professionals. *Journal of Applied Psychology, 69,* 135-146.**

This study described the development of the "accomplishment record" method, which resulted in self-reported descriptions of accomplishments in areas that were highly relevant for making selection or promotion decisions. The accomplishments were reliably rated using specially prepared rating scales and guidelines. To test the validity of the accomplishment record method, 329 attorneys prepared accomplishment records according to specified instructions and also completed a battery of more typical assessment methods. Scores on the accomplishment record inventory were unrelated to the traditional assessment measures (e.g., aptitude tests, grades and honors), but they correlated with job performance. The accomplishment record had also been found to correlate with self-perceptions of success, hard work and self-assurance and with the length of time spent practicing a profession. This accomplishment record also appeared to work equally well for predicting the job performance of females, minorities and white males.

**Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three 'alternative' selection measures. *Personnel Psychology, 36,* 261-276.**

A content-oriented validation strategy was used to develop three alternative selection inventories that were designed to reflect the job content of positions held by attorneys employed with a large federal agency. These inventories and three traditional assessment methods were completed by 329 agency attorneys as part of a concurrent validation study. Criterion-related validities of two traditional inventories (a background inventory and an interest and opinion inventory) and one alternative inventory (an accomplishment record inventory) were statistically and practically significant. The special features and advantages of the accomplishment record inventory as an alternative selection procedure were discussed.

**Shoenfeldt, L. F. (1999). From dustbowl empiricism to rational constructs in biodata. *Human Resource Management Review, 9,* 147-167.**

This paper presents an attempt to use different types of biographical data scales in a concurrent validation to predict service orientation. Over 867 service employees, along with a large number of applicants, completed a 137-item biographical questionnaire constructed to cover 15 aspects of service delivery identified through an extensive job analysis. Biographical data scales were developed using different techniques, some relying on statistical procedures and others relying on rational judgment. They were validated against supervisory ratings of service orientation and overall performance, along with three personnel measures: days absent, times absent and times tardy. Several types of scales showed positive validity results with the performance measures. The results were promising in terms of both the prediction and understanding of customer service orientation.

**Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.). (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction.* Palo Alto, CA: CPP Books.**

The information derived from biographical data enables professionals to predict the future career choice, performance, safety behaviors and turnover of employees. This is done by assessing prospective employees' past performance and behaviors that are relevant to the target job of interest. This book is a comprehensive reference that provides information for understanding and using biographical data in selection in the public and private sectors (including selection of blue-collar and federal government employees and work team members), career counseling and development, and job classification.

## Integrity Tests

**Camara, W. J., & Schneider, D. L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist, 49,* 112-119.**

This article described two independent reports, completed by the American Psychological Association (APA) and the U.S. Congress Office of Technology Assessment, that examined scientific and measurement issues concerning integrity testing. Data were presented on a variety of tests collected by surveying test publishers. The survey data provided a view of the industry's scope (e.g., test audience, user screening, score reporting) that was not available elsewhere. The article also addressed unresolved issues regarding integrity testing that had a wide range of implications for the profession of psychology, the testing industry and public policy (e.g., cutting scores, user screening and training and test marketing practices).

**Goldberg, L. R., Grenier, R. M., Guion, L. B., Sechrest, L. B., & Wing, H. (1991).** *Questions used in the prediction of trustworthiness in pre-employment selection decisions: An APA Task Force Report.* **Washington, D.C.: American Psychological Association.**

This report presented the findings of an APA task force that was appointed to review available commercial tests used for the purpose of assessing the honesty and integrity of prospective employees. Specifically, the task force was asked to gather and examine data relevant to the scientific and social-policy considerations associated with the development and use of assessments of dishonesty, theft and related behaviors. The report focused exclusively on commercially published instruments used for pre-employment selection decisions about the trustworthiness (vs. untrustworthiness) of job applicants. It did not consider assessments that required special apparatus or were used to assess current employees.

**Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance.** *Journal of Applied Psychology* **(monograph), 78, 679-703.**

The authors conducted a comprehensive examination of the validity of integrity tests based on 665 validity coefficients. Results indicated that integrity test validities were substantial for predicting job performance and counterproductive behaviors on the job, such as theft, disciplinary problems and absenteeism. The estimated mean validity of integrity tests for predicting supervisory ratings of job performance was .41. Results of validity studies conducted on applicants using external performance measures (i.e., excluding self-reports) indicated that integrity tests predicted the broad criterion of organizationally disruptive behaviors better than they predicted employee theft alone. The authors concluded that integrity test validities were positive across situations and settings.

**Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection.** *Personnel Psychology, 49,* **787-829.**

This article reviews journal articles, books, book chapters, law review articles, convention papers and dissertations regarding integrity testing for personnel selection. Developments include an examination of professional and congressional inquiry into this area of testing, rapid growth of the validity database, new insight into similarities and differences between different tests, and links to the Big Five personality dimensions. Inquiries into relationships with other constructs are reviewed, as are applicant reactions to these tests. The effects of the Civil Rights Act of 1991 and the Americans with Disabilities Act are considered. In the domain of criterion-related validity, the cumulative database has grown dramatically, and the pattern of findings continues to be consistently positive.

## Structured Interviews

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50,* 655-702.

> This article reviewed the research literature in order to describe and evaluate the many ways selection interviews can be structured. Fifteen components of structure were identified that may enhance either the content or the evaluation process of the interview. Each component was described, and the different ways in which they had been used in previous studies were discussed. Then, each component was critiqued in terms of its impact on reliability, validity and user reactions. Finally, recommendations for research and practice were presented. The authors concluded that interviews could be easily enhanced by using the various components of structure. They recommended that improvement of this popular selection procedure be a high priority for future research and practice.

Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41,* 25-42.

> This article proposed an employee-interviewing technique that included the following steps: (1) develop questions based on a job analysis; (2) ask the same questions of each candidate; (3) use examples and illustrations; (4) have a panel record and rate answers; (5) administer the process to all candidates; and (6) emphasize job relatedness, fairness and documentation. When the interview was used to hire 149 entry-level production employees, it revealed high interrater reliability and predictive validity as well as evidence for test fairness and utility.

Eder, R. W., & Ferris, G. R. (Eds.). (1989). *The employment interview: Theory, research, and practice.* Newbury Park, CA: Sage Publishing.

> This book provides a comprehensive treatment of theory, research and practice relevant to the employment interview. The topics covered include the interviewer's decision-making process, applicant strategies and employment interview validity. There is also a concluding commentary summarizing the volume's implications for theory building, research methods and effective practice.

**Judge, T. A., Higgins, C. A, & Cable, D. M. (2000). The employment interview: A review of recent research and recommendations for future research.** *Human Resource Management Review, 10,* 383-406.

> This is a comprehensive review that examined recent research on the employment interview process. The authors began with a review of the traditional areas of interview research: reliability, validity, structured interviews, interviewer differences, equal employment opportunity issues, impression management and decision-making processes. Next, they reviewed and discussed more recent developments in interview research, such as the use of the interview as a means of assessing person-organization fit and applicant reactions to the employment interview. Throughout the review, they suggested topics for future research.

## Physical Fitness and Physical Abilities Tests

**Blakley, B. R., Quinones, M. S., Crawford, M. S., & Jago, I. A. (1994). The validity of isometric strength tests.** *Personnel Psychology, 47,* 247-274.

> This article examined six studies in which isometric strength tests were used as part of selection procedures. The studies represented seven jobs across various industries with a total sample size of 1,364 individuals. The relationship between performance on four isometric strength tests and both supervisory ratings of physical performance and performance on work simulations was examined. Results indicated that isometric strength tests were valid predictors of both types of performance measures across all of the jobs examined. In addition, the four of the six tests were found to correlate more strongly with work simulations than with supervisory ratings of physical performance.

**Campion, M. A. (1983). Personnel selection for physically demanding jobs: Review and recommendations.** *Personnel Psychology, 36,* 527-550.

> In this review article, the authors argued that improvements in personnel selection systems for physically demanding jobs were needed due to equal employment opportunity (EEO) considerations, concern for worker physical well-being and the lack of alternative procedures. The paper addressed the special EEO sensitivities of physical abilities selection and reviewed the literature from a variety of disciplines on (1) the physiological background underlying the selection strategies; (2) the assessment of human physical abilities; (3) the measurement of physical requirements of jobs; and (4) the physical abilities personnel selection studies reported in the literature.

Hogan, J. (1991). Physical abilities. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology,* vol. 2 (pp. 753-831). Palo Alto, CA: Consulting Psychologists Press.

This chapter identifies factors that affect the ability to perform in the workplace, including physiological responses, training, ability and injury, and workload. It reviews procedures and methods for designing personnel selection systems for physically demanding jobs and emphasizes the unique physical ability content of job analyses, performance measures and assessments. In addition, the chapter summarizes legislative decisions designed to protect women, handicapped persons and older workers in light of their implications for physical ability assessments by employers. Finally, it argues that the need to improve development and implementation of fair physical ability selection procedures and the need to understand the relationship between physical performance and other organizational effectiveness outcomes are important issues for future research.

## Situational Judgment Tests

Chan, D., & Schmitt, N. (1997). Video-based versus paper and pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82,* 143-159.

This study examined the effects of race, reading comprehension, method of assessment, face validity perceptions and performance on a situational judgment test. The situational judgment test was administered via a videotape and via a paper-and-pencil instrument. A total of 241 psychology undergraduates (113 blacks and 128 whites) participated in the study. The results showed that the racial differences in situational judgment test performance and face validity reactions to the test were substantially smaller with the video-based method of testing compared with the paper-and-pencil method. Implications of the findings were discussed in the context of research on adverse impact and examinee test reactions.

Motowidlo, S. J., Dunnette, M. D., & Carter G. (1990). An alternative selection procedure: A low fidelity simulation. *Journal of Applied Psychology, 75,* 640-647.

From critical-incident analysis and judgments by subject-matter experts, a low-fidelity simulation was developed for selecting entry-level managers in the telecommunications industry. The simulation presented applicants with descriptions of work situations and five alternative responses for each situation. It asked them to select one response they would most likely make and one they would least likely make in addressing each situation. In a sample of 120 management incumbents, simulation scores correlated from .28 to .37 with supervisory ratings of performance. These results show that samples of hypothetical work behavior can predict performance, without the props, equipment or role players often required by high-fidelity simulations, such as work-sample tests or assessment centers.

Weichmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410-417.

Using three different samples, the authors assessed the increases in the validity produced by situational judgment inventories (SJIs), relative to job knowledge, cognitive ability, job experience and conscientiousness, in the prediction of job performance. The SJI was a valid predictor in three samples and produced substantial increases in validity in two samples. Relative to the other predictors, SJI's correlation with performance, controlling for the other predictors, was superior in most comparisons. Subgroup differences on the SJI also appeared to be less than those for cognitive ability and job knowledge measures, but greater than the differences observed in conscientiousness. The authors concluded that the SJI should prove to be a valuable additional measure in the prediction of job performance, but also suggested several additional areas of research.

## Work Sample Tests

Asher, J. J., & Sciarrino. J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology, 27,* 519-533.

This study reviewed the validity evidence for a wide sample of motor and verbal work sample tests, which were designed to measure on-the-job behaviors. Motor tests were shown to have higher validities than verbal tests when job proficiency was the performance measure, but this pattern was reversed when success in training was the performance measure. These validities were somewhat lower than those for biographical information but higher than those for other predictors. The authors discussed a number of possible explanations for these findings.

Howard, A. (1983). Work samples and simulations in competency evaluation. *Professional Psychology: Research and Practice, 14,* 780-796.

This article argues that the evaluation of professional competence in psychology—by licensing boards, the American Board of Professional Psychology or other boards—has not taken full advantage of expertise in psychological measurement within the profession. The advantages and disadvantages of work samples and simulations for such purposes are discussed and compared with those of paper-and-pencil tests. Examples from various professions and from psychological research in personnel selection are presented. A comparison of work samples with simulations shows greater advantages for the latter, but a combination of exercises in an assessment center model is recommended. An illustration of how examinations of psychological competence at the licensing and the specialty board levels could be improved by incorporating work samples and simulations is presented.

## Assessment Centers

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analyses of assessment center validity. *Journal of Applied Psychology, 72,* 493-511.

This paper examined the validity of assessment centers across a large number of studies. Specifically, the authors calculated the average validity of 107 validity coefficients from 50 assessment centers and found an average validity of .37. Validities were sorted into five categories based on the performance measures used in the study and four categories based on the purpose of the assessment. Higher validities were found in studies in which ratings of potential were the performance measures, and lower validities were found in studies where the assessment center was used to make promotion decisions. The authors also found that the validities were higher when (1) the percentage of females who were being assessed was high; (2) when several evaluation devices were used; (3) when assessors were psychologists rather than managers; (4) when peer evaluation was used; and (5) when the study was methodologically sound. Age of those being assessed, whether feedback was given, days of assessor training, days of observation, percentages of minority candidates being assessed, and criterion contamination did not have any effect on assessment center validities.

Thornton, G. C., III. (1992). *Assessment centers in human resources management.* Boston: Addison-Wesley.

While assessment centers have been used for numerous human resource management functions, all assessment centers are not alike. This book shows how assessment for managerial selection/promotion, diagnosis of managerial training needs and team building among managerial groups are done differently. The author also shows how the assessed dimensions, observations and participant feedback all must be tailor-made to the specific application of the assessment center method.

Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance.* New York: Academic Press.

This book focuses on examining the assessment center experience. It traces the historical development of multiple assessment procedures with emphasis on those advances relevant to assessment center; critiques all of the published and unpublished research on assessment centers; integrates assessment center procedures into several theories of measurement and human judgment; and presents new models of job analysis, the nature of managerial work, work-sampling assessment methods and the process of human judgment based on the assessment center experience.

## Adverse Impact

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50,* 707-721.

> A common assumption exists that including predictors that demonstrate smaller group differences with others that demonstrate larger group differences will help alleviate the adverse impact observed. The purpose of this paper was to answer the question, "If two or more predictors are combined that have smaller and larger group differences, what will be the magnitude of group differences and, consequently, of adverse impact?" To answer this question, a set of tables, figures and formulas were presented that highlighted variables influential in affecting how combinations of predictors influenced observed group differences. A number of conclusions were drawn that clarified the extent to which combining predictors with smaller and larger group differences affected subsequent adverse impact.

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

> Cognitively loaded tests of knowledge, skill and ability often contribute to decisions regarding education, jobs, licensure or certification. Users of such tests often face difficult choices when trying to maximize both the performance and ethnic diversity of chosen individuals. The authors describe the nature of this quandary, review research on different strategies to address it and recommend using selection materials that assess the full range of relevant attributes using a format that minimizes verbal content as much as is consistent with the outcome one is trying to achieve. They also recommend the use of test preparation, face-valid assessments and the consideration of relevant job or life experiences. Regardless of the strategy adopted, however, they suggest that it is unreasonable to expect that one can maximize both the performance and ethnic diversity of selected individuals.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies. A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49,* 549-562.

This study examined the effects of selection process variables on employee performance and on minority hiring. Simulated data were used to estimate the effects of 14 selection procedures, including the use of within-group norming (i.e., selecting the highest scoring candidates within each racial or gender group rather than selecting the highest scoring candidates overall), which was restricted by the Civil Rights Act of 1991. The authors examined how selection rules that did and did not include within-group norming fared in terms of tradeoffs between performance and minority representation. The results showed that the preferred strategy depended on the relative value the organization placed on performance versus minority representation. Results also showed that the effects of different screen-then-select selection strategies varied as a result of the selection ratios at the screening and selection stages, thus precluding simple conclusions about the merits of each selection strategy.

## Applicant Reactions

Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3,* 11-19.

Applicant reactions to selection procedures were examined in terms of the satisfaction and/or violation of 10 procedural justice rules. The authors first collected 237 critical incidents describing fair and unfair treatment during selection processes from 31 individuals who had recently participated in job search and hiring. The critical incidents were categorized into 10 procedural justice rules, and the distribution of these incidents was examined for different hiring outcomes and different selection procedures. The most common procedural concerns reflected selection procedure job relatedness and the interpersonal treatment applicants had received. Accepted applicants were primarily concerned about consistency of treatment, while rejected applicants were more concerned about receiving timely feedback and blatant bias. Ease of faking was the primary procedural concern among applicants who took honesty and personality tests, while job relatedness was the primary concern among applicants who took ability and work sample tests. The authors concluded the paper by discussing future research issues and offering practical suggestions for minimizing applicants' negative reactions to selection processes.

**Hausknecht, J. P., Day, D. V., & Thomas, S.C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis.** *Personnel Psychology, 57,* 639-683.

An updated theoretical model of applicant reactions to selection procedures was proposed and tested. Results from 86 independent samples (48,750 individuals) indicated that applicants who held positive perceptions about selection were more likely to view the organization favorably and report stronger intentions to accept job offers and recommend the employer to others. Applicant perceptions were positively correlated with actual and perceived performance on selection tools and with self-perceptions. The average correlation between applicant perceptions and gender, age and ethnic background was near zero. Face validity and perceived predictive validity were strong predictors of many applicant perceptions, including procedural justice, distributive justice, attitudes toward tests and attitudes toward selection. Interviews and work samples were perceived more favorably than cognitive ability tests, which were perceived more favorably than personality inventories, honesty tests, biodata and graphology. The discussion identified remaining theoretical and methodological issues as well as directions for future research.

**Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. (1993). Applicant reactions to selection procedures.** *Personnel Psychology, 46,* 49-76.

The premise of this research was that applicant reactions to selection procedures may be of practical importance to organizations' attractiveness to candidates and to selection procedure validity and utility. In part one of a two-part study, 110 newly hired entry-level managers and 44 recruiting-employment managers viewed sample items or brief descriptions of 14 selection tools. They judged simulations, interviews and cognitive tests with relatively concrete item-types (such as vocabulary, standard written English and mathematical word problems) as significantly more job-related than personality, biodata and cognitive tests with relatively abstract item types (such as quantitative comparisons and letter sets). A measure of new managers' cognitive abilities was positively correlated with their perceptions of the job relatedness of selection procedures. In part two, the reactions of 460 applicants to a range of entry-level to professional civil service examinations were found to be positively related to procedural and distributive justice perceptions and willingness to recommend the employer to others.

## Americans With Disabilities Guidance

Daley, L., Dolland, M., Kraft, J., Nester, M. A., & Schneider, R. (1988). *Employment testing of persons with disabling conditions.* Alexandria, VA: International Personnel Management Association.

This is a monograph that is based on a symposium presented at both the 1987 IPMAAC Conference and the 1987 IMPA International Conference. The first section is devoted to a review of research bearing on the accommodation of tests and testing procedures for disabled examinees. Section two presents the experiences of two large eastern states and their programs of accommodation testing. Sections three and four are commentaries on the material presented in the first two sections and reflect the perspectives of the public sector manager and of the disabled community. An appendix presents model guidelines for accommodated testing of the disabled.

Eyde, L. D., Nester, M. A., Heaton, S. M., & Nelson, A. V. (1994). *Guide for administering written employment examinations to persons with disabilities.* Washington, D.C.: U.S. Office of Personnel Management.

This paper provides guidance for administering written examinations to persons with disabilities. It also includes information on legal requirements for reasonable accommodations and offers detailed guidelines for testing applicants who are deaf or hard of hearing or who have vision or motor impairments. The paper includes a checklist for physical accessibility of test sites and information on effective personal interactions with persons with disabilities. Although the guide was primarily intended for federal government test administrators, most of the information covered is useful for any individual who administers written tests.

Williams, S. K. (Jan/Feb 2003). *Tips for minimizing abuses under the Americans with Disabilities Act.* Alexandria, VA: Society for Human Resource Management.

This paper provides a discussion of the law, its interpretation and implications for usage of various selection techniques in relation to the Americans with Disabilities Act of 1990. It covers physical exams, psychological exams, health questionnaires, physical agility tests, drug and alcohol tests, specific oral and written questions (e.g., about the existence of a disability, about ability to perform job-related functions), and specific requests (e.g., how you would perform job-related functions with or without reasonable accommodation). Indications are given about what inquiries can be made about disabilities for pre- and post-offer external applicants and for internal employee applicants.

## Legal Guidance

Azar, B. (1994). Could "policing" test use improve assessments? *APA Monitor,* p. 16.

This article discusses APA's Standards for Educational and Psychological Testing. It discusses issues surrounding the enforcement of the standards as well as the possibility of developing more specific standards. The article describes the potential positive and negative consequences if changes were made to the standards and the enforcement practices.

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

This document outlines principles adopted by the Society for Industrial and Organizational Psychology (SIOP) of the American Psychological Association for the validation and use of personnel selection and assessment procedures, including performance appraisal. The Principles specify SIOP policy on conducting validation research using principles of good practice in the choice, development and evaluation of personnel selection procedures.

Uniform guidelines on employee selection procedures. (1978). *Federal register, 43,* 38295-38315.

These guidelines incorporate a single set of principles that are designed to assist employers, labor organizations, employment agencies and licensing and certification boards to comply with requirements of federal law prohibiting employment practices that discriminate on grounds of race, color, religion, sex and national origin. They are designed to provide a framework for determining the proper use of tests and other assessment procedures (including performance appraisal) and to inform employers how the Equal Employment Opportunity Commission would evaluate personnel practices to ensure adherence to federal law.

## General References

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W. Borman and Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

In this chapter, the authors argue that selection criteria should embrace a domain of organizational behavior broader than just technical task performance. Specifically, performance measures should be expanded to also include contextual activities. Contextual work activities include non-technical behaviors such as helping co-workers when needed and putting in whatever effort is necessary to get the job done. After discussing what contextual performance is and how it differs from task performance, the authors present four streams of research that illustrate aspects of contextual performance. Finally, the authors discuss various individual difference variables that may be useful in predicting contextual job performance.

Heneman, H.G., III, & Judge, T.A. (in press). *Staffing organizations, 5E.* Middleton, WI: Mendota House, and New York: McGraw-Hill.

> This is a comprehensive, user-friendly book dealing with all aspects of the staffing process. It is based on a comprehensive staffing model and includes the following components: (1) staffing models and strategy; (2) staffing support systems (legal compliance, planning, job analysis); (3) core staffing systems (recruitment, selection, employment); and (4) staffing system and retention management. The book is filled with up-to-date research, useful examples and best business practices in the staffing area. In addition, each chapter concludes with in-depth applications (cases and exercises) that enhance skills and provide practice in key staffing activities and related decision-making.

Kehoe, J. (2000). *Managing selection in changing organizations.* San Francisco: Jossey-Bass.

> This edited book provides managers and HR practitioners with practical guidance on making decisions about employee staffing. The chapters are authored by leading researchers and practitioners with extensive experience in the staffing area who offer proven strategies for the design and management of effective selection processes in organizations. The book examines selection in its organizational, social and legal contexts and helps human resource professionals forge links between staffing and other functions such as training, development, recruitment and resourcing.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. A. (1990). Project A validation results: The relationships between predictor and criterion domains. *Personnel Psychology, 43,* 335-353.

> In this study, a predictor battery of cognitive ability, perceptual-psychomotor ability, temperament/personality, interest and job outcome preference measures was administered to 4,039 enlisted soldiers in nine Army occupations. Relationships between the predictor measures and five components of job performance were analyzed. Scores from the cognitive and perceptual-psychomotor ability tests provided the best prediction of job-specific and general technical task proficiency, while the temperament/personality measures best predicted giving extra effort, supporting peers and exhibiting personal discipline. Scores from the interest inventory correlated more highly with technical task proficiency than with demonstrating effort and peer support.

Murphy, K. (1996). *Individual differences and behavior in organizations.* San Francisco: Jossey-Bass.

> This edited book integrates existing research and stimulates new ways of thinking about how individual differences affect people's behaviors and experiences in organizations. It provides insights into what influences behavior on the job. Leading psychologists, who authored the individual chapters in this book, examine the entire spectrum of individual differences and show how they affect personality, vocational interests, job performance and work dysfunctions.

**Potosky, D., & Bobko, P. (2004). Selection testing via the Internet: Practical considerations and exploratory empirical findings. *Personnel Psychology, 57,* 1003-1034.**

This article presents equivalence information and practical lessons concerning selection testing via the Internet. The authors identify several issues associated with measurement and validity, the role of individual characteristics, respondents' reactions and behaviors, and other considerations concerning Internet test administration. They also report results from an exploratory study of the correlation between paper-and-pencil and Internet-administered cognitively oriented selection tests (including timed and untimed, proctored tests). The results suggest modest degrees of cross-mode equivalence for an untimed situational judgment test (r = .84) and for a timed cognitive ability test (r = .60). Further, some types of items (math, verbal, spatial) in the timed cognitive ability test seem to play a different role in the reduced cross-mode equivalence. New issues regarding the perception of, and reaction to, items presented via the Internet are presented, and a variety of practical issues are discussed.

**Ryan, A. M., & Tippins, N. T. (2004). Attracting and selecting: What psychological research tells us. *Human Resource Management, 43,* 305-318.**

This article reviews research on what selection tools work, what recruitment strategies work, how selection-tool use relates to workforce diversity and what staffing and recruiting processes lead to positive applicant perceptions. Knowledge and implementation gaps in these areas are also discussed, and key research findings are presented. To aid the HR manager in evaluating how well a staffing system fits with current research knowledge, a list of audit questions is presented. The goal is to assist the reader in not only understanding the present gaps between research and practice in recruitment and selection, but also to develop skills for employing research in HR practice.

**Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resources practices: Correspondence between research and practice. *Human Resource Management, 41,* 149-174.**

In this study, 5,000 human resource professionals were surveyed regarding the extent to which they agreed with various HR research findings. Responses from 959 participants suggested that there were large discrepancies between research findings and practitioners' beliefs in some areas, especially staffing. In particular, practitioners placed far less faith in intelligence and personality tests as predictors of employee performance than HR research would recommend. Practitioners were somewhat more likely to agree with research findings when they were at higher organizational levels, had SPHR certifications and read the academic research literature. Suggestions were made for more effective dissemination of HR research findings.

**Schmitt, N., & Borman, W. C. (1995).** *Personnel selection in organizations.* **San Francisco: Jossey-Bass.**

This book presents emerging issues in personnel research and practice. In the book, leading experts highlight the personnel selection issues that will receive increasing attention in the years ahead. The authors explore key subjects, including job analysis, performance measure development, biographical and personality measures, the concept of validity, the changing demographics of the work population, the decline of the manufacturing economy, and the development of small organizations. The book examines specific topics such as recruitment and retention, structured versus unstructured interviews, the ethics and effectiveness of computerized psychological testing, perceptions of selection fairness, productivity, turnover, and absenteeism. It also covers broader concerns, including downsizing and retirement, selection and staffing as a corporate strategy, promoting job and life satisfaction, organizational citizenship, and commitment.

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

This article summarizes the practical and theoretical implications of 85 years of research in personnel selection. It presents the validity of 19 selection procedures for predicting job and training performance and the validity of paired combinations of general mental ability (GMA) and the 18 other selection procedures. Overall, the three combinations with the highest validity and utility for predicting job performance are GMA plus a work sample test (mean validity of .63), GMA plus an integrity test (mean validity of .65) and GMA plus a structured interview (mean validity of .63). A further advantage of the latter two combinations is that they can be used for both entry-level selection and selection of experienced employees. The practical utility implications of these summary findings are substantial, and the implications of these findings for the development of theories of job performance are discussed.